# Genetic Diversity of Hepatitis B and Hepatitis C Viruses in Ethiopia

Gadissa Bedada Hundie

# Genetic Diversity of Hepatitis B and Hepatitis C Viruses in Ethiopia

Gadissa Bedada Hundie

The research described in this thesis was conducted at the Department of Viroscience, Erasmus MC, Rotterdam, The Netherlands within the framework of the Erasmus Molecular Medicine Post Graduate School. Parts of the studies in this thesis were financially supported by Netherland Fellowship programmes, Nuffic (grant number CF8605). The financial support for printing by I&I is gratefully acknowledged.

ISBN: 978-94-6295-924-8
Cover design: Gadissa Bedada Hundie
Lay-out and print by: ProefschriftMaken // www.proefschriftmaken.nl

© Gadissa Bedada Hundie, 2018. All rights reserved. No part of this thesis may be reproduced or transmitted, in any form or by any means, without the permission of the author.

# Genetic Diversity of Hepatitis B and Hepatitis C Viruses in Ethiopia

Genetische diversiteit van hepatitis B en hepatitis C virussen in Ethiopië

Thesis

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus

Prof.dr. H.A.P. Pols

and in accordance with the decision of the Doctorate Board.

The public defense shall be held on

Tuesday 19 June 2018 at 11:30 hours

By

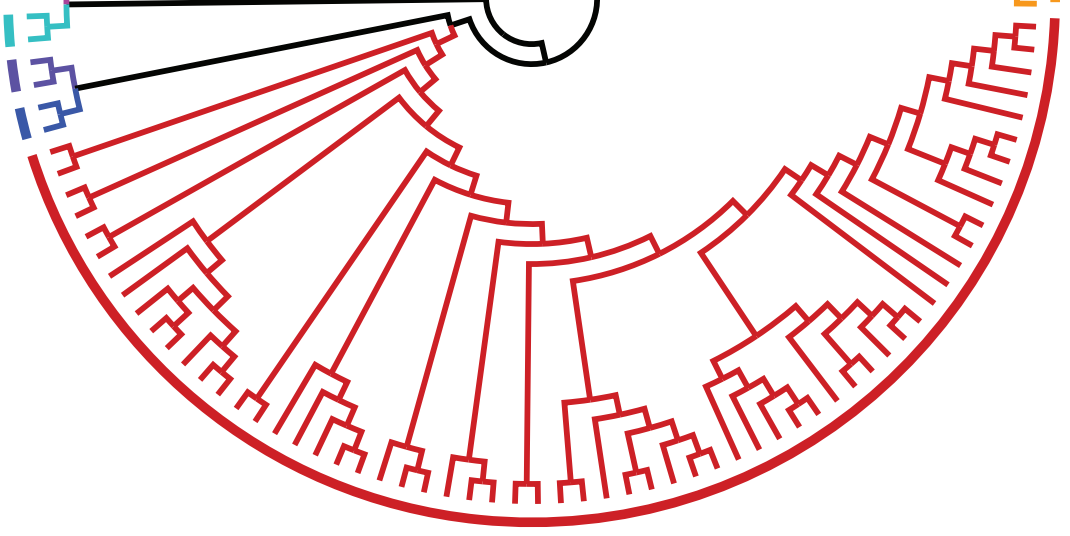
Gadissa Bedada Hundie
Born in Addis Alem, Ethiopia

**Erasmus University Rotterdam**

**DOCTORAL COMMITTEE:**

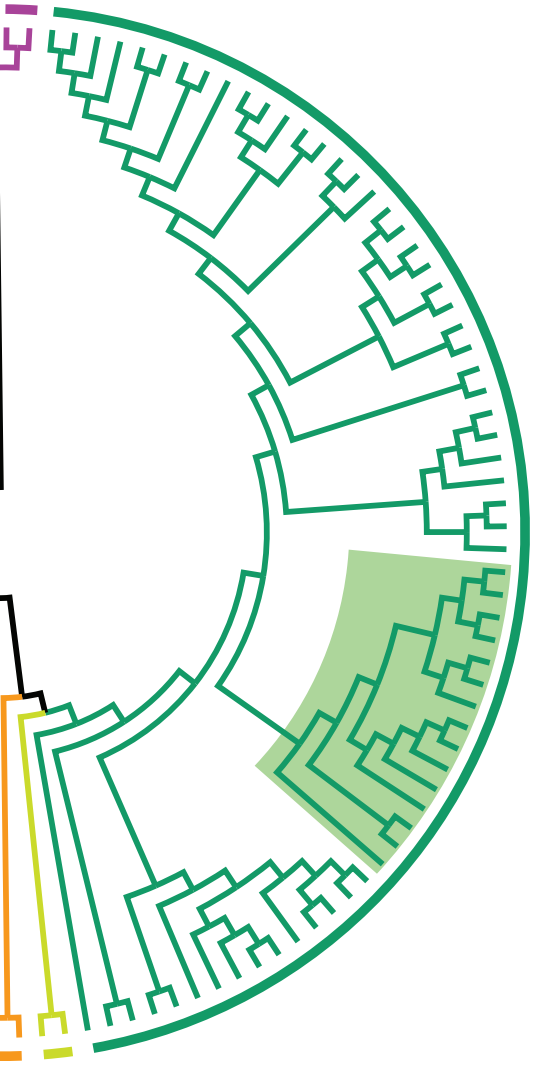
| | |
|---|---|
| Promotor: | Prof.dr. M.P.G. Koopmans |
| Other members: | Prof.dr. C.A. Boucher |
| | Prof.dr. R.A. de Man |
| | Prof.dr. H.L. Zaaijer |
| Copromotor: | Dr. B.L. Haagmans |

## TABLE OF CONTENTS



# Chapter 1

## General Introduction



### Viral hepatitis

Viral hepatitis is a group of infectious diseases that primarily affect the liver and cause serious illness and death to hundreds of millions of people worldwide. To date, there are five distinct hepatitis viruses, referred to as types A, B, C, D, and E around the globe. The history of modern study on viral hepatitis was dating back to the year 1963 when Nobel Prize winner Baruch Blumberg (1925-2011) and his team identified an unusual antigen from an Australian Aborigine blood, which they named the Australian antigen [1]. Some years later this was recognized to be the hepatitis B virus surface antigen (HBsAg), the first specific marker of viral hepatitis [2]. Thereafter, hepatitis B virus became a driving force for the discovery of hepatitis A and D in the 1970s, hepatitis C in 1989 [3], and hepatitis E in 1990 [4]. Except for their liver tissue tropism, all the five viral hepatitis are unrelated viruses belonging to different families, have different epidemiological profiles and vary in terms of their impact. Hepatitis B is a DNA virus whereas A, C, D, and E are RNA virus. Hepatitis A and E typically cause acute infections (although hepatitis E virus can be chronic in immunocompromised patients), usually obtained by fecal-oral transmission (including food and waterborne spread), while hepatitis B, C, and D are blood-borne viruses that are transmitted by parenteral or mucosal contact with infected blood or other body fluids, and cause both acute and chronic infections [5]. Infections with viral hepatitis remain major global health problems for decades and the morbidity and mortality because of viral hepatitis is increasing. In 2013 alone, viral hepatitis caused 1.45 million deaths, making it the seventh most important cause of death worldwide [6]. Together hepatitis B and C accounted for 96% of viral hepatitis-related mortality and 91% of viral hepatitis-related disability life-years. Hence, this thesis focuses exclusively on hepatitis B and C viruses, a leading cause of infectious disease-related morbidity and mortality, particularly in low-income countries.

## HEPATITIS B VIRUS

### Epidemiology

Hepatitis B virus (HBV) infection is a serious global public health problem causing significant morbidity and mortality. It is estimated that 2 billion people, one-third of the world's population, have been infected with HBV and more than 250 million are chronic carriers [7,8]. The largest number of chronic carriers live in the Western Pacific region (> 95 million, prevalence estimate 5.26%) and sub-Saharan Africa (> 75 million, prevalence estimate 8.83%), which together account for 70% of the global HBV burden [8]. Despite the availability of vaccine

and antiviral treatment, HBV still remains among the top 10 global leading causes of mortality and morbidity [9]. Each year approximately 1 million people die from chronic HBV-related consequences, including liver cirrhosis, end-stage liver disease, and hepatocellular carcinoma (HCC) [10]. HBV ranks tenth among all causes of mortality and HBV- related HCC is the sixth most common and second deadliest cancer worldwide [11-13]. Moreover, 30% of liver cirrhosis and 53% of HCC cases globally are attributed to HBV infection [14]

HBV prevalence (as measured by serology) and mode of transmission vary markedly worldwide, and the epidemiology of HBV changes over time. It is highly endemic (≥ 8% seroprevalence) in sub-Sharan Africa, South-East Asia, China, most Pacific Islands, and the Amazon basin, where ~45% of HBV-infected individuals live. The Middle East, Mediterranean region, Eastern Europe, and South America are considered areas of intermediate prevalence (2-7%) while the United States, Western and Northern Europe, and Australia are regions of low prevalence (< 2%) [7,15]. Roughly, 43% of infected individuals live in intermediate prevalence areas and 12% live in low prevalence areas [7]. HBV is transmitted by percutaneous and mucosal exposure to blood or body fluids from infected persons. The common routes of transmission related to this exposure are vertical transmission (defined as transmission during or immediately after pregnancy), or transmission through unsafe injection practices, sexual intercourse, and blood transfusions. In high prevalence areas, perinatal or household transmission during early childhood are the major modes of infection [7,15]. However, in sub-Saharan Africa, perinatal transmission is less common since most women of fertile age are HBeAg negative and early childhood horizontal transmission is the main route of HBV infection. In addition, transfusion-transmission is substantial in this region because of the frequent use of paid or replacement donors and incomplete donor screening coverage [16,17]. In intermediate prevalence areas, multiple modes of transmission like perinatal, household, sexual, and injection drug use contribute to the infection burden, while sexual and injection drug use in adults are the major routes of spread in low prevalence areas [15]. Overall, the risk of developing chronic HBV infection is 90% when infected perinatal, 20 to 30% when infected at childhood, and less than 1% when first infected as adults[18].

**Basic virology**

Hepatitis B virus is the prototype member of the family *hepadnaviridae*, a group of small, hepatotropic, DNA viruses that share similar genomic organization and replication strategy. The family *Hepadnaviridae* contains two genera; the *Orthohepadnaviruses* that infect mammals (human, woodchucks, ground squirrel and bats HBV), and the *Avihepadnaviruses* that infect birds (ducks, goose, and

herons HBV) [9,19]. Human HBV shares 70% nucleotide sequence homology with other mammalian hepadnaviruses and around 40% sequence homology with avian hepadnaviruses [20,21]. The human HBV is morphologically diverse and exists in human blood in three different viral structures: as infectious virions (Dane particles), and as noninfectious subviral particles that are either spherical (particles of 20 nm diameter) or filamentous (particles of 22 nm diameter but of variable length). The noninfectious subviral particles are composed of hepatitis B surface antigen (HBsAg) and host-derived lipids but lack genomic DNA. The infectious virion is a 42 nm double-shelled sphere and consists of an outer lipid envelope and an inner icosahedral nucleocapsid [9,22]. The envelope is composed of a double lipid layer and three envelope proteins, while the nucleocapsid is a 27 nm particle assembled from 240 viral capsid proteins that enclose the viral genome [12].

The viral genome is a relaxed, circular, partially double-stranded DNA (rcDNA) of approximately 3.2 kb and organized into four open-reading frames (ORFs) (Figure 1). The largest ORF (P) encodes the viral polymerase, which is functionally divided into four domains — the terminal protein, a spacer region, the reverse transcriptase (RT), and the ribonuclease H domain [22]. The polymerase enzyme catalyzes many functions during HBV replication, including viral RNA binding, RNA packaging, protein priming, template switching, DNA synthesis and RNA

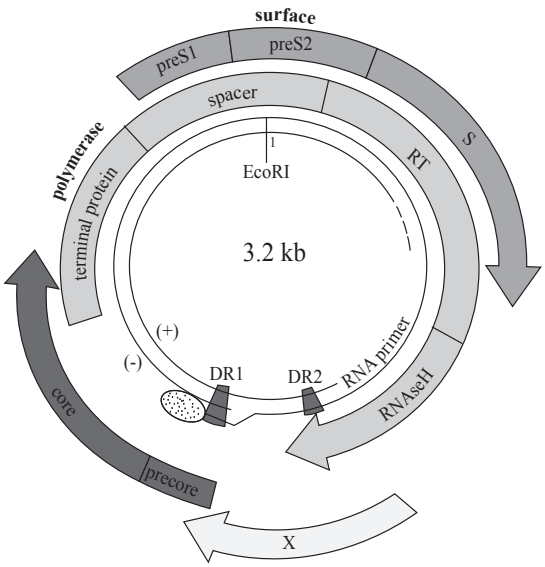



**Figure 1. HBV genome organization.** The genome nucleotide numbering is based on the unique EcoRI restriction enzyme site. The four different open reading frames encoded by the genome, the S, core, polymerase, and X, are indicated by gray color palette.

degradation [23]. The second largest ORF (S) encodes three viral envelope proteins — large (L), middle (M) and small surface antigen (HBsAg) protein — that are encoded by three in-frame genes (the pre-S1, pre-S2 and S gene, respectively). HBsAg is the serological marker for current HBV infection and the main antigenic determinant for serotyping and diagnosis [9]. The antigenic determinant referred as the "a" determinant is located between amino acid positions 124 and 147 of HBsAg and plays an important role in diagnosis and immunoprophylaxis. The pre-S1 domain is a key determinant in hepatocyte receptor binding (residue 21-47) and initiation of viral infection [24]. The preS2 region contains only 165 nt and shows the highest nucleotide polymorphism among the S regions. The third ORF (C) encodes the precore, also known as HBV E antigen (HBeAg), and the core protein (HBcAg), which makes up the viral capsid. HBeAg is a non-structural protein secreted by infected cells into the blood and serves as a surrogate marker for active high-level viral replication [12]. The core protein contains numerous T and B-cell immune recognition epitopes and plays a critical role in the interactions between HBV and the host immune response, which in turn shapes the course of the disease [25]. The fourth ORF (X) encodes the HBV X protein (HBx), a small regulatory protein (154 amino acids) with multiple functions, including signal transduction, cell cycle progress and viral replication [13].

In addition to the four ORFs, there are other elements that are important components of the HBV genome. The short RNA-primer is used for the DNA positive strand synthesis during rcDNA repair to form a covalently closed circular DNA. The DR1 and DR2 represent 11 base pair direct repeats that play a pivotal role in template switches during viral replication.

The HBV replication cycle is complex and unique because it replicates via reverse transcription of an RNA intermediate like retroviruses. The initial step in HBV replication involves the attachment and entry of mature virions to host hepatocytes through the sodium taurocholate cotransporting polypeptide (NTCP) receptor on the cell membrane [24]. Once in the cytoplasm, the nucleocapsid is released from the viral envelope and transported to the nucleus where the rcDNA is repaired by host and viral polymerase to form a covalently closed circular DNA (cccDNA). The cccDNA, which forms a minichromosome, is a template for the transcription of all viral RNAs including the pregenomic RNA (pgRNA, a 3.5 kb RNA intermediate, which serves as a translation template for HBV core and polymerase proteins and is also the template for reverse transcription to viral DNA) (Figure 2). These transcripts are then exported into the cytoplasm where they are translated to form the viral proteins. In the cytosol, RT binds to pgRNA and triggers assembly of the core proteins into immature RNA-containing nucleocapsids. The immature nucleocapsids then undergo a process

of maturation whereby pgRNA is reversed transcribed by RT to make the viral rcDNA genome. The newly formed nucleocapsids are either re-imported to the nucleus to replenish the cccDNA pool or enveloped and secreted as mature virions [12,22,26]. A schematic overview of HBV replication cycle is shown in Figure 2.

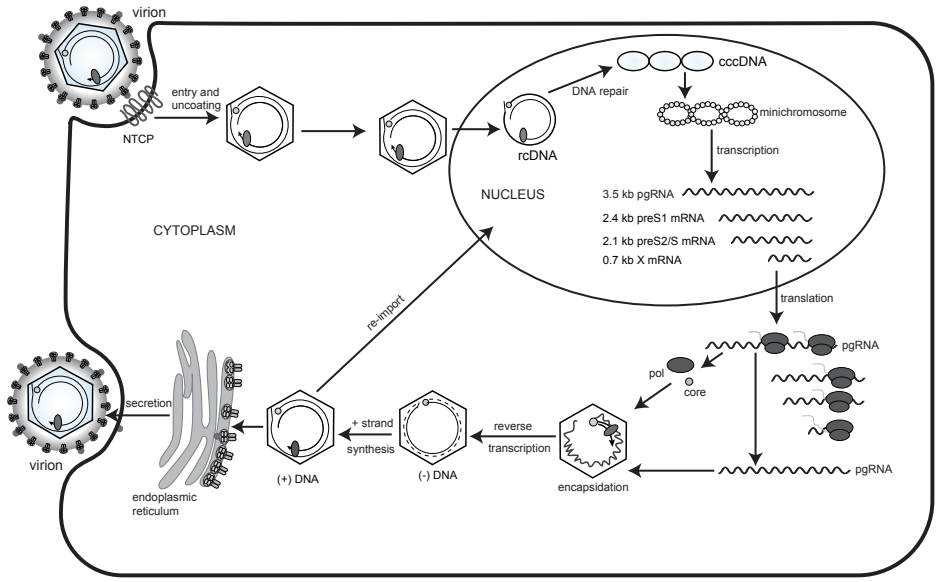



**Figure 2. Hepatitis B virus replication cycle.** Enveloped virion enters hepatocytes through NTCP, followed by uncoating, and nuclear transport of the rcDNA. The rcDNA is repaired to yield cccDNA, which serves as template for transcription of pgRNA & mRNAs. These RNAs are exported to cytoplasm for protein translation.

## HBV genetic diversity

Similar to other viruses, genetic variability and genome evolution are common features of the HBV life cycle. HBV is unique among DNA viruses because its polymerase lacks proof-reading activity and uses an RNA (pgRNA) intermediate during its replication. These features increase the probability of random errors during genomic replication, leading to change in genome sequence over time [9]. It has been estimated that HBV genome evolves at an error rate of ~ $10^{-3}$ to $10^{-6}$ nucleotide substitutions/site/year, which is approximately 100 times higher than that of other DNA viruses [27]. The accumulation of these mutations over a long period of time results in a large amount of genetic diversity in the form of genotypes, subgenotypes, serotypes, and mutants. In addition, HBV genomes evolve by recombination [28]. Prior to the definition of genotypes, HBV was

classified by serotyping, which discriminates HBV strains into nine serotypes, *ayw1*, *ayw2*, *ayw3*, *ayw4*, *ayr*, *adw2*, *adw4*, *adrq+* and *adrq-*, based on the main antigenic determinant 'a' and two mutually exclusive determinant pairs (d/y and w/r) in the HBsAg protein. The molecular bases for both the d/y and w/r changes were mediated through a shift from Lysine to Arginine at positions 122 and 160, respectively [29,30]. The advent of molecular biology later enables classification of HBV strains by genomic sequences rather than by serotyping. Okamoto et al [31] first analyzed 18 full-length HBV genome sequences and classified them into 4 distinct genotypes, A to D, based on greater than 8% divergence between genotypes. Since then, six additional genotypes, E to J, have been identified [30,32-36]. HBV genotypes are further divided into subgenotypes based on nucleotide divergence of more than 4% but less than 8% over the complete genome [30,37]. By this criteria, different studies worldwide have identified several subgenotypes within genotypes, A (A1-A7), B (B1-B9), C (C1-C16), D (D1-D9), F (F1-F4) and I (I1 and I2) [30,35,38-41]. It is noteworthy to mention that genotype C has 16 subgenotypes and is probably the oldest genotype while genotypes E, G, and H have no subgenotypes, suggesting their recent origin [42]. There is a significant association between HBV serotypes and genotypes: *adw* is associated with genotypes A, B, F, G, and H, *adr* with C and *ayw* with D & E [37] although many exceptions exist (Table 1).

Besides differences in genome sequences, genotypes also differ in their genome length, the size of open reading frames, and the proteins decoded (Table 1). Genotypes B, C and I have a genome size of 3215 nucleotides whereas genotypes A, E, and G have a genome size of 3221, 3212, and 3248 nucleotides, respectively. Genotypes D and J have the smallest genome size of 3182 nucleotides [43]. The differences in genome size are due to in-frame insertions or deletions. For instance, genotypes A and G are characterized by 6 and 36 nucleotides insertion in the core gene, respectively while genotype D and J have 33 nucleotides deletion at the N-terminus of the preS1 region [27,35]. As shown in both *in vitro* and *in vivo* study, these virological features have marked genotype-dependent differences in the expression of HBV DNA levels (viral replication) and antigens (HBeAg, HBsAg) [44]. These, in turn, may have contributed to differences among genotypes in clinical and epidemiological settings.

**Table 1. Characterstic of HBV genotypes and their correletion with serotypes**

| Genotype | Genome length (nt) | Differentiating features | Protein | | | Subgenotypes | Serotypes |
|---|---|---|---|---|---|---|---|
| | | | preS1 | Pol | Core | | |
| A | 3221 | 6 nt insertion near the 3′ end of core gene | 119 | 845 | 185 | A1, A2<br>A3 - A6<br>A7 | *adw2*<br>*ayw1*<br>*ayw1/adw2* |
| B | 3215 | | 119 | 843 | 183 | B1- B3, B5 - B9<br>B4 | *adw2*<br>*ayw1/ dw2* |
| C | 3215 | | 119 | 843 | 183 | C1-C3, C6 -15<br>C4<br>C5<br>C16 | *adr*<br>*ayw2/ayw3*<br>*adw2*<br>*ayr* |
| D | 3182 | 33 nt deletion at the 5′ end of the preS1 region | 108 | 832 | 183 | D1, D4, D7<br>D2<br>D3, D5 | *ayw2*<br>*ayw3*<br>*ayw2/ayw3* |
| E | 3212 | 3 nt deletion at the 5′ end of the preS1 region | 118 | 842 | 183 | - | *ayw4* |
| F | 3215 | | 119 | 843 | 183 | F1<br>F2 - F4 | *adw4/ayw4*<br>*adw4* |
| G | 3248 | 36 nt insertion at the 5′ end of core gene and a 3 nt deletion at the 5′ end of preS1 region; stop codons at positions 2 & 28 in precore | 108 | 842 | 195 | - | *adw2* |
| H | 3215 | | 119 | 843 | 183 | - | *adw4* |
| I | 3215 | | 119 | 843 | 183 | I1<br>I2 | *adw2*<br>*ayw2* |
| J | 3182 | 33 nt deletion at the N-terminus of the preS1 region | 108 | 832 | 183 | - | *ayw3* |

It is increasingly evident that recombination is a very important source of HBV genetic variability and evolution, with possible clinical implications [21,45]. HBV intergenotypic recombinants are generated either from two or more genotypes. As most recently reviewed by Araujo [43], nearly 60% of the recombinants detected were either B/C or C/D hybrids. Other recombinants identified included A/D, D/E, A/B/C, A/C, A/C/G, A/E, A/G, B/C/U (U = unknown genotype),

C/F, C/G, C/J, D/F, and F/G hybrids. In addition, a large-scale phylogenetic analysis by Shi and colleagues [46] showed that only genotypes H and J showed no recombination, whereas genotype I was composed entirely of recombinants (A/C/G) and 93% (803/860) of genotype B were recombinants (mainly with C in Asia where both genotypes are prevalent). Recombination may provide a selective advantage to the viral strains, and the recombinants over time become the dominant strains of HBV and persist in a population in respective geographic areas. Generally, HBV recombinants were reported in almost all regions of the world with similar circulation patterns as their original genotypes [38,40,41,43,47-50]. The B/C recombinants are most prevalent in East and Southeast Asia, A/D and A/E recombinants in Africa and C/D recombinant in Tibet (96% of the isolates sequenced were C/D recombinants) [51]. Although not commonly reported, Simmonds and Midgley showed the evidence of frequent intragenotypic recombination between members of genotypes A, D, and F/H variants, but not in B and C, in HBV strains of Southeast Asia [45].

**Geographical distribution of HBV genotypes and subgenotypes**

HBV genotypes and subgenotypes have a distinct ethno-geographic distribution worldwide [30,52] (Figure 3). Genotype A is widespread around the world but is most prevalent in sub-Saharan Africa, Europe, and North America [27]. Genotypes B and C are most common in Asia and Oceania [30]. Genotype D is predominant in the Mediterranean areas, the Middle East, and India, while genotype E is endemic to western and central Africa [52]. Genotype F is most frequent in South and Central America whereas genotype H is endemic to Central America. Little is known about the distribution of genotype G although sporadic cases have been reported in Europe, USA, and Japan chiefly among homosexual men [53]. Genotype I has been reported in Vietnam and Laos and genotype J in a Japanese patient [53].

Like genotypes, HBV subgenotypes also display distinct geographical distributions. Subgenotype A1 is dominant in Southern and Eastern Africa, A3 to A7 are common in Central and Western Africa, and A2 is dominant in Europe and North America [54]. Subgenotype B1 is common in Japan, B2 to B5 in East and Southeast Asia [30] whereas B6 is found in indigenous populations living in Arctic regions, including Alaska, Northern Canada, and Greenland. Subgenotypes B7, B8, and B9 are mainly found in Southeast Asia islands [55]. While subgenotype C4 is restricted to Australian Aborigines, C1 is widespread in East Asia, C2 in China and Southeast Asia, and C3 in Oceania [30]. The geographical distributions of subgenotypes D1-D4 are less defined than those for genotypes A, B, C and F [30]. Subgenotype D1 dominates in the Middle East and Asia, D2 in Europe, D3 in South Africa, Asia and Europe, D4 in Oceania, D5and D9 in India, D7 in North

Africa, and D8 in Niger [53,56]. Having a knowledge of the prevailing genotype and subgenotype in a given geographic region is important as these genotypes and subgenotypes differ in disease progression, clinical outcomes, prognosis and response to antiviral therapies [54,57,58]. For instance, in Africa, genotype A1 is associated with HCC in young men who are usually HBeAg-negative and without underling cirrhosis [59]. Most studies (mainly in Asia) showed that patients infected with genotype C have more severe liver disease, including cirrhosis and HCC, than those with genotype B [54,60]. Similarly, antiviral treatment response to standard or pegylated interferon is more favorable in patients infected with genotype A than D and in B than those with genotype C [61,62]. It is important to note that the rates of disease progression, clinical and treatment outcomes for each HBV genotype are also influenced by environmental, host and other viral factors.

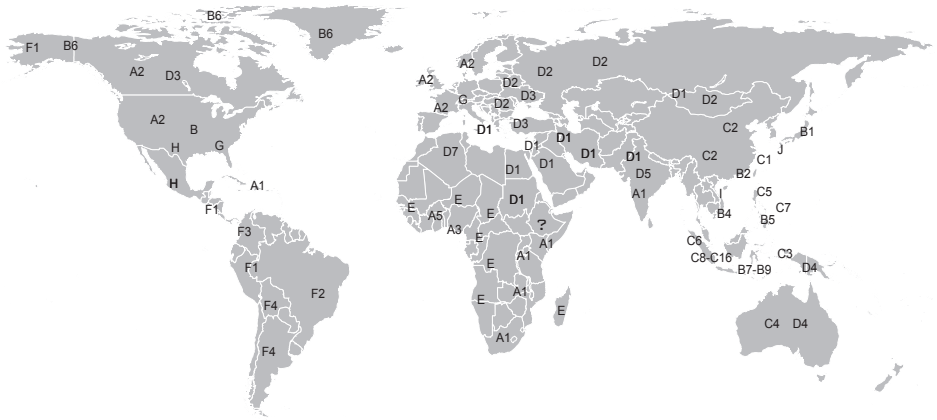



**Figure 3. Global distribution of HBV genotypes and subgenotypes.** In Africa, genotypes A, D and E predominant.

## HBV quasispecies variants

HBV quasispecies variants — defined as a population of genetically distinct but closely related variants — emerge in response to external selective pressures such as host immune responses, prophylactic or therapeutic responses (vaccination, antiviral treatments) [21]. They are generally less fit than the normal genotypic variants (wild type) and do not emerge as dominant viral populations in patients in the absence of selective pressures. However, under the external selective pressure, the minor mutants can be selected and become the dominant variants [28]. Some of these quasispecies variants include the preS1 and S2 mutants, the HBsAg vaccine escape mutants, the core, precore (PC) and basal core promoter (BCP) mutants, and the polymerase drug-resistant mutants [21].

Many of the mutations affecting the preS1 and preS2 domains of the envelope proteins are deletions [28]. Some of these deletions are harmful to the virus while others are beneficial for its survival [63]. For instance, a mutation in the N-terminus of the preS1 affects the virus as this domain is a key determinant for virus receptor binding and entry [24]. In addition, some preS1mutations can result in intracellular retention of the preS1 protein, which inhibits virion secretion. As a result, most of such mutants are usually found as minor viral populations [21]. In contrast, deletion in the preS2 and in the C-terminus of preS1 result in reduced expression of HBV surface proteins and help viral persistence by eliminating HLA-restricted B-cell and T-cell epitopes [63]. Many of such mutants have clinical importance, including influencing the risk of HCC, fulminant hepatitis, and cirrhosis [63-65]. The HBsAg, in particular the major hydrophilic region (MHR; amino-acids 99-169), is the main target for viral neutralization [28]. Mutation in the MHR can lead to immune and vaccine escape mutants that exist as minor/dominant viral populations depending on selective pressure. Several immune- and vaccine escape variants of clinical importance have been identified worldwide, but G145R is by far the most common vaccine escape variant, and the most studied [28,63]. Further, variants within MHR region have been shown to be associated with the occurrence of HBsAg escape mutants that are undetectable by current commercial diagnostic assays [66,67].

Various naturally occurring mutants have been described in the BCP (nt1814-1900) and PC (nt1742–1849) regions, some of which are defective variants and require helper virus while others are competent stable variants [21,65]. The most common PC mutation is a G to A transition at nucleotide position 1896 (G1896A), which creates a premature stop codon at codon 28 that completely abolishes HBeAg synthesis but increase viral replication [21]. The most frequent BCP mutation is the double A1762T and G1764A nucleotide substitution, which decreases HBeAg production by up to 70% through suppressing the transcription of precore mRNA, but enhances viral replication [28,65]. The consequent decrease or abolishment of the HBeAg production lead to selection of anti-HBe antibody escape mutants and subsequently favoring the persistence of HBV infection [68]. Several studies have shown that the persistent BCP and PC mutants are clinically important because they increase the risk of HCC, cirrhosis, fulminant hepatitis, and because they are difficult to treat [64,69-71].

The RT domain which is the main catalytic center of the polymerase enzyme is the main target for most antiviral drugs, particularly the nucleos(t)ide analogs. Mutations within RT may affect the replication capacity of the viruses, leading to the emergence of drug-resistant mutants that can exist as minor or major variants [28,72]. For example, rtM204I/V, rtL180M and rtA181T/V mutants confer resistance to lamivudine and telbivudine whilst rtA181T/V and rtN236T mutants confer

resistance to adefovir and tenofovir [28,65,73]. HBV harboring I169T and M250V, or T184G and S202I variants have been associated with entecavir resistance [74]. In addition, drug-resistance mutations in the polymerase may result in the production of mutations and stop codons in the HBsAg envelope and lead to altered viral secretion, infectivity, antigenicity, and vaccine escape mutants because of the overlap between the two genes [73]. For instance, A181T/V mutations in the RT region can cause W172* (stop codon), W172L and L173F mutations while M204V/I mutations result in I195M, W196*, W196S and W196L mutations in the HBsAg [28,73].

## HEPATITIS C VIRUS

### Epidemiology

Hepatitis C virus (HCV) is a global health problem that causes substantial morbidity and mortality. It is estimated that more than 185 million people are chronically infected with HCV worldwide and at great risk of developing liver disease [75]. Each year, 3 to 4 million people are newly infected [76] and around 700,000 people die from HCV-related liver diseases [77]. HCV is a major global health burden because it accounts for 20%-30% of acute hepatitis, 70%-80% chronic hepatitis that subsequently leading to 20-40% cirrhosis, 25% HCC, and 30-40% liver transplants worldwide [78,79]. Moreover, HCV is one of the most common causes of death in HIV-positive patients on highly active antiretroviral therapy [80]. Furthermore, the most recent estimation of disease burden shows an increase in global HCV seroprevalence over the last 15 years from 2.3% to 2.8%, with great variation in different geographic areas. Africa has the highest HCV burden with an overall prevalence of 5.3% or an estimated 32 million chronic carriers [75]. Within Africa, Central Africa has the highest estimate prevalence at 6%, followed by Western Africa 2.4% while Southern and Eastern Africa have the lowest estimated at 1.6% [5]. Most of the European countries, America, and Southeast Asia have a prevalence < 2.5% [78], whereas in the Western Pacific region the prevalence is 3.9% [78]. The prevalence of HCV in the Middle East and Central Asia countries varies between 1 and 12% with an overall prevalence rate of 4.7% [81]. Worldwide, Egypt has the highest HCV prevalence (>15%) and the United Kingdom and Scandinavia countries have the lowest prevalence (0.01-0.1%) [82].

### Basic virology

Hepatitis C virus is a single-stranded positive-sense RNA virus classified in the *Hepacivirus* genus within the *Flaviviridae* family [83]. It was discovered in 1989 by

the joint efforts of researchers at the centers for disease control and prevention (CDC) and Chiron Cooperation [84]. Until recent time it has been considered the sole species within the genus *Hepacivirus*. However, most recently numerous genetically diverse *hepaciviruses* have been identified in bats, dogs, cows, horses, primates and rodents [85], which now classify the genus into 14 recognized species (*Hepacivirus A–N*) [86]. Humans are the only natural host and apparent reservoir of HCV although it can be transmitted experimentally to non-human primates [83]. HCV is spherical in shape, 40–100 nm in diameter [87] and consists of a single RNA genome encapsidates in core proteins, which in turn enclosed by the envelope glycoproteins, E1 and E2 [88]. The HCV genome interacts with the capsid protein to form the nucleocapsid that is surrounded by a lipid membrane in which the envelope glycoproteins are anchored.

The HCV genome is a linear, non-segmented, single-stranded RNA of ~ 9.6 kb in length (Figure 4). It encompasses a single open reading frame encoding a long polyprotein of 3000 amino acids, which proteolytically cleaved by cellular and viral proteases into 10 proteins. These include 3 structural proteins (core, E1, and E2), and 7 nonstructural (NS) proteins ( p7, NS2, NS3, NS4A, NS4B, NS5A, and NS5B) [89,90]. These proteins have different functions and are encoded in the gene order of 5′-C-E1-E2-p7-NS2-NS3-NS4A-NS4B-NS5A-NS5B-3′ (Figure 4). The ORF is flanked in 5′ and 3′ by two highly conserved untranslated regions (UTRs) of 341 and 230 nucleotides, respectively [81] that are essential for viral replication and translation. The core protein is a highly basic multifunctional protein, which is responsible for packaging viral RNA to form a viral nucleocapsid [81,88]. The viral envelope, E1 (33–35 kDa) and E2 (70–72 kDa) are type I transmembrane glycoproteins that mediate viral entry and fusion [88,91]. In addition, the E2 gene constitutes the hypervariable region 1 (HVR1), which is the most variable region of the genome and appears to be involved in immune escape and disease progression. p7, an integral membrane protein, is a member of the viroporin family which plays a critical role in capsid assembly and envelopment. NS2 (21–23 kD) is a non-glycosylated transmembrane protein that together with NS3 forms a metalloproteinase which cleaves the site between NS2 and NS3 and contributes to virus particle assembly [81,91]. NS3 is a 67 kDa multifunctional protein with a serine protease, viral helicase and NTPase activities that antagonize the host innate immune response [88,92]. The NS4A (8 kDa) is a cofactor of NS3 [91] while NS4B is a membrane-associated protein that serves as an anchor for the viral replication complex [88]. NS5A is a 58 kDa phosphoprotein that appears to play a vital role in RNA replication and regulation of cellular pathways. The NS5B protein (68 kDa) is the viral RNA-dependent RNA polymerase (RdRp), an enzyme that catalyzes viral replication [88].

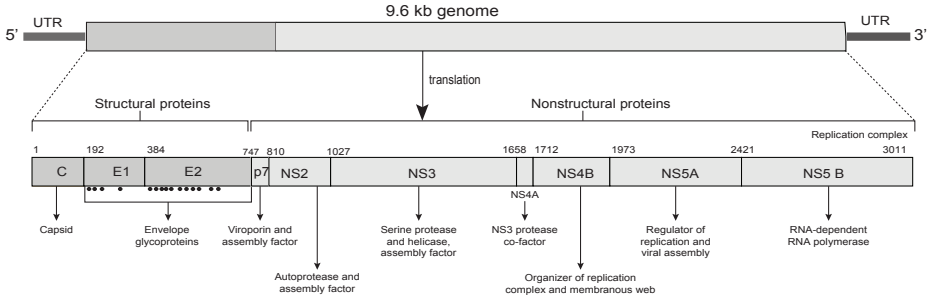



**Figure 4. HCV genome organization and polyprotein.** The genome is composed of a single open reading frame flanked by 5' and 3' untranslated regions. The ~ 9.6 kb genome is translated in a single polypeptide of 3000 amino acids, which proteolytically cleaved by cellular and viral proteases into 10 proteins.

## HCV genetic diversity

Genetic variability is a distinctive characteristic of HCV and exists at several levels throughout the genome. This high genetic diversity is primarily caused by the error-prone nature of the RdRp that lacks proofreading ability, combined with the high viral replication rate ($10^{10}$- $10^{12}$ virions per day) [83,93]. The resulting genetic heterogeneity is reflected by the high mutation rates, 1.5–2.0 x$10^{-3}$ nucleotide substitutions per site per year [90], which led to the classification of HCV into different groups. Through phylogenetic and sequence analyses of the entire genome, HCV has been classified into seven confirmed genotypes, 1-7 [94]. These genotypes differ from each other by 30–35% at the nucleotide level [95]. All genotypes except 5 and 7 are further divided into multiple epidemiologically distinct subtypes, denoted by lower-case letters 1a, 2c, 3d, 6f, etc., that differ at least by 15 % at nucleotide level [94,96]. At present, there are 67 confirmed, 20 provisional, and 21 unassigned subtypes in HCV [94]. While genotypes are easy to discrete genetically, subtypes discrimination is less clear, particularly in areas of high HCV genetic diversity such as sub-Saharan Africa and Southeast Asia. This highlights that much of genetic characterization is needed in this part of the world.

In addition to genotypes and subtypes, HCV circulates in infected individuals as a complex mixture of genetically distinct but closely related variants referred to as quasispecies, which differs by 1–3% at nucleotides level [76]. This quasispecies exists as a continuously evolving mutant spectrum, i.e., an array of viral variants, which confers the advantage of a rapid adaptation to a changing environment, a strategy utilized by the virus to escape selective forces such as antiviral agents or the host immune system [97]. Study revealed that there are $10^{10}$–$10^{12}$ variants produced per day in each infected individual, which leads to highly individualized quasispecies

that are unique because of the individual's immune response to HCV, and that may play a role in disease progression and treatment response [98].

Although not common like in many other RNA viruses and HBV, a few naturally occurring HCV recombinants have been documented worldwide [46,99]. To date, around ten inter-genotype recombinants have been identified, of which only one (2k/1b) is represented by multiple isolates originated in different countries [99,100]. It is likely that recombination in HCV might be underestimated due to difficulties in detection of recombinant HCV strains [101].

**Geographical distribution of HCV genotypes and subtypes**

HCV genotypes and subtypes display distinct geographical distribution and prevalence worldwide [83,96] (Figure 5). HCV genotype 1 is the most prevalent genotype worldwide accounting for 46% of all HCV cases. Genotype 3 is the second most prevalent with 22% of all HCV cases, followed by genotype 2 (13%), genotype 4 (13%), genotype 6 (5.4%), and genotype 5 (< 1%) [96,102]. The global distribution of these genotypes shows geographic variations, which reflect differences in transmission history and ethnicity. Overall, genotypes 1-3 are widely distributed across the globe, while 4 – 7 have relatively restricted geographic distribution (Figure 5) [96]. HCV genotype 1 is the most dominant genotype in North America (70% of cases), Japan (75%) and Europe (50–70%); although genotypes 2 and 3 are also prevalent in these regions [83,102]. On the other hand, genotype 2 is predominant in West Africa, genotype 3 in South Asia and parts of Scandinavia, genotype 4 in Central and North Africa and the Middle East, genotype 5 in South Africa, and genotype 6 in Southeast Asia [96], while genotype 7 only presents in Central Africa [103]. Like genotypes, most subtypes also show variation in their geographic distribution. Subtypes, 1a, 1b, 2a, 2b, 2c, and 3a are widely distributed across the globe and account for the majority of HCV infections worldwide, therefore termed epidemic. They are thought to have spread rapidly during the twentieth century via infected blood and blood products, intravenous drug use, and invasive medical procedures [96,104,105]. In contrast, the other HCV subtypes have long-term local persistence and restricted geographic distribution, hence endemic [106]. For instance, subtypes belonging to genotypes 1, 2, and 4 are endemic in Africa, while those belonging to genotypes 3 and 6 are endemic in South and South East Asia [96,107].

Africa is known for its high HCV genetic diversity pool and thought to be the origin of this virus. Almost all the known HCV genotypes circulate in the continent and six of them are proposed to have originated here: genotype 1 and 2 in West Africa [106,108], 4 and 7 in Central Africa [103,105], 3 and 5 in South Africa [109,110]. Within Africa, HCV genotype distribution exhibits two epidemiological patterns

— one characterized by high genetic diversity, distributed in West Africa, with genotypes 1 and 2 [106], and the other in Central and Norther Africa, with genotype 4 [105]. A single subtype, 5a, dominates in Southern Africa [102], and very little is known about the genotype distribution in Eastern Africa, including Ethiopia.

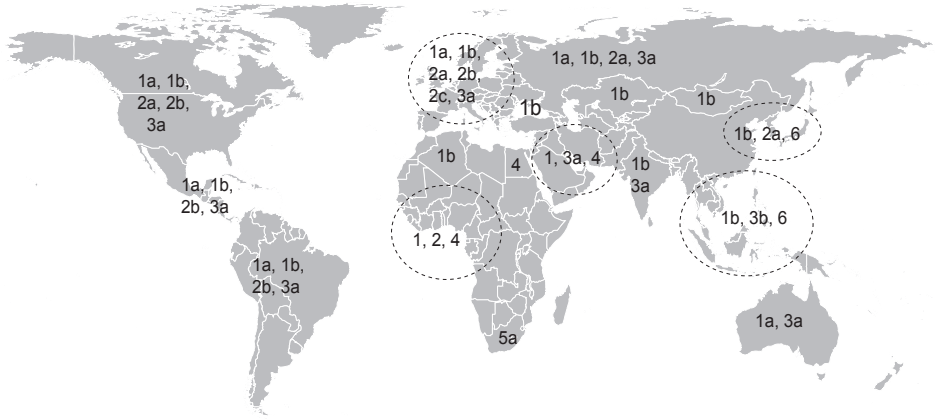



**Figure 5. Global distribution of HCV genotype and subtype**

### Viral hepatitis B and C infection in Ethiopia

Ethiopia is a large, culturally diverse country in the Horn of Africa located at a strategic geographic position between Africa and Eurasia. It shares borders with Eritrea to the north, Djibouti and Somalia to the east, Sudan and South Sudan to the west and Kenya to the south. It has a land area of 1.1 million square kilometers and is home to more than eighty different ethnic groups. With a population of nearly 100 million persons, it is the second most populous country in Africa and the twelfth populous in the world. Around 80% of its population resides in rural areas where health infrastructures are limited, and the delivery of essential health services remains challenging. As a result, little information is available on the public health burden due to infectious diseases in general and hepatitis viruses in particular. Owing to the asymptomatic nature of HBV and HCV infections for decades, most people don't know that they are infected. Also, the cause of deaths usually remains undocumented in Ethiopia.

Nevertheless, a clinical study performed in the 1980s showed that liver disease accounted for 12% of hospital admissions and 31% of deaths of hospitalized patients in Ethiopia [111,112]. Moreover, seroepidemiological studies, although quite small in number, show heterogeneity in seroprevalence estimates for HBV (2.1 to 25.0%) and HCV (0.7 to 13.3%), even within specific geographical areas [113]. Most of these reports are either from a limited number of subjects, or are biased for

specific groups and may not represent the true national burden of viral hepatitis in the general population. As a result, the burden of viral hepatitis in Ethiopia is not accurately known. Besides, there is a lack of awareness among the general public concerning viral hepatitis, with underdiagnosis and therefore undertreatment. Moreover, HBV and HCV are blood borne pathogens, and practices contributing to their transmissions such as shared materials during traditional tattooing, scarification and body piercing, circumcision and female genital mutilation are common in Ethiopia. There is a continuous influx of refugees from South Sudan where HBV prevalence is hyperendemic (prevalence estimate 22.38%, 95% CI 20.10–24.83), further adding to the unknown reservoir of carriers [8]. As a result of these factors, transmission of these viruses may go unnoticed.

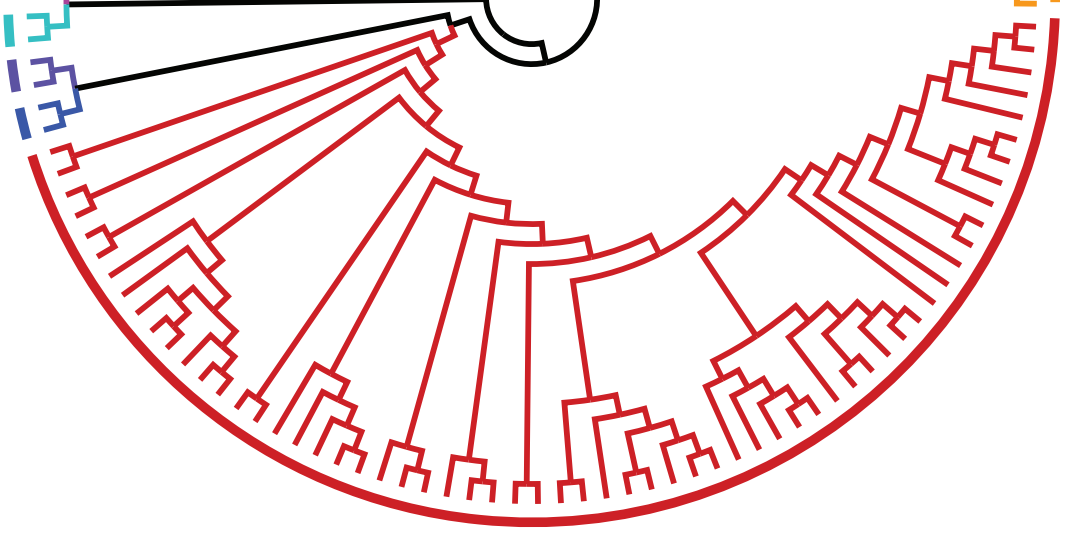
It is well-known that HBV and HCV show great genetic diversity [53,94]. As shown by many studies, this genetic diversity has an impact on clinical outcomes and treatment responses [28,60,98]. However, there is very limited information regarding the genetic diversity of these viruses in Ethiopia. To establish effective strategies for the control and prevention of these viruses, knowledge of the prevailing genotype and subgenotype in Ethiopia is highly important. Thus, this thesis focused on this and other related issues as outlined below.

**Outline of the thesis**

Since HBV and HCV infections can go without clinical symptoms for decades, the prevalence and disease burden are generally underestimated. In addition, their genetic diversity is not fully studied. Therefore, this thesis in its first part, chapter 2-5, focuses on understanding the viral diversity and seroepidemiology of HBV and HCV in Ethiopia, to provide a baseline for predicting clinical outcomes, disease progression, and response to antiviral. Then the second part, chapter 6 and 7, of the thesis is focusing on a further understanding of HBV/HCV quasispecies variants through in-depth molecular analyses.

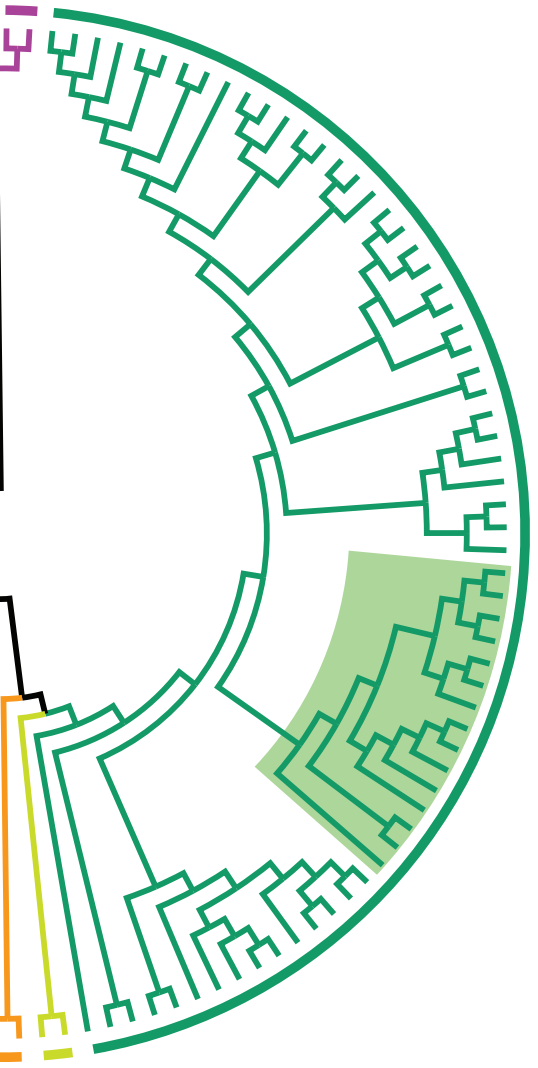
Thus, in **chapter 2** of this thesis, we describe the seroepidemiology of HBV and HCV in five distinct geographic regions in Ethiopia that differ in socioeconomic, culture and ethnicity to gain insight into the burden of these viruses in the general population. **Chapter 3** focuses on the molecular epidemiology and genetic diversity of HBV in Ethiopia and emphasizes the main genotypes, subgenotypes, serotypes and mutants that circulate, based on study conducted using S gene sequence of 391 sera obtained from blood donors living in five different geographic regions. Here we discussed differences in HBV genotype/subgenotype distribution among Ethiopian regions and compare those with data from other African countries and the world. **Chapter 4** describes in-depth complete genome sequence analysis of 29 selected HBV strains; 10 genotypes A and 19 genotypes

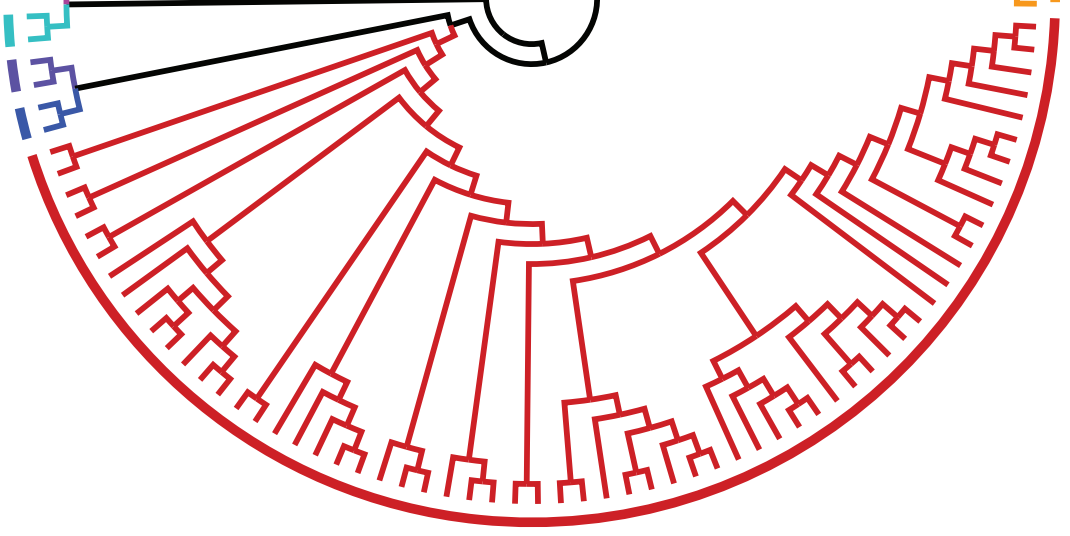
D. These strains were selected based on their unique features in the previous fragment analysis including separate clusters, divergent strains and unusual strains (like D2 which is less common in Africa) to classify them into subgenotypes and to gain more insight into the molecular diversity and evolution of HBV in Ethiopia. Here, a novel subgenotype D10 was identified and discussed thoroughly and the two genotype A1 clades and the D2 strains were further elaborated. In addition, mutations in the basal core promoter (BCP), precore (PC) and preS regions that have a role in the clinical manifestation of disease following HBV infection were characterized and two natural recombinant strains noted. Then **Chapter 5** describes HCV genetic diversity in Ethiopia; the main genotypes and subtypes and their implication in the future development of treatment guidelines and patient management. Thereafter **Chapter 6** discusses the HBV quasispecies variants and their association with clinical characteristics such as HBeAg and viral load. It further describes HBV variants variation between HBeAg positive and negative patients and among different genotypes. **Chapter 7** investigates the pre-existence of direct-acting antivirals (DAA) resistance-associated variants in the NS5B region of treatment-naïve patients. Further, it assesses whether there is a difference in quasispecies variants distribution between plasma and liver compartments of the same patient. Finally, the main findings of this thesis are summarized and discussed in **Chapter 8**.



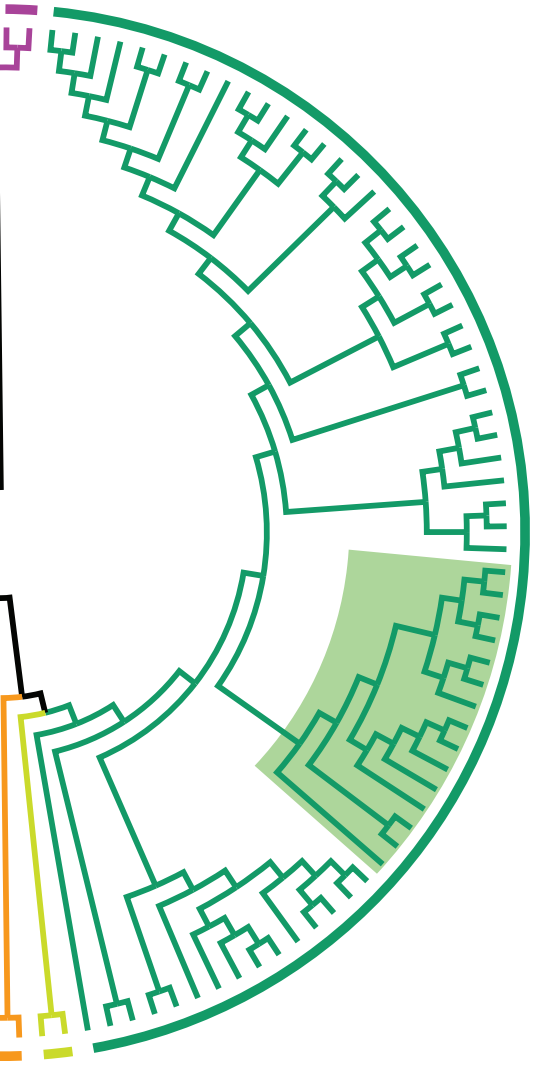
# Part I

## Genetic diversity of HBV and HCV in Ethiopia





# Chapter 2



## Seroepidemiology of hepatitis B and C virus infections among blood donors in Ethiopia

Gadissa Bedada Hundie, V. Stalin Raj, Daniel GebreMichael, Bart L. Haagmans

*J Med Virol. 2017;9999:1–4. Doi:10.1002/jmv.24770*

## ABSTRACT

This study was conducted to investigate the prevalence of hepatitis B virus (HBV) and hepatitis C virus (HCV) infections among blood donors in different regions in Ethiopia. A total of 56,885 sera were tested for HBsAg and anti-HCV antibodies. Of these, 3.9% were found HBsAg-positive, 0.52% anti-HCV-positive and 0.054% dual positive. HBV prevalence was relatively higher in Adama (5.91%) than Gondar (4.05%), Jimma (3.87%), Addis Ababa (3.75%), and Tigray (3.7%); and in males (4.64%) than females (2.1%). Overall, HBV and HCV prevalence increased with age. In conclusion, HBV and HCV seroprevalence among blood donors in Ethiopia is intermediate and low, respectively.



**Keywords**

Ethiopia; HBV; HCV; Seroepidemiology

## INTRODUCTION

Hepatitis B virus (HBV) and hepatitis C virus (HCV) are a major public health problem in all parts of the world. Worldwide, over 240 million people are infected with HBV and 185 million with HCV [8,75]. HBV and HCV are the leading cause of infectious disease-related morbidity and mortality. According to a recent study by Stanaway *et al.*, the annual number of deaths attributable to viral hepatitis rose by 63% from 1990 to 2013 — from 0.89 million to 1.4 million. Of these, HBV and HCV accounted for 96% of hepatitis-related mortality in 2013 alone, with the highest mortality rates seen in Africa and Asia [6].

Although HBV and HCV infections are a global health problem, there is a distinct geographical variation in their prevalence, with regions having high (≥ 8%), intermediate (2-7%) and low (< 2%) endemicity. Africa is among the highly endemic regions, with an HBV prevalence rate over 8%. Overall, this continent shares 30% of the global HBV burden, with over 75 million chronic carriers [8]. Moreover, Africa has the highest HCV burden with a prevalence of 5.3% and an estimated 32 million chronic carriers [75]. Like most African countries, Ethiopia has been categorized among the highly endemic areas for viral hepatitis. However, due to the lack of a nationwide survey and readily available routine serological tests for patients, the exact prevalence of these viruses in Ethiopia remains unknown. Although a few studies have been performed, these were mainly either restricted to subjects from risk groups, HIV patients, a small sample size [114-117] or done several years ago [118]. Hence, large-scale studies are needed to monitor the up-to-date prevalence of these viruses and to implement appropriate preventive measures. In this multicentric cross-sectional study, we aimed to determine the prevalence of HBV and HCV infections among healthy blood donors to gain insight into the burden of these viruses in the general population of Ethiopia.

## MATERIALS AND METHODS

In a cross-sectional study, a total of 56,885 serum samples was collected from healthy blood donors living in five distinct geographical regions in Ethiopia. Between March 2013 and April 2014, 46,052 samples were collected from Addis Ababa, 3,488 samples from Adama, 2,986 samples from Gondar, 465 samples from Jimma and 3,894 samples from Tigray regions. Serum sample collections were carried out at laboratories in national and regional blood banks. Blood donors who fulfilled the national blood bank criteria to be eligible for donation were included. Those who did not meet the inclusion criteria: age < 18 or >

65 years, history of long-term medication use, family history of liver disease, surgery, repeated blood donation < 1 year and pregnant women were excluded from the study. A blood sample of 10 ml was collected from each donor using a sterile BD vacutainer, and the serum was separated by centrifugation and placed in sterile serum storage vials. All samples were screened for hepatitis B surface antigen (HBsAg) and anti-HCV antibodies by using a fourth-generation Enzyme-linked immunosorbent assay (ELISA) kits (DIALB Diagnostics GmbH, Vienna, Austria), according to the manufacturer's instructions. The study was approved by the National Research Ethics Review Committee at the Federal Ministry of Science and Technology, Ethiopia.

## RESULTS

Of the 56 885 blood donors tested in this study, 40 204 (70%) were males and 16 681 (30%) were females, aged 18-62 years. The prevalence of HBV, HCV and HBV/HCV coinfection is shown in Table 1. Of the 56 885 samples screened, 2,217 were HBsAg positive, giving the overall nationwide HBV prevalence of 3.9%. When we compare each region, HBV prevalence was relatively high in Adama (5.91%) followed by Gondar (4.05%), Jimma (3.87%), Addis Ababa (3.75%) and Tigray (3.7%). With regard to gender, there was a significant difference between males and females in HBV seropositivity. HBsAg prevalence was roughly two times higher in males than females (4.64% vs. 2.1%).

Notably, a very low prevalence of anti-HCV antibodies was found in the present study. The overall prevalence of HCV was 0.52%, corresponding to 294 anti-HCV positive individuals out of 56,885 tested. No difference was observed in anti-HCV positivity between males and females. Regarding the prevalence of HBV/HCV coinfection, the presence of both HBsAg and anti-HCV antibodies were seen in 31 out of 56,885 (0.054%) individuals. Of these, 23 were males and 8 were females.

Age-related prevalence of HBV and HCV was also assessed in this study. As shown in Figure 1 the prevalence of HBV and HCV increases with age with a peak in the age group 40-49 (5.33% for HBV and 0.81% for HCV). This peak is then followed by a gradual decrease in prevalence in the older group > 50 years old (to 4.57% for HBV and to 0.78% for HCV).

**Table 1. Prevalence of HBV and HCV infection and coinfection by region and gender**

| | No. Tested | HBsAg positive N (%) | Anti-HCV-Ab positive[a] N (%) | Coinfection N (%)[b] | Total N (%) |
|---|---|---|---|---|---|
| **Region** | | | | | |
| Addis Ababa | 46052 | 1728 (3.75) | 248 (0.54) | 29 (0.063) | 2005 (4.35) |
| Adama | 3488 | 206 (5.91) | 13 (0.37) | 0 | 219 (6.28) |
| Gondar | 2986 | 121 (4.05) | 15 (0.50) | 1 (0.033) | 137 (4.59) |
| Jimma | 465 | 18 (3.87) | 3 (0.65) | 0 | 21 (4.52) |
| Tigray | 3894 | 144 (3.7) | 15 (0.39) | 1 (0.026) | 160 (4.11) |
| Total | 56885 | 2217 (3.9) | 294 (0.52) | 31 (0.054) | 2542 (4.47) |
| **Gender** | | | | | |
| Male | 40204 | 1866 (4.64) | 196 (0.49) | 23 (0.06) | 2085 (5.19) |
| Female | 16681 | 351 (2.1) | 98 (0.59) | 8 (0.05) | 457 (2.74) |
| Total | 56885 | 2217 (3.9) | 294 (0.52) | 31 (0.054) | 2542 (4.47) |

[b]Positive for both HBsAg and Anti-HCV antibody; [a]Ab, antibody

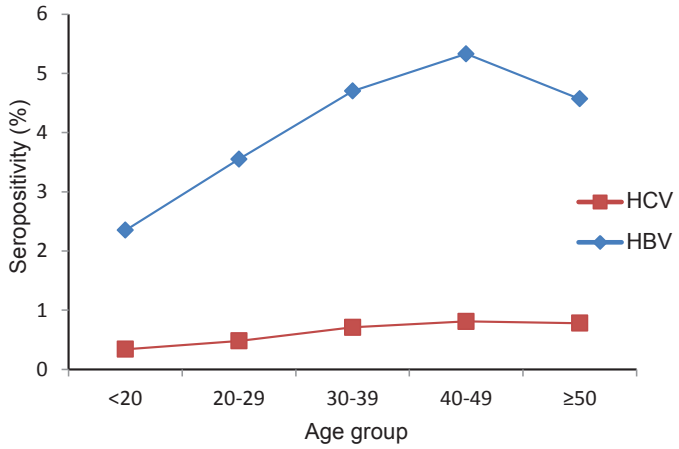



**Figure 1. Age-specific prevalence of HBsAg and anti-HCV antibody among blood donors**

## DISCUSSION

Information on the epidemiology of viral hepatitis in Ethiopia is still lacking. To the best of our knowledge, this is the first large-scale nationwide epidemiological study of HCV and HBV infections from Ethiopia. Herein we report the prevalence of HBV and HCV in different geographical regions— Addis Ababa, Adama, Gondar, Jimma, and Tigray. These regions are approximately 100 to 1,500 km

distant from each other and differ in their socioeconomic, culture and ethnicity. The overall prevalence of HBsAg was 3.9%, and anti-HCV was 0.52%. These prevalence rates classify the country as an intermediate (2–7%) endemic area for HBV infection and low (< 1.5%) for HCV [8,75].

Region wise HBsAg positivity rates ranged from 3.7% in the northern Tigray region to 5.91% the eastern Adama region. The centrally located Addis Ababa city (capital city) had HBsAg prevalence of 3.75%, whereas the southwestern region Jimma had 3.87%. These results are in concurrence with previous studies reported from Addis Ababa and Gondar [114,115,119], but lower than those studies from Jijiga [120], Sekota [117], Wolaita Sodo [116] regions, and a nationwide study (10.8%) [118]. This may be attributable to the difference in the study period, study subjects, and sample size. For instance, the previous nationwide study was conducted three decades ago and the sample size was limited to 5270 young males. Whereas the majority of study subjects at Jijiga blood banks were replacement donors [120] and those studies by Gebreegziabher et al. and Bistetegen et al. were limited to 482 and 390 samples, respectively, included risk groups and suspected patients [117], our findings were drawn from a large sample size consisting of only voluntary donors.

The HBsAg prevalence among blood donors in this study is comparable with reports from neighboring countries; 4.2% in Sudan [121] and 2.58% in Eritrea [122], but lower than other Central and Western African countries [123]. Overall, blood donors showed lower HBsAg prevalence compared to the general population [124-127] might be due to exclusion of risk groups with clinical signs and symptoms of hepatitis in blood donors.

Regarding the HCV infection, the overall prevalence in the present study was 0.52%, which is significantly lower than the previously estimated 2.0% [75]. Interestingly, the prevalence rate was constantly lower than 1% across all regions. Our result is in line with HCV prevalence among blood donors in Southeast African region (0.43%) but lower than Central (2.61%) and West (3.18) Africa regions [128]. It is noteworthy that HBV/HCV coinfection is not uncommon in endemic areas because of the shared modes of transmission. Many studies have reported HBV/HCV coinfection prevalence rates, ranging from 0.7 to 15%, as recently reviewed by Konstantinou and Deutsch [129]. In our study, we found only 31 (0.054%) coinfections out of 56, 885 samples tested, which is significantly lower than the aforementioned studies. This might be due to differences in study subjects as HBV/HCV coinfection is more frequently found in high-risk populations, including injecting drug users, hemodialysis patients, organ transplant patients, HIV-positive and β-thalassemia patients [129], whereas our study subjects are healthy blood donors.

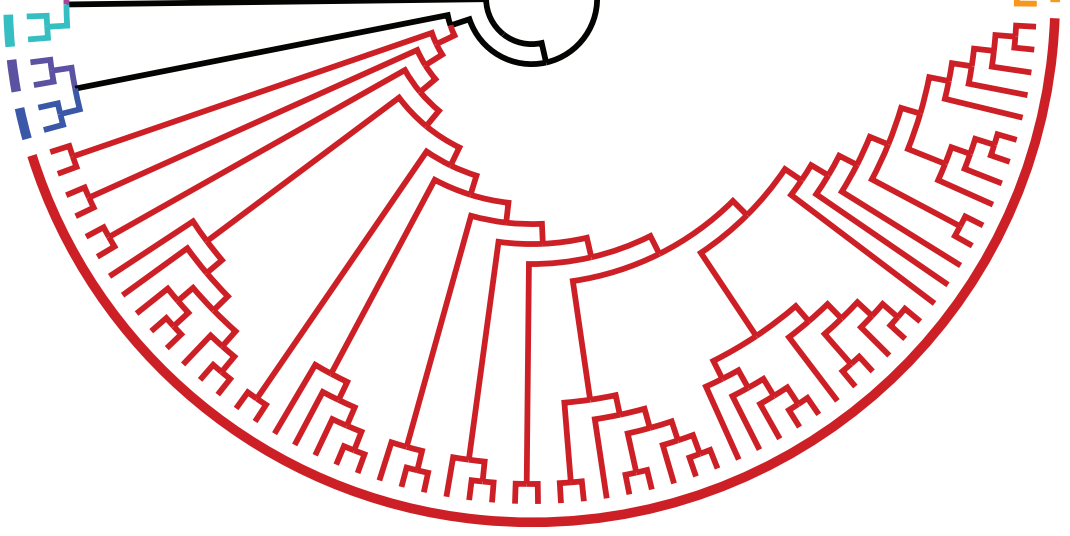
Another important observation in this study was gender-related differences in the prevalence of HBV infection. The prevalence of HBV infection was significantly higher in males than in females. This was comparable with findings reported elsewhere [115,124,125]. This difference might be due to the higher exposure to occupational HBV risk factors in males than females, or because females are more likely than males to produce hepatitis B surface antibody (anti-HBs) in response to HBV infection [130]. Nevertheless, no difference was observed in anti-HCV positivity between males and females.

With regard to age-related analysis, the prevalence of both HBV and HCV infection increased with age from 2.35% to 5.33% for HBV and from 0.34% to 0.81% for HCV. The highest prevalence had been seen among persons 40–49 years of age. Whereas, in the older group over 50 years, the prevalence of both viruses decreased progressively. This could be probably due to deaths caused by chronic hepatitis-related cirrhosis and hepatocellular carcinoma in this age group [131]. Moreover, 25% of people who acquire HBV as children will develop primary liver cancer or cirrhosis as adults. This is the case in sub-Saharan Africa, where transmission predominantly occurs in infants and children by perinatal and horizontal routes [8]. Thus, it is important to note that many people are at high risk of developing viral hepatitis-related chronic liver disease at the later age, due to the fact that antiviral therapy is currently not available in Ethiopia.

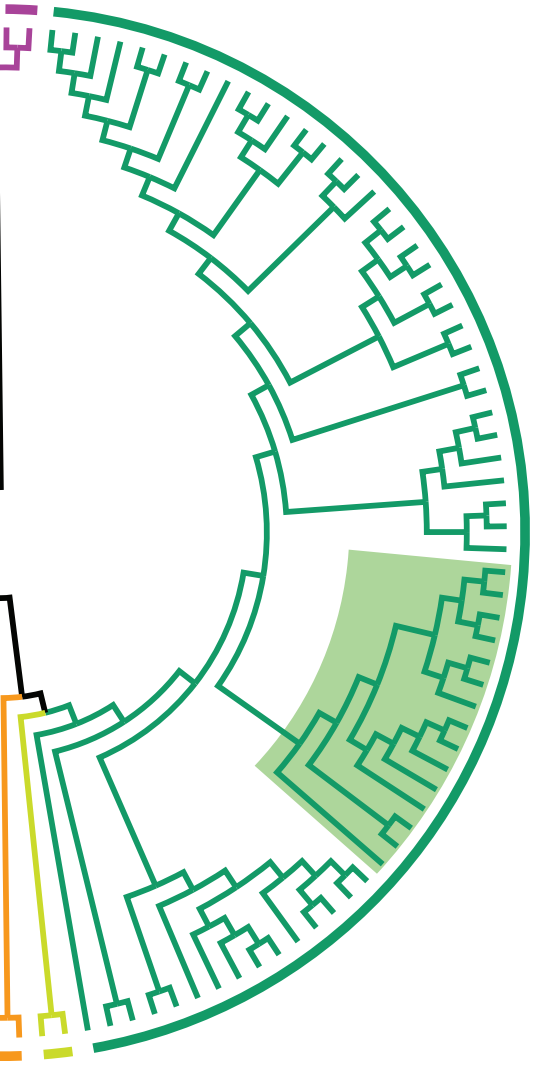
In conclusion, this study shows that the prevalence rate of HBV is intermediate and that of HCV is low in Ethiopia and their prevalence more or less similar across the different regions. Although these findings are useful for making estimates about the overall viral hepatitis burden in Ethiopia, it might underestimate the prevalence in the general population due to exclusion of risk groups with clinical signs and symptoms of hepatitis in blood donors. Thus, we strongly encourage a need for national-level surveillance and public campaigns to raise awareness.

## ACKNOWLEDGEMENTS

We thank our colleagues from the regional and national blood banks of Ethiopia for collecting the samples and data. This work was supported by Nuffic grant, Netherlands Fellowship Programmes (grant no. CF8605).



# Chapter 3



## Molecular Epidemiology and Genetic Diversity of Hepatitis B Virus in Ethiopia

Gadissa Bedada Hundie , V. Stalin Raj, Daniel Gebre Michael, Suzan D. Pas, A.D.M.E. Osterhaus, Marion P. Koopmans, Saskia L. Smits, Bart L. Haagmans

*J Med Virol. 2016 Jun;88(6):1035-1043.*

## ABSTRACT

Although hepatitis B virus (HBV) infection is hyperendemic in Ethiopia and constitutes a major public health problem, little is known about its genetic diversity, genotypes and circulation. The aim of this study was to determine the molecular epidemiology and genetic diversity of HBV in Ethiopia, using 391 serum samples collected from HBsAg-positive blood donors living in five different geographic regions. The HBV S/pol gene was amplified, sequenced, and HBV genotypes, subgenotypes, serotypes and major hydrophilic region (MHR) variants were determined. Phylogenetic analysis of 371 samples (95%) revealed the distribution of genotypes A (78%) and D (22%) in Ethiopia. Further phylogenetic analysis identified one subgenotype (A1) within genotype A, and 4 subgenotypes within genotype D (D1; 1.3%, D2; 55%, D4; 2.5% and D6; 8.8%). Importantly, 24 isolates (30%) of genotype D formed a novel phylogenetic cluster, distinct from any known D subgenotypes, and two A/D recombinants. Analysis of predicted amino-acid sequences within the HBsAg revealed four serotypes: *adw2* (79%), *ayw1* (3.1%), *ayw2* (7.8%), and *ayw3* (11.6%). Subsequent examination of sequences showed that 51 HBV isolates (14%) had mutations in the MHR and 8 isolates (2.2%) in the reverse transcriptase known to confer antiviral resistance. This study provides the first description of HBV genetic diversity in Ethiopia with a predominance of subgenotypes A1 and D2, and also identified HBV isolates that could represent a novel subgenotype. Furthermore, a significant prevalence of HBsAg variants in Ethiopian population is revealed.



**Keywords**

HBV; Genotype; Epidemiology; Ethiopia; Mutation; Genetic diversity

## INTRODUCTION

Despite the availability of vaccine and antiviral treatments, hepatitis B virus (HBV) infection remains a major global health problem, causing substantial mortality and morbidity. The World Health Organization (WHO) estimates that there are 240 million chronic HBV carriers worldwide [132], more than 65 million of them living in Africa [133]. Ethiopia, a sub-Saharan African country with a population of ~95 million, is classified by the WHO as one of the African countries with the highest rate of HBV infection [134]. It is a geographically diverse nation, home to over eighty different ethnic groups, and bordered by six different countries. Over 70% of the population shows evidence of past infection with HBV and 7–14% are chronic carriers [115,135]. Even though a safe and effective vaccine against HBV has been available since 1982, Ethiopia implemented the HBV vaccine into its national immunization program only recently, in 2007. In general, hepatitis B is one of the most neglected infectious diseases in Ethiopia, and overall awareness of the population about HBV is very low. Because of this late implementation, lack of awareness, and the absence of a national HBV control program, the burden of hepatitis B likely remains substantial and molecular data on HBV genetic diversity is lacking.

HBV, a member of the family *Hepadnaviridae*, is a circular, partially double-stranded DNA virus with a genome size of 3.2 kb [27]. HBV shows high genetic diversity and has been classified into ten genotypes (A-J) [30,35,36,136], forty subgenotypes [20,27,30], and nine serotypes (*adw2, adw4, ayw1, ayw2, ayw3, ayw4, adrq+, adrq– and ayr*) [29,30]. HBV genotypes and subgenotypes differ in clinical outcomes and response to antiviral therapy [54,58] and have distinct geographic distributions [27,52]. Some genotypes and subgenotypes are restricted to particular geographic regions while others are distributed globally.

Besides genotype diversity, HBV genomic variations in the hepatitis B surface antigen (HBsAg) which are selected during the infection course are of clinical and public health importance. HBsAg contains a central region called the Major Hydrophilic Region (MHR; aa 100 to 170) [137]. MHR is highly immunogenic and is one of the main targets for neutralizing antibodies. Mutations within MHR have been associated with HBsAg detection failure, vaccine and immunotherapy escape [138,139]. The epidemiology of these mutants has been studied in many HBV endemic countries [140], however, not in Ethiopia.

## OBJECTIVES

Information about HBV genetic diversity is essential not only for epidemiological reasons but also to understand the routes of HBV transmission and to establish efficient strategies for preventing and controlling HBV infection. Here we aimed to study the distribution of HBV genotypes, subgenotypes, serotypes and HBsAg variants in Ethiopia, a hyper-endemic HBV region where HBV molecular data is lacking, to obtain more insight into the molecular epidemiology of HBV.

## MATERIALS AND METHODS

### Study area and sample collection

A total of 391 serum samples were collected from HBsAg-positive blood donors at blood bank centers in five geographically distinct regions of Ethiopia: Addis Ababa, (central), Adama (central-east), Gondar (north-west), Mekelle (north), and Jimma (south-west). Serum samples were collected between March 2013 and April 2014, and stored at −80°C until use. The study was approved by the National Research Ethics Review Committee at the Federal Ministry of Science and Technology, Ethiopia.

### Serological Assays

Serological markers for HBsAg and anti-HCV antibodies were tested using commercially available ELISA kits (DIALB Diagnostics GmbH, Vienna, Austria). Blood was also screened for anti-HIV-1 and -HIV-2 antibodies by ELISA kit (Vironostika HIV Ag/Ab, Bio-Merieux, Boxtel, The Netherlands) and anti-*Treponema Pallidum* antibodies by Rapid Plasma Reagin Test (RPR) (DIALB Diagnostics GmbH, Vienna, Austria). All serological tests were carried out twice according to the manufacturer's instructions. HBsAg-positive serum samples from patients co-infected either with HCV or HIV were excluded from the present HBV study.

### HBV DNA extraction, quantification, amplification and sequencing

Viral DNA was extracted from 200 µl serum using the QIAamp MinElute virus spin kit (Qiagen, Hilden, Germany), according to the manufacturer's instructions. Serum HBV DNA levels were quantified using an in-house quantitative real-time PCR (qPCR) as described previously [141,142]. AcroMetrix® HBV low, medium and high controls (Applied Biosystems, life technology, USA) were used as standard reference. The HBV S-gene (nucleotides 56 to 805) was amplified by one-round

PCR for samples with viral loads >3.5 log ($3 \times 10^3$ IU/ml) and by a nested PCR for those with viral loads ≤ 3.5 log with primers S01F (5'- CCTGCTGGTGGCTC-CAGTTC-3'), S05R (5'-TGGTAACAGCGGTATAAAGGGACT-3'), S03F(5'-GCTGGTGGCTCCAGTTCAGGAACA-3') and S07R (5'-AAGGGACTCAA-GATGYTGTACAGA- 3'). Amplicons were sequenced bi-directionally using Big Dye Terminator v3.1 kits and an ABI PRISM 3130xl sequencer (Applied Biosystems).



**Phylogenetic analysis and HBV typing**

Sequence alignments were created with ClustalW implemented in MEGA version 6.0 [143]. Phylogenetic trees were constructed using PhyML v 3.0 [http:// www.atgc-montpellier.fr/phyml/]. Briefly, a maximum likelihood (ML) phylogenetic tree was inferred using the Hasegawa-Kishino-Yano (HKY85) nucleotide substitution model with gamma distributed rates with invariant sites (G+I). A tree searching operation was performed using the best of nearest neighbor interchange (NNI) and subtree pruning and regrafting (SPR) branch-swapping algorithm. The statistical robustness and reliability of the branching order within each phylogenetic tree were confirmed with a bootstrap analysis (1000 replicates). The obtained tree was visualized in FigTree v1.4.2 [ http://tree.bio.ed.ac.uk/software]. The tree was rooted using a midpoint rooting. HBV recombinant analysis was performed by jumping profile Hidden Markov Model (jpHMM) [144]. HBV serotypes were determined as described previously [30,145].

**HBV mutations analysis in the S and RT regions**

The analysis of the amino acid substitutions within the MHR (aa 99 to 170) was performed by comparison of the 371 sequences in this work with the amino-acid sequence of similar genotypes retrieved from GenBank (AY934770 and AB188243). Distinction between naturally occurring variants and mutations within each genotype was made based on published data. The corresponding substitutions in the reverse transcriptase (RT) were determined according to the Stanford database (http://hivdb.stanford.edu/).

**Nucleotide sequence accession number**

All sequences reported in this study were submitted to GenBank (accession numbers KP310929 to KP311299).

**Statistical analysis**

Categorical data were analyzed using the chi-square ($\chi 2$) test, and or Fisher's exact 2-sided test, if appropriate. A $p$-value ≤ 0.05 was considered significant.

## RESULTS

### Geographic distribution and diversity of HBV genotypes

Serum samples ($n$ = 391) collected from HBsAg-positive blood donors (344 male and 47 female; mean age; 30.8 years, range, 18 to 60 years) were analyzed. HBV DNA levels were quantified in 383 (98%) samples, and the median serum viral load was 785 IU/ml (range, 7 IU/ml to $3\times10^8$ IU/ml). The remaining 8 (2%) samples showed no detectable viral DNA. The majority of samples (78%) had viral load below $10^4$ IU/ml (Figure 1). Samples ($n$ = 12) that were low positive by qPCR (< 40 IU/ml) were not confirmed by nested PCR.

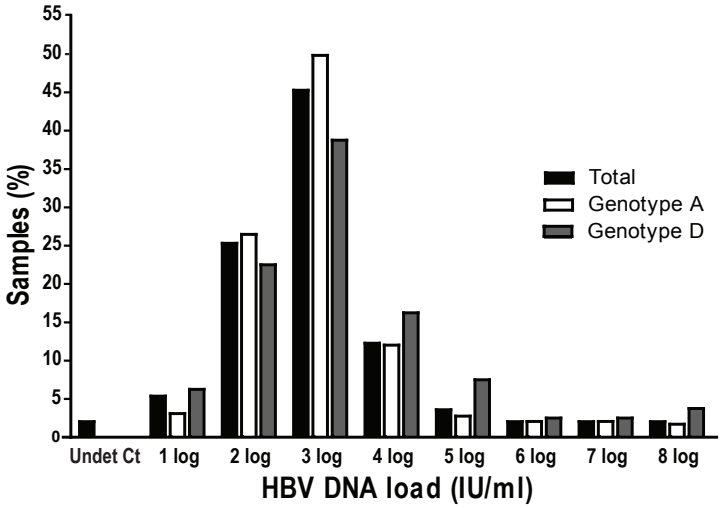



**Figure 1. HBV viral load distribution of 391 Ethiopian HBsAg-positive blood donor samples.** Shown are the total population of donors (black bars), donors infected with HBV genotype A (white bars), and donors infected with HBV genotype D (gray bars). Undet, undetectable.

Three hundred and seventy-one strains were successfully amplified and sequenced. All 371 HBV isolates were classified into two genotypes: A ($n$ = 291; 78%) and D ($n$ = 80; 22%). HBV genotype distribution in Ethiopia differed by geographic region (Figure 2). Genotype A prevails in all regions, but was predominant in Addis Ababa (85%), Adama (91%) and Jimma (93%), while genotype D is more prevalent in Mekelle (40%) and Gondar (34%) area. Of note, genotype D prevalence was significantly higher in the two northern regions as compared to the other three regions (odds ratio (OR) 3.46, 95% confidence interval (CI) 2.07-5.81, $P$ < 0.001). No significant differences were observed between genotypes with respect to donor's age and gender (Figure S1 A and B).

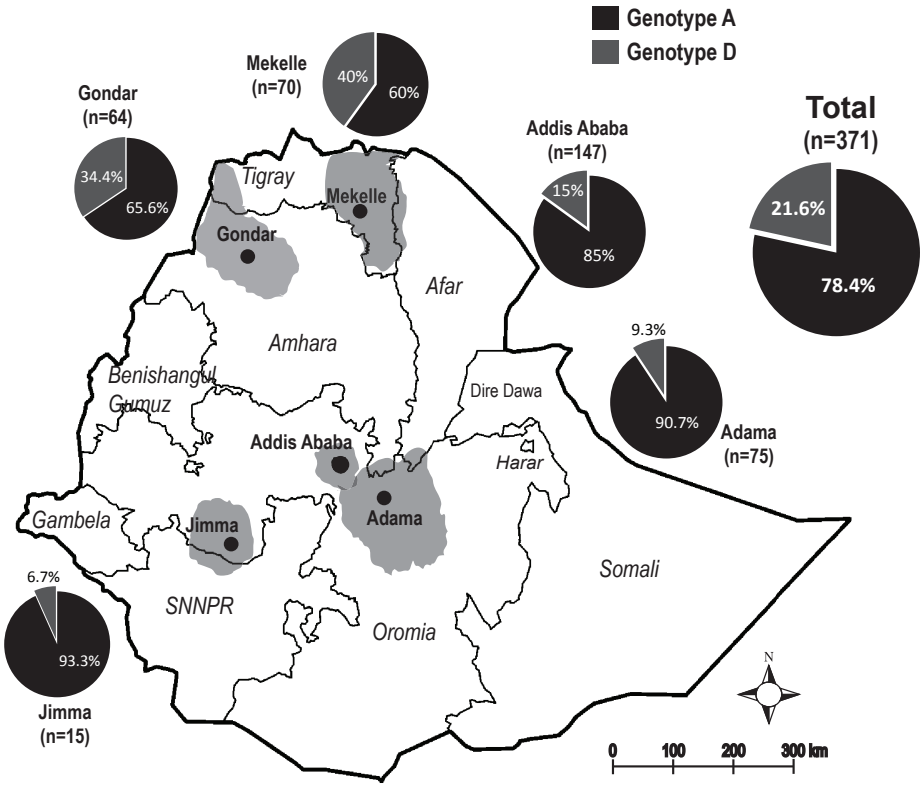
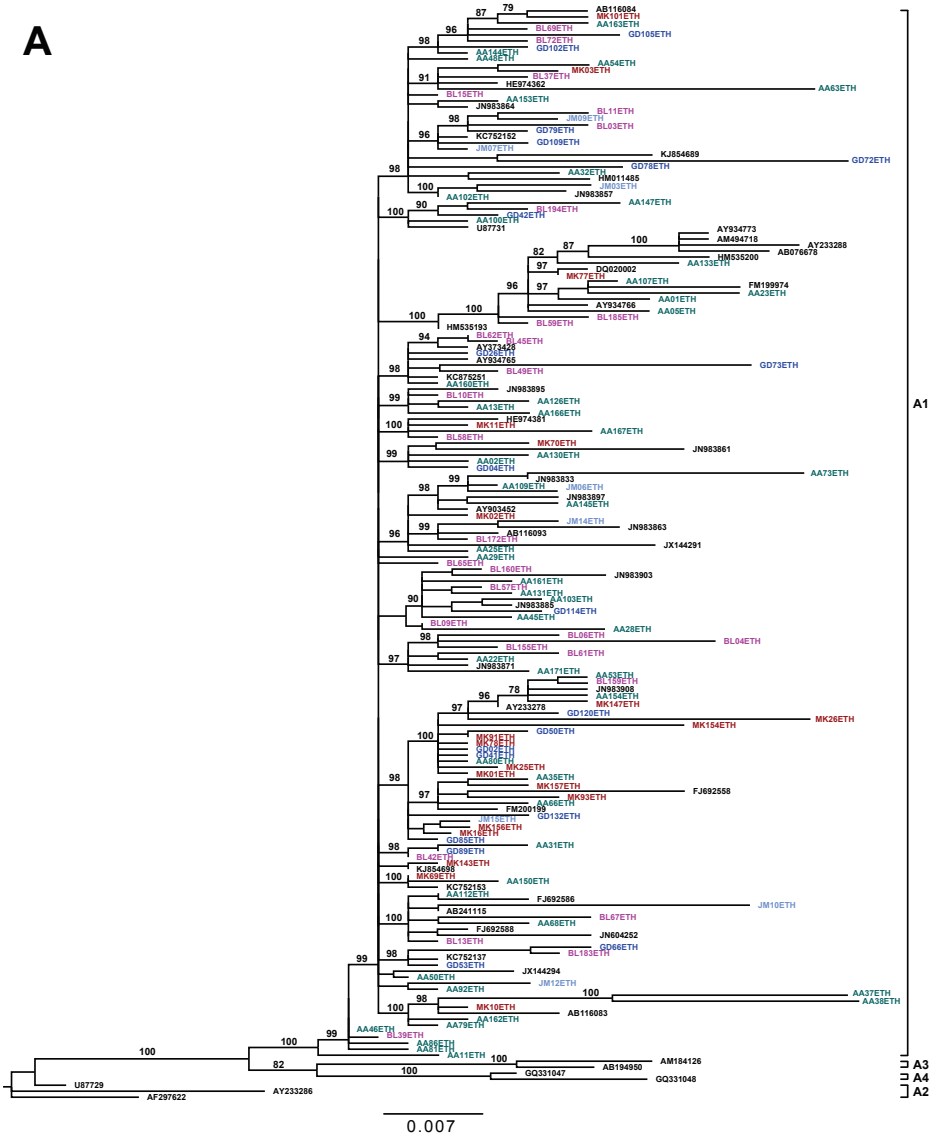



**Figure 2. Geographic distribution of HBV genotypes in five regions of Ethiopia.** Total number of samples analyzed per region are indicated in brackets and percentages of corresponding genotypes in pie chart. Small black dot indicate locations of blood bank centers, and the gray area shows approximate area coverage of sample collection (70-400 km).

Phylogenetic analysis classified all genotype A isolates into subgenotype A1 (Figure S2). Figure 3A shows 130 representatives of subgenotype A1 Ethiopian isolates selected from different clusters and the five geographic regions based on the Figure S2 analysis. A1 isolates from each of the study regions clustered into different HBV/A1 branches suggesting that more than one strain is circulating in each region (Figure 3A). Genotype D isolates split into four phylogenetic groups: D1 ($n$ = 1), D2 ($n$ = 44), D4 ($n$ = 2), D6 ($n$ = 7), (Figure 3B). Interestingly, the remaining 24 (30%) isolates formed a novel unclassified cluster (Figure 3B), while two outliers (BL66ETH & MK141ETH) were D/A recombinants (data not shown). The majority of Ethiopian genotype D belong to subgenotype D2 (55%).


A
AB116084
MK101ETH
AA163ETH
BL69ETH
GD105ETH
BL72ETH
GD102ETH
AA144ETH
AA48ETH
AA54ETH
MK03ETH
BL37ETH
HE974362
AA63ETH
BL15ETH
AA183ETH
JN983864
BL11ETH
JM09ETH
BL03ETH
GD79ETH
KC762152
GD109ETH
JM07ETH
KJ854689
GD72ETH
GD78ETH
AA32ETH
HM011485
JM03ETH
AA102ETH
JN983857
AA147ETH
BL194ETH
GD42ETH
AA100ETH
U87731
AY934773
AM494718
AY233288
AB076678
HM535200
AA133ETH
DQ020002
MK77ETH
AA107ETH
FM199974
AA23ETH
AA01ETH
AY934766
AA05ETH
BL185ETH
BL59ETH
HM535193
BL62ETH
BL45ETH
AY373426
GD39ETH
AY934765
GD73ETH
BL49ETH
KC875251
AA160ETH
JN983895
BL10ETH
AA126ETH
AA13ETH
AA166ETH
HE974381
MK11ETH
AA167ETH
BL58ETH
MK70ETH
JN983861
AA130ETH
AA02ETH
GD04ETH
AA73ETH
JN983833
AA109ETH
JM06ETH
JN983897
AA145ETH
AY903452
MK02ETH
JM14ETH
JN983863
AB116093
BL172ETH
JX144291
AA25ETH
AA29ETH
BL65ETH
BL160ETH
JN983903
AA161ETH
BL57ETH
AA131ETH
AA103ETH
JN983885
GD114ETH
AA45ETH
BL99ETH
AA28ETH
BL06ETH
BL04ETH
BL155ETH
BL61ETH
AA32ETH
JN983871
AA171ETH
AA53ETH
BL159ETH
JN983908
AA154ETH
MK147ETH
AY233278
GD120ETH
MK26ETH
MK154ETH
GD50ETH
MK91ETH
MK94ETH
MK95ETH
GD37ETH
AA80ETH
MK25ETH
MK01ETH
AA35ETH
MK157ETH
FJ692558
MK93ETH
AA66ETH
FM200199
GD132ETH
JM15ETH
MK156ETH
MK10ETH
GD85ETH
AA31ETH
GD89ETH
BL42ETH
MK143ETH
KJ854698
MK69ETH
AA150ETH
KC762153
AA112ETH
FJ692586
JM10ETH
AB241115
BL67ETH
AA68ETH
FJ692588
JN604252
BL13ETH
GD66ETH
BL183ETH
KC762137
GD53ETH
JX144294
AA50ETH
JM12ETH
AA92ETH
AA37ETH
AA38ETH
MK10ETH
AB116083
AA162ETH
AA79ETH
AA46ETH
BL39ETH
AA86ETH
AA81ETH
AA11ETH
AM184126
AB194950
GQ331047
GQ331048
U87729
AY233286
AF297622
A1
A3
A4
A2
0.007

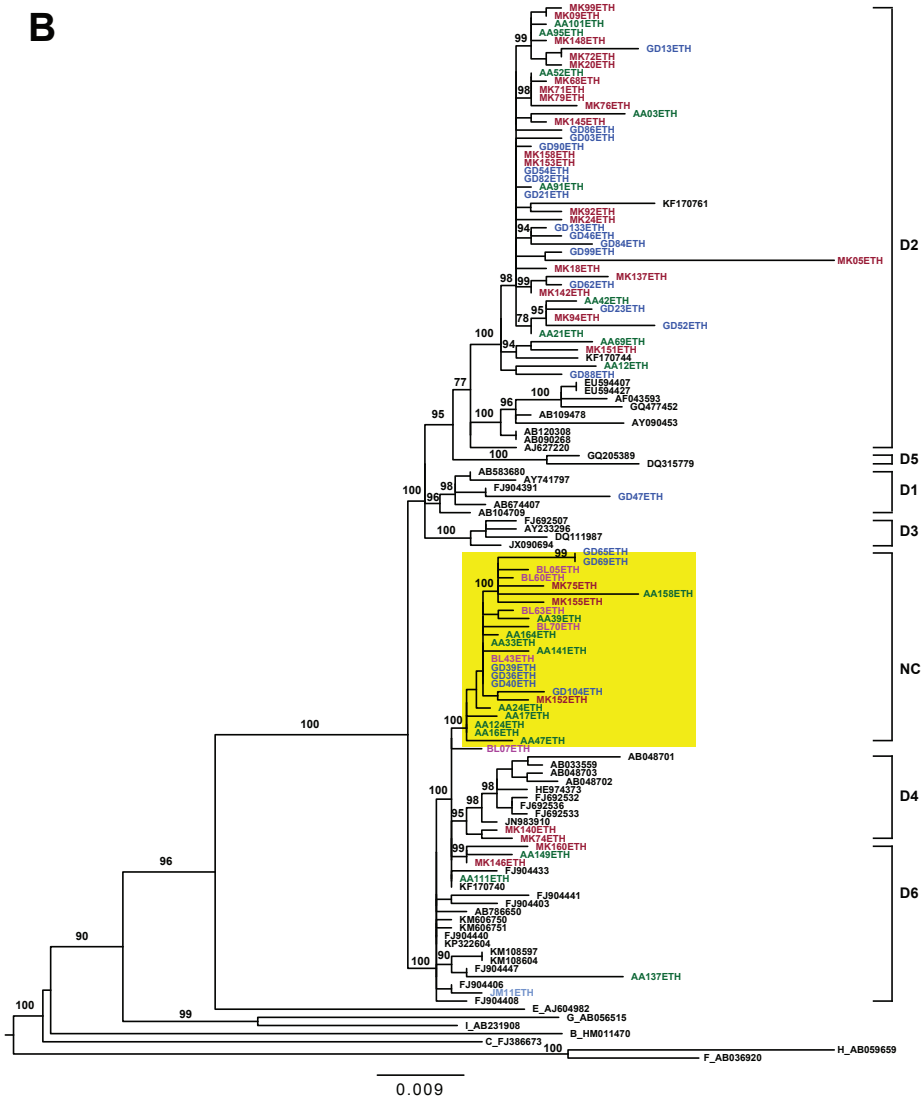



**Figure 3. Maximum Likelihood phylogenetic tree of HBV Ethiopian isolates belonging to (A) genotype A, 130 isolates and (B) genotype D, 78 isolates.** Phylogenetic analysis inferred from Maximum Likelihood tree (HKY85) model) based on HBV S/pol gene sequences (700 nt) of Ethiopian HBV isolates and reference sequences retrieved from the GenBank database representing HBV genotypes (A-I) and all genotype A and D subgenotypes (A1-A4, and D1-D6). Ethiopian HBV isolates are shown in color (color corresponding to their geographic origin: Green = Addis Ababa; Pink = Adama; Red = Mekelle; Blue = Gondar; Light blue = Jimma), and reference sequences retrieved from GenBank shown by their accession number. Numbers in the tree represent the percentages of bootstrap values (1000 replicates). The novel subgenotype D cluster (NC) is highlighted in yellow.

## HBV serotypes

HBV serotype distribution in the studied population was as follows: *adw2* (293/371; 79%), *ayw3* (43/371; 11.6%), *ayw2* (33/371; 8.9%), and *ayw1* (2/371; 0.5%). All genotype A isolates classified into serotype *adw2*. For genotype D, 43 (53.8%) isolates belonged to *ayw3*, 33 (41.2%) were *ayw2*, and two each belonged to *ayw1* and *adw2* (found among D/A recombinant isolates BL66ETH and MK141ETH) (Table 1). Because of the association between genotype and serotype (*adw2* versus other serotypes, OR 142.86, 95% CI 37.03-500, $P < 0.001$), the distribution of serotypes varied according to geographic location, like the genotypes (Table S1, OR 3.5, 95% CI 2.08-5.90, $P < 0.001$). The distribution of serotypes was not significantly different among age groups or gender (Figure S1 C & D).

**Table 1. Relationship between HBV genotype and subtypes**

| Genotype | Number of samples | Subtype, Number (%) | | | |
|---|---|---|---|---|---|
| | | *adw2* | *ayw1* | *ayw2* | *ayw3* |
| A | 291 | 291 (100) | 0 | 0 | 0 |
| D | 80 | 2 (2.5) | 2 (2.5) | 33 (41.2) | 43 (53.8) |
| **Total** | **371** | **293 (79)** | **2 (0.5)** | **33 (8.9)** | **43 (11.6)** |

## HBV mutation in HBV S and RT regions

Several mutations were detected in the MHR of HBsAg (Figure 4). Fifty-one out of 371 (13.8%) samples were found to have mutation (Table 2). Twenty-six isolates (7%) had mutations that have potential impact on the detection of HBsAg: M103I, L109P, P120S, T126I, A128V, Q129R/H, M133T/I, F134L, A159V, E164G. Mutations associated with vaccine or immunotherapy escape, P120S, T126I, Q129H/R, N/T131S/I, M133T/I, Y134F, T140I, T143M, and D144E, were detected in 20 (5.4%) isolates (Figure 4). Furthermore, one amino-acid deletion in isolate JM10ETH (Figure 4a) and a 7 amino-acid deletion in isolate AA12ETH (Figure 4b) was detected. Interestingly, we observed a stop codon W172* in GD88ETH , which was shown to be associated with low level secretion of HBsAg [146]. Distribution of MHR mutations by geo-demographic and genotype/serotype are presented in Table 2. MHR mutations were more prevalent in genotype D than genotype A (OR 2.27, 95% CI 1.20-4.30, $P = 0.016$).

Sequence analysis of the RT region revealed 8 (2.2%) isolates had mutations (Table 2). The most prevalent resistant mutations were L180M (1.35%) and M204V (1.35%). Five isolates had a combination of L180M + M240V mutation, one isolate had I169T mutation, and the remaining two isolates had T184A and A194T mutation.

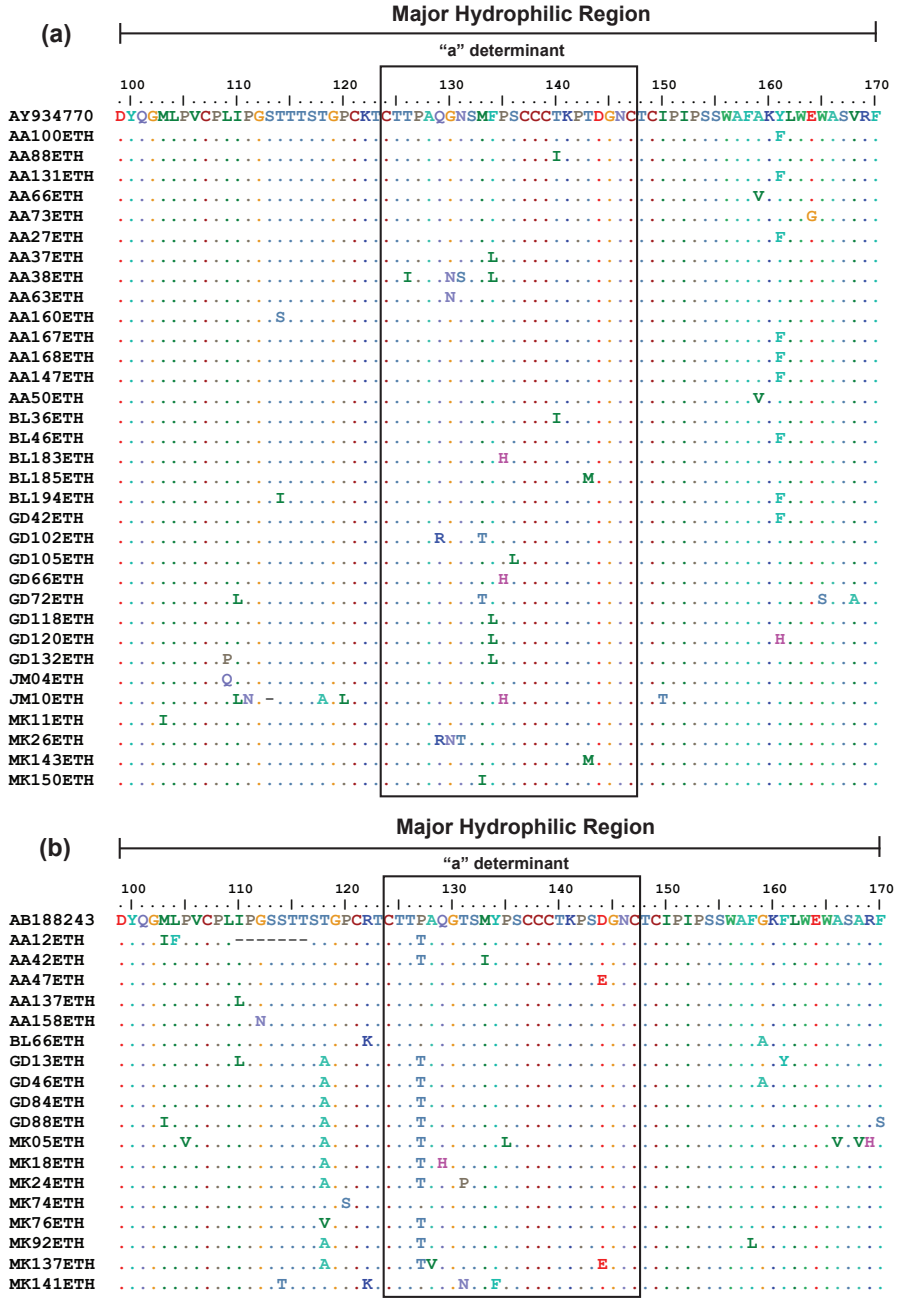



**Figure 4. Distribution of amino acid mutations detected within the major hydrophilic region (99–170) of the HBsAg among Ethiopian HBV isolates.** The amino acid variations in the "a determinant" (124-147) region are boxed (solid line). (**a**) MHR amino-acid sequence alignment of genotype A, (**b**) MHR amino-acid sequence alignment of genotype D.

**Table 2. Distribution of MHR mutations in the HBV S gene and RT mutations by geographic region, demographic variables, genotype, and serotype**

| | Number of samples | Number of samples with MHR mutations | MHR mutation (%) | "a" determinant mutation (%) | Number of Samples with RT mutation |
|---|---|---|---|---|---|
| **Total** | 371 | 51 | 13.8 | 6.8 | 8 |
| **Region** | | | | | |
| Addis Ababa | 147 | 19 | 12.93 | 4.1 | |
| Adama | 75 | 6 | 8 | 4 | |
| Gondar | 64 | 12 | 18.75 | 10.94 | 7 |
| Mekelle | 70 | 12 | 17.14 | 11.43 | 1 |
| Jimma | 15 | 2 | 13.33 | 6.67 | |
| **Gender** | | | | | |
| Male | 325 | 43 | 13.23 | 6.46 | 7 |
| Female | 46 | 8 | 17.39 | 8.70 | 1 |
| **Age** | | | | | |
| < 20 | 13 | 2 | 15.38 | 7.69 | |
| 20-29 | 168 | 22 | 13.1 | 6.55 | 4 |
| 30-39 | 131 | 21 | 16.03 | 7.63 | 2 |
| 40-49 | 45 | 3 | 6.67 | 4.44 | |
| ≥ 50 | 14 | 3 | 21.43 | 7.14 | 2 |
| **Genotype** | | | | | |
| A | 291 | 33 | 11.34 | 6.19 | 7 |
| D | 80 | 18 | 22.5 | 8.75 | 1 |
| **Subtype** | | | | | |
| adw | 293 | 35 | 11.95 | 6.48 | 7 |
| ayw | 78 | 16 | 20.51 | 7.69 | 1 |

## DISCUSSION

Increasing evidence shows that HBV genotypes and subgenotypes display distinct geographic distributions and impact clinical and antiviral therapy outcomes [35,52,58,133,147,148]. Moreover, HBV genotypes and serotypes are useful tools in epidemiological and transmission studies [27]. To date, HBV genetic diversity and genotype distribution pattern in Ethiopia has not been investigated. In this comprehensive and nationwide study we therefore determined the HBV molecular epidemiology and genetic diversity in Ethiopia for the first time. Subsequently, we also assessed the prevalence of mutation in the major hydrophilic region

(MHR) and reverse transcriptase (RT) regions. The study subjects comprise a geographically, socially, economically and ethnically diverse population.

Phylogenetic analysis showed presence of HBV genotypes A (78%) and D (22%) in the Ethiopian population, with genotype A being predominant. This is consistent with previous studies of HBV genotype distribution in neighboring countries and other regions of Africa [30,50,133,149-151]. Although genotype E has been reported in some east-African countries we did not find genotype E in Ethiopia. Most importantly, HBV genotype distribution in Ethiopia showed a distinctive epidemiological pattern. Genotype D was more prevalent in the Northern region of the country (37%) compared to the Central region (12%). In contrast, genotype A predominantly prevails in the Central region. As reported in neighboring countries, genotype A is the dominant strain in Somalia, Kenya and Uganda [133,149], whereas genotype D is the most prevalent strain in Sudan [151]. Our findings shows that Ethiopia is at the border of the genotype A and D geographical distribution in Africa.

HBV genotype A is highly heterogeneous and has been classified into four subgenotypes, A1, A2, A4 and quasi-subgenotype A3 [152,153]. Subgenotypes A1, A4, and quasi-subgenotype A3 are found mainly in Africa, whereas A2 prevails in Europe and North America [154]. All genotype A isolates from Ethiopia belonged to subgenotype A1 similar as in other eastern and southern Africa countries [30,133,150,155,156]. Two distinct subgenotype A1 clades have been recognized: an African and an Asian-American clade [154]. We found Ethiopian A1 isolates to be the most divergent: they cluster in both the African and the Asian-American clades (Figure 3A).

Genotype D was the second most dominant strain in Ethiopia, and its prevalence was higher in northern parts of the country. This genotype has been found universally, but predominates in Europe, North Africa, and Middle East [20,56]. Genotype D isolates from Ethiopia showed a genotypic heterogeneity and discriminated into four subgenotypes, D1, D2, D4 and D6 [56]. Surprisingly, 55% of genotype D isolates belonged to subgenotype D2, which is most prevalent in East Europe including Russia and the Baltic States [157], but less common in Africa. Nevertheless, Ethiopian D2 isolates clustered separately and more distantly to D2 of European and Asian origin. Subgenotype D6 is second most dominant in our population. It has been suggested that the putative geographic area harboring D6 covers North Africa and as far south as the Central African Republic, which borders the sub-Saharan area known to be endemic for HBV/E [39]. However, the relatively high prevalence of D6 in Ethiopia showed that the geographic area of this strain extends further than proposed—as far as eastern Africa. Notably, D1 the predominant subgenotype in North Africa and Middle East [50,56,133,151], is rare

and found only in one isolate. Interestingly, about 30% of genotype D isolates in this study clustered in a novel monophyletic phylogenetic branch with a high statistical bootstrap support (100%) that is distinct from any known subgenotype. Further studies at complete genome level is needed to determine whether the currently unclassifiable isolates belong to a novel subgenotype [53,158]. Another interesting observation in this study was the presence of recombinant genotypes (A/D) and the rare subgenotype D4 in two isolates. The two HBV recombinant isolates were identified by jumping profile Hidden Markov Model [144]. Relative to position on HBV reference genome AM282986 numbering, isolate BL66ETH had sequence of genotype A from nucleotides 75-288 and of genotype D from nucleotides 289-772, whereas isolate MK141ETH had sequence of genotype A from nucleotides 472-573 and of genotype D from nucleotides 75-471 and 574-772.

Importantly, the present study was also the first to describe the HBV serotypes prevalent in Ethiopia. Serotype *adw2* was the most predominant (79%) and found in all subgenotype A1 isolates and in the D/A recombinant isolates (BL66ETH and MK141ETH). Serotype *ayw2* was found in all isolates belonging to subgenotypes D1, D4, D6 and in most of the unclassifiable isolates. All D2 isolates were *ayw3* apart from one isolate specifying *ayw2*. This result is consistent with previous data reporting the relationship between serotype and genotypes/subgenotypes [30,37]. Notably, we also noted that the unclassifiable isolates were not only phylogenetically closely related to subgenotypes D4 and D6, but also that they encoded similar serotype *ayw2*.

In this study, the HBsAg variants were analyzed and found that 51 (13.8%) out of 371 isolates had mutations within the major hydrophilic region. Studies have shown that MHR mutants are associated with failure of HBsAg detection, vaccine and immunotherapy escape [66,67,138,159]. In this study, MHR variants known to be involved in diagnostic failures were found in 7% of isolates although none of them had failed to be detected by the commercial kit utilized. Moreover, variants associated with immunotherapy or vaccine escape, P120S [159], T126I [160], Q129H [161-163], Q129R [163], G130N [164], T131I [163], M133I [165], M133T [165,166], F134L [166], Y134F, T140I, D144E [164], were detected among 20 isolates in this study. The natural occurrence of such variants in vaccine and treatment naïve population is possible as reported by several studies [164,166]. However, a high presence of immune-escape mutants in this study should be considered in the immunization program as they have the potential to spread undetected and thereby affect the effectiveness of the national vaccination program and treatment.

Mutations related to antiviral therapy resistance were identified mainly in samples from Gondar region (7/64, 11%). The best-described drug resistance

mutation— the lamivudine resistance L180M and M204V— were both detected in five isolates [167,168]. Other antiviral resistance mutations, I169T, T184A and A194T, were also detected in three isolates. Importantly, all mutants were detected among therapy-naive subjects. Since there is no HBV therapy in Ethiopia, the current findings might be naturally occurring mutants or therapy-failure mutant strains introduced from elsewhere.

In conclusion, this study is the first to establish the distribution of HBV genotypes, guessed subgenotypes and serotypes in Ethiopia. The study shows the prevalence of two genotypes and various subgenotypes, with a predominance of subgenotypes/serotypes A1/*adw2* and D2/*ayw3*. Importantly, a cluster of HBV isolates which could be represent a novel HBV subgenotype and also detected recombinants strains that should be further investigated by applying the full length genomes of identified strains. Moreover, a significant level of HBsAg variants and RT mutants in vaccine and treatment naïve population were detected. Our findings will therefore have major impact on the design of HBV prevention, treatment, and diagnosis in Ethiopia.



**Acknowledgements**

We thank Claudia Mulders for excellent technical assistance and our colleagues from the Ethiopian Blood Bank Centers for collecting the samples and data. This work was supported by Nuffic grant from the Netherlands Fellowship Programmes (grant no. CF8605).

## SUPPORTING INFORMATION

**Table S1. Distribution of HBV subtypes in the five geographic regions of Ethiopia**

| Region | Subtype, Number (%) | | | | | | **Total** |
|---|---|---|---|---|---|---|---|
| | Subtotal | | | Subtotal | | | |
| | adw | adw2 | ayw | ayw1 | ayw2 | ayw3 | |
| Addis Ababa | 125 (85.0) | 125 (85.0) | 22 (15.0) | 0 | 14 (9.5) | 8 (5.4) | **147** |
| Adama | 69 (92.0) | 69 (92.0) | 6 (8.0) | 0 | 6 (8.0) | 0 | **75** |
| Gondar | 42 (65.6) | 42 (65.6) | 22 (34.4) | 2 (3.1) | 5 (7.8) | 15 (23.4) | **64** |
| Jimma | 14 (93.3) | 14 (93.3) | 1 (6.7) | 0 | 1 (6.7) | 0 | **15** |
| Mekelle | 43 (61.4) | 43 (61.4) | 27 (38.6) | 0 | 7 (10.0) | 20 (28.6) | **70** |
| **Total** | **293 (79.0)** | **293 (79.0)** | **78 (21.0)** | **2 (0.5)** | **35 (9.4)** | **43 (11.6)** | **371** |

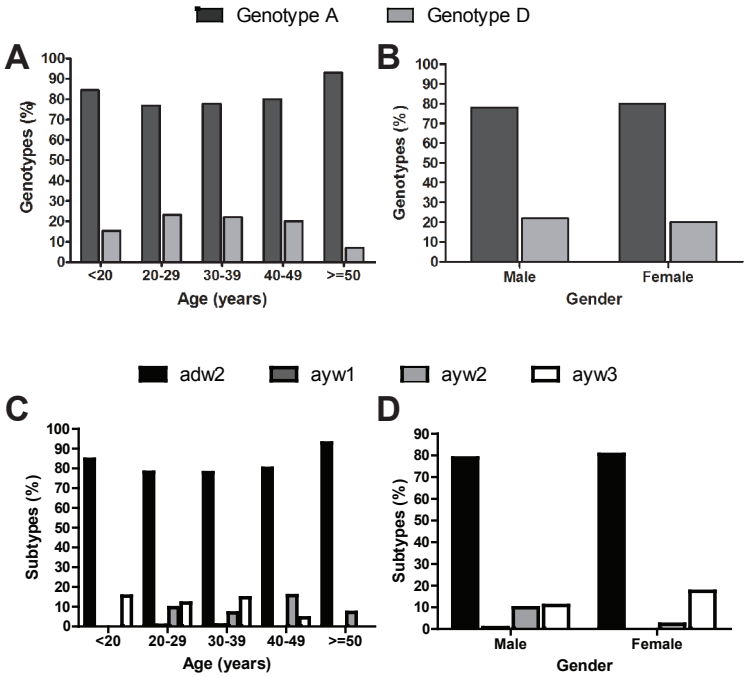
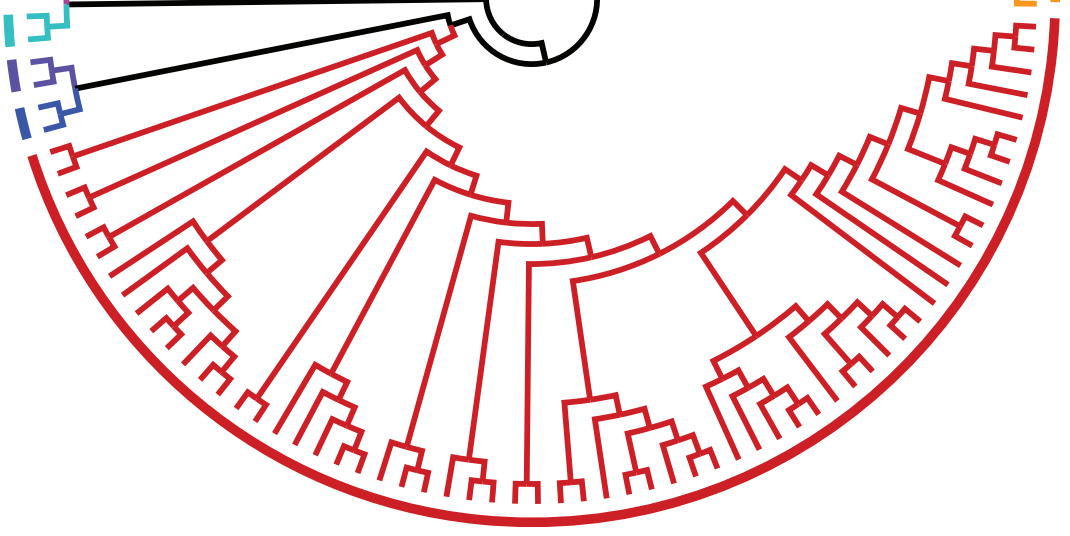
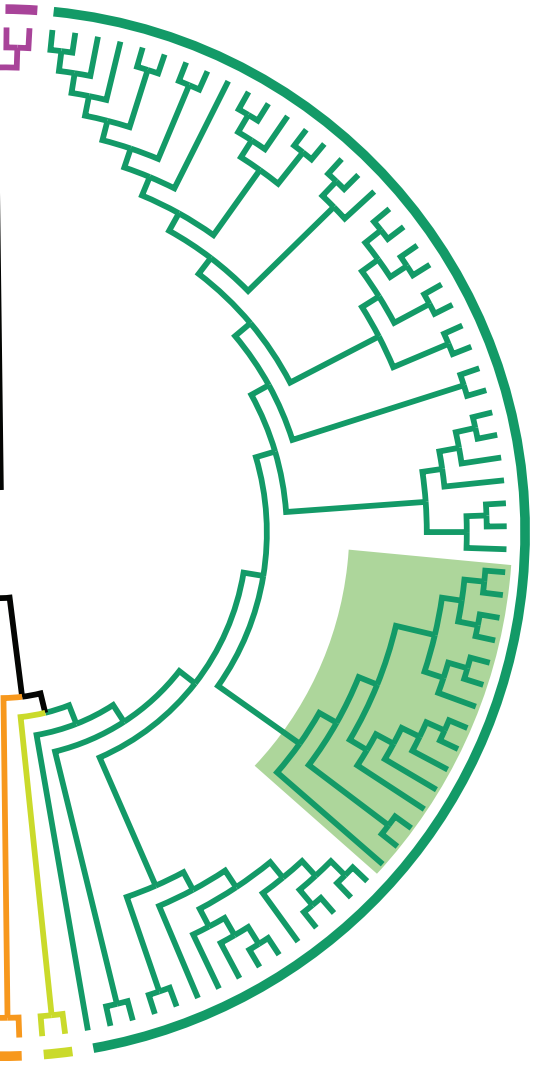



**Figure S1. Distribution of HBV genotypes and subtypes by age and gender.** HBV genotypes (A, B) and subtypes (C, D) were determined and distributed according age (A, C) or gender (B, D).


**Figure S2. Maximum Likelihood phylogenetic tree of 371 Ethiopian isolates and 110 HBV reference sequences.** Phylogenetic analysis inferred from Maximum Likelihood tree (HKY85 model) based on HBV S/pol gene sequences (691 bp) of Ethiopian HBV isolates and reference sequences retrieved from the GenBank database representing HBV genotypes (A-I) and all genotype A and D subgenotypes (A1-A4, and D1-D6). Ethiopian HBV isolates are shown in color (color corresponding to their geographic origin: Green = Addis Ababa; Pink = Adama; Red = Mekelle; Blue = Gondar; Light blue = Jimma), and reference sequences retrieved from GenBank shown by their accession number. Numbers in the tree represent the percentages of bootstrap values (1000 replicates). The subgenotype A1 outgroup cluster, the novel subgenotype D cluster (NC) and the D/A recombinant strain are highlighted in yellow, orange and green, respectively.



# Chapter 4



## A novel hepatitis B virus subgenotype D10 circulating in Ethiopia

G. B. Hundie, V. Stalin Raj, D. GebreMichael, S. D. Pas, M. P. Koopmans, A.D.M.E. Osterhaus, S. L. Smits, B. L. Haagmans

*J Viral Hepat 2017; 24: 163–173.*

## SUMMARY

Hepatitis B virus (HBV) is genetically highly divergent and classified in ten genotypes and forty subgenotypes in distinct ethno-geographic populations worldwide. Ethiopia is a country with high HBV prevalence; however, little is known about the genetic variability of HBV strains that circulate. Here, we characterize the complete genome of 29 HBV strains originating from five Ethiopian regions, by 454 deep sequencing and Sanger sequencing. Phylogenetically, ten strains were classified as genotype A1 and nineteen as genotype D. Fifteen genotype D strains, provisionally named subgenotype D10, showed a novel distinct cluster supported by high bootstrap value and >4% nucleotide divergence from other known subgenotypes. In addition, the novel D10 strains harbored nine unique amino acid signatures in the surface, polymerase and X genes. Seventy-two percent of the genotype D strains had the pre-core premature stop codon G1896A. In addition, 63% genotype A and 33% genotype D strains had the basal core promoter mutations, A1762T/G1764A. Furthermore, four pre-S deletion variants and two recombinants were identified in this study. In conclusion, we identified a novel HBV subgenotype D10 circulating in Ethiopia, underlining the high genetic variability of HBV strains in Africa.



## KEYWORDS

Ethiopia; genetic variability; genotype and subgenotypes; hepatitis B virus; novel subgenotype D10


Abbreviations: BCP, basal core promoter; HBV, hepatitis B virus; ORF, open reading frames; PC, pre Core; RDP, recombination detection program

## INTRODUCTION

Despite advances in prevention, diagnosis and treatment, hepatitis B virus (HBV) infection remains a major global health problem, especially in Africa and Asia, where seroprevalence is high [8]. HBV is the prototype of the family *Hepadnaviridae* with a partially double-stranded circular DNA genome of 3.2 kb organized into four overlapping open reading frames (ORFs: S, for the surface or envelope gene; C, for the core gene; X, for the regulatory X gene; and P, for the polymerase gene) [27]. Because its polymerase lacks proof-reading activity and its genomes undergo frequent recombination, HBV exhibits a high degree of genetic heterogeneity [95]. To date, ten genotypes, A–J, defined by an intergroup divergence of greater than 7.5% over the entire genome sequence have been established [30,32,33,36,37]. Moreover, extensive phylogenetic analyses of HBV genotypes in different studies worldwide have identified several subgenotypes within genotypes, A (A1–A6), B (B1–B9), C (C1–C16), D (D1–D9), F (F1–F4), and I (I1–I2), based on full genome divergence of > 4 % but < 7.5 % [20,30,37-41,53]. HBV genotypes and subgenotypes differ in clinical and therapeutic outcomes [58,60] and have distinct ethno-geographic distribution worldwide [30,52].

Numerous studies carried out in different African countries have also unveiled the emergence of a trend in the geographic distribution of genotypes and subgenotypes. Subgenotype A1 is dominant in southern, central and eastern Africa [30,133,150,155,156,169]. Quasi subgenotype A3 [153], previously designated as genotypes A3, A4, A5, and A7 [40,170], is mainly found in west Africa. Subgenotype D1 is the most prevalent subgenotype in north Africa [133], D2 in Ethiopia [169], D3 in South Africa [133], D7 in Tunisia [39] and the recently described D8 in Niger [38]. Genotype E is the most prevalent strain in central and western Africa [30,38]. However, unlike genotypes A and D, it has very low genetic diversity and does not segregate into distinct subgenotypes. Moreover, in countries where several genotypes circulate many HBV recombinant strains have been described [38,40,49,50].

Ethiopia is a large, culturally diverse, multiethnic country in the horn of Africa located at a strategic geographical position between Africa and Eurasia. It was the main route of human migration out of Africa as well as the origin of modern humans [171-175]. Ethiopia has an intermediate to high HBV endemicity, with HBsAg prevalence of 5.8-6.3% in the general population [8]. However, so far there is limited data available regarding the molecular diversity of HBV in this country [169]. Our previous study was limited to the small S gene sequence analysis and several strains remained unclassified. The aim of this study was to characterize the complete genomes of representative HBV strains of different clades originating from five regions of Ethiopia to classify them into subgenotypes and to gain insight into

the molecular diversity and evolution of the strains in circulation. With 454 deep sequencing and Sanger sequencing, we identified a novel prevalent subgenotype D, tentatively named D10, currently circulating in Ethiopia.

## MATERIALS AND METHODS

### Samples

Twenty-nine HBV strains, 10 genotype A and 19 genotype D obtained from healthy blood donors, were selected for complete genome sequence analyses from the 371 strains previously investigated [169]. These HBV strains originated from five regions — Addis Ababa ($n$ = 11), Adama ($n$ = 6), Gondar ($n$ = 5), Mekelle ($n$ = 6) and Jimma ($n$ = 1) (Table 1). The criteria for representative strain selection were inclusion of strains with identical cluster, divergent and/or separate clusters based on the small S-gene sequences phylogeny, and those with viral load of ≥ 2 log. Samples were stored at −80°C until use. The study was approved by the National Research Ethics Review Committee at the Federal Ministry of Science and Technology, Ethiopia.

### DNA extraction, amplification, and complete genome sequencing

Viral DNA was extracted from 200 µl serum using High Pure Viral Nucleic Acid Kit (Roche Diagnostics, Mannheim, Germany) and QIAamp DNA Mini kit (Qiagen, Hilden, Germany), according to the manufacturer's instructions. Serum HBV DNA levels were quantified using an in-house quantitative real-time PCR (qPCR) as described previously [141,142]. AcroMetrix® HBV low (112 IU/mL to 335 IU/mL), medium (2800 IU/mL to 8900 IU/mL) and high ($5.63 \times 10^6$ IU/mL to $1.78 \times 10^7$ IU/mL) controls (Applied Biosystems, life technology, USA) were used as the standard reference. Next-generation 454 deep sequencing were performed on 13 serum samples with a high viral load (> 4 log) as described previously [176] with modifications. The complete HBV genome for samples with low viral load (≤ 4 log) was amplified in 4 overlapping fragments by nested PCR using primers and protocols described in Table S1. The amplified HBV fragments were sequenced bi-directionally using Big Dye Terminator version 3.1 cycle sequencing kit on an ABI PRISM 3100 genetic analyzer (Applied Biosystems).

### Phylogenetic analyses

The sequences obtained were assembled with the Genomic Workbench software version 6.0.1 (CLC Bio, Aarhus, Denmark). Multiple sequence alignments were performed using CLUSTAL W implemented in MEGA version 6.0 [143].

Phylogenetic trees were constructed by the neighbor-joining method using Kimura two-parameter model and including reference strains of genotypes A–H and all known A and D subgenotypes. To confirm the reliability of the phylogenetic tree topologies, bootstrap resampling and reconstruction were carried out 1,000 times. The obtained tree was visualized in FigTree v1.4.2 (http://tree.bio.ed.ac.uk/software). The Kimura two-parameter and maximum likelihood models were used to calculate genetic divergence and pairwise distances within or between a given genotype and/or subgenotype. All sequences retrieved from GenBank and used in this study were listed in the supplementary material (Table S2).

### Amino acid analyses

All nucleotide sequences in this study and those retrieved from GenBank data bases were translated to amino acids and aligned using CLC Bio software version 6.0.1. Individual subgenotypes were searched for conserved amino acid signatures in the four different open reading frames of the HBV genome.



### Recombination analysis

The detection of recombinant strains was carried out by the recombination detection program (RDP) version 4.69 [177], with the RDP, Geneconv, Chimaera, MaxChi, BootScan, SisScan, 3Seq, and LARD methods. A multiple-comparison-corrected *p*-value cutoff of 0.01 was used throughout the analyses. Only potential recombination events detected by three or more out of the eight independent recombination detection methods implemented in RDP4.69 were considered as genuine recombination events. The analysis was based on a pairwise distance model with a window size of 200, step size of 20, and 1000 bootstrap replicates. These potential recombinant sequences were further validated with SimPlot v3.5.1 [178] and jumping profile Hidden Markov Model (jpHMM) [144].

### Nucleotide sequence accession numbers

The GenBank accession numbers for sequences determined in this study are KX357622-KX357650.

## RESULTS

### Phylogenetic analyses of complete HBV genome sequences from Ethiopia

Of the 29 HBV samples selected, complete genome sequences were successfully determined for 28 HBV samples (9 genotype A and 19 genotype D viruses) and for one sample (AA11ETH), only 2230 nucleotide was amplified, due to shortage

of serum. The demographic and molecular characteristics of 29 sequences are summarized in Table 1. A total of 28 complete and 1 nearly complete HBV genome sequences were phylogenetically analyzed with HBV reference sequences of genotypes A-H retrieved from the GenBank database. Phylogenetically, ten strains grouped with genotype A and 19 strains with genotype D (Figure 1). All the 10 genotype A strains were further classified into subgenotypes A1, but separated into two clades. The majority of strains (9 strains) clustered in the Asian-American clade (aac) and one strain in the African clade (ac) (Figure 1). Interestingly, the nine A1 strains are phylogenetically related to strains from Japan, Haiti, Brazil, Philippines and Somalia. The mean intergroup nucleotide divergence between Ethiopian A1 strains (excluding AA46ETH) and strains from those countries were: 1.7% (Japan), 1.9% (Haiti and Somalia), 2% (Brazil), and 2.1% (Philippines), while its divergence from sub-Saharan African strains was 3%.

The nineteen complete genotype D sequences clustered into three distinct groups (Figure 1). Three strains previously shown to belong to subgenotype D2 (Table 1), were also classified within D2 at the complete genome level. As shown in Figure 1, of the remaining sixteen strains, fifteen D strains clustered separately from all other D subgenotypes and formed a distinct novel cluster with a strong bootstrap support (100%), whereas one strain (MK141ETH) was an outgroup. As inclusion of the recombinant strains into phylogenetic analysis might change the topology of the tree, a second phylogenetic tree was constructed for genotype D by omitting D8 and D9 (Figure 2). In the second phylogenetic tree, the MK141ETH clustered with subgenotype D7 (D7 is a major parent as seen later from recombinant analysis) but all the fifteen novel strains clustered separately again. Because these novel strains formed a distinct monophyletic cluster with a significantly high bootstrap support (100%), we tentatively assigned them to a novel subgenotype, D10.

**Table 1. Demographic and molecular characteristics of 29 HBV sequences**

| Strain[a] | Age | Sex | Viral load log IU mL$^{-1}$ | Genotype and subgenotype[b] | | | Nucleotide length (bp) | Positions of pre-S2 deletions |
|---|---|---|---|---|---|---|---|---|
| | | | | S gene (previous study)[d] | Full genome (this study) | Serotype | | |
| AA33ETH | 33 | Male | 5 | D/ND | D10 | *ayw2* | 3182 | n/a |
| AA95ETH | 29 | Male | 6 | D2 | D2 | *ayw3* | 3182 | n/a |
| AA124ETH | 28 | Male | 3.5 | D/ND | D10 | *ayw2* | 3182 | n/a |
| AA141ETH | 30 | Male | 3 | D/ND | D10 | *ayw2* | 3158 | 32-55 |
| AA164ETH | 40 | Male | 3 | D/ND | D10 | *ayw2* | 3182 | n/a |
| BL43ETH | 45 | Male | 8 | D/ND | D10 | *ayw2* | 3182 | n/a |
| BL63ETH | 29 | Male | 3 | D/ND | D10 | *ayw2* | 3282 | n/a |
| BL60ETH | 35 | Male | 2 | D/ND | D10 | *ayw2* | 3182 | n/a |
| BL70ETH | 22 | Male | 3 | D/ND | D10 | *ayw2* | 3182 | n/a |
| GD39ETH | 23 | Male | 8 | D/ND | D10 | *ayw2* | 3182 | n/a |
| GD65ETH | 33 | Male | 4 | D/ND | D10 | *ayw1* | 3173 | 50-55 & 3172-3174 |
| GD69ETH | 22 | Male | 4 | D/ND | D10 | *ayw2* | 3182 | n/a |
| GD82ETH | 34 | Female | 8 | D2 | D2 | *ayw3* | 3182 | n/a |
| GD104ETH | 32 | Male | 3 | D/ND | D10 | *ayw2* | 3182 | n/a |
| MK20ETH | 42 | Male | 6 | D2 | D2 | *ayw3* | 3182 | n/a |
| MK74ETH | 30 | Male | 5 | D4 | D10 | *ayw2* | 3182 | n/a |
| MK75ETH | 30 | Male | 4 | D/ND | D10 | *ayw2* | 3182 | n/a |
| MK152ETH | 46 | Male | 3 | D/ND | D10 | *ayw2* | 3176 | 50-55 |
| MK141ETH | 25 | Male | 2 | D/A | D/A/E | *adw2* | 3212 | n/a |
| AA11ETH | 35 | Male | 3 | A1 | A1 | *adw2* | 2230[c] | n/a |
| AA46ETH | 25 | Male | 2 | A1 | A1/D | *adw2* | 3221 | n/a |
| AA81ETH | 30 | Male | 2 | A1 | A1 | *adw2* | 3223 | n/a |
| AA86ETH | 40 | Male | 3 | A1 | A1 | *adw2* | 3221 | n/a |
| AA152ETH | 24 | Male | 6 | A1 | A1 | *adw2* | 3170 | 5-55 |
| AA162ETH | 23 | Male | 7.5 | A1 | A1 | *adw2* | 3221 | n/a |
| BL13ETH | 26 | Female | 7 | A1 | A1 | *adw2* | 3221 | n/a |
| BL185ETH | 33 | Male | 8 | A1 | A1 | *adw2* | 3221 | n/a |
| JM16ETH | 22 | Male | 7 | A1 | A1 | *adw2* | 3221 | n/a |
| MK156ETH | 22 | Male | 7 | A1 | A1 | *adw2* | 3221 | n/a |

[a]Strain origin; the first two letters indicate region (AA= Addis Ababa, BL= Adama, GD= Gondar, JM= Jimma, MK= Mekelle) and the last three letters country (ETH= Ethiopia).

[b]ND, not determined; n/a, not applicable; [c]AA11 partial sequence (2230 nt, excluding the polymerase C terminal and the X gene). [d]Hundie et al.[169]

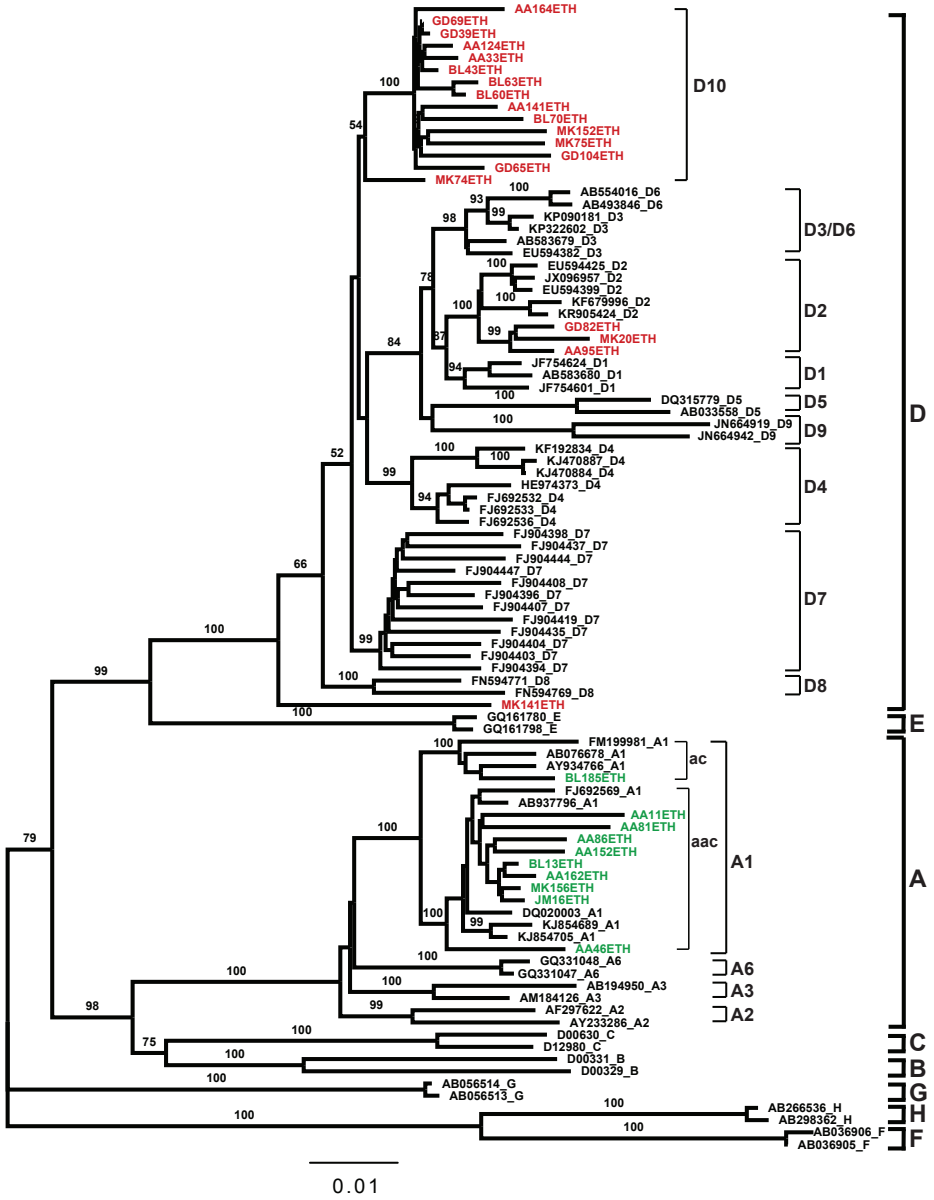



**Figure 1. Phylogenetic tree constructed by the neighbour-joining method, based on the complete sequences of HBV strains.** Ethiopian sequences determined in this study are indicated by color (green = genotype A: red = genotype D). The reference sequences used in this comparison are indicated by their accession numbers, followed by the letter of the corresponding genotype/ subgenotype. Bootstrap resampling was carried out 1000 times and is indicated on the respective branches as a percentage

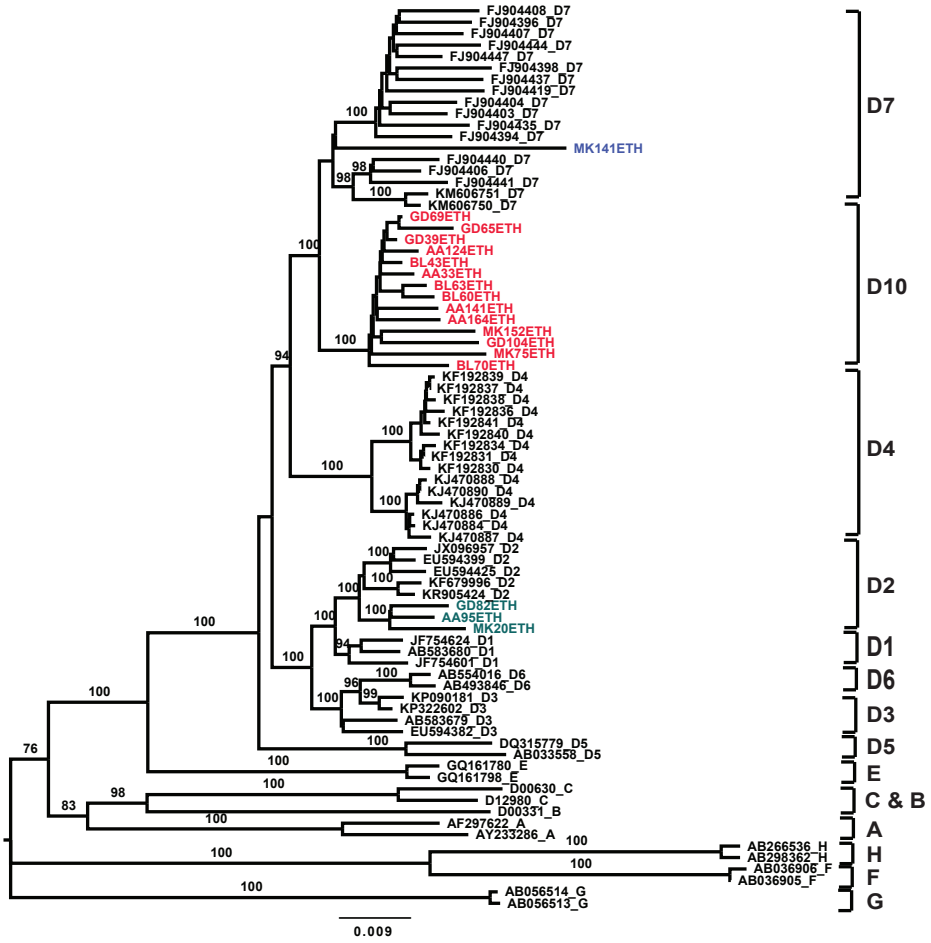



**Figure 2. Phylogenetic analysis of Ethiopian HBV strains belonging to genotype D compared with reference sequence of each genotype (A-H) and seven subgenotypes of D (D1–D7).** Genetic distances were estimated using the Kimura two-parameter model, and phylogenetic tree was constructed using the neighbor-joining method. GenBank accession numbers of all published sequences used, followed by the letter of the corresponding genotype or subgenotype, are indicated. Ethiopian strains are shown in color and the novel subgenotype D10 depicted in red. Bootstrap resampling was carried out 1000 times and is indicated on the respective branches as a percentage.

## Genetic characteristics of the novel HBV subgenotype D10

To further explore the characteristics of the strains belonging to the novel D10 subtype, phylogenetic trees were constructed for each of the four ORFs; preS/S, P, preC/C, and X, and the representative sequences of subgenotypes D1 to D7 were used as references. As depicted in Figure 3A-D, the HBV D10 strains formed a distinct monophyletic cluster in all four individual ORFs (preS/S, P, preC/C,

and X) supported by high bootstrap values. This result further confirmed that the D10 strains cluster distinctly not only at complete genome level but also when analyzing each of the four individual ORFs. It is noteworthy that D10 strains were distantly close to strains belonging to subgenotypes D1 and D2 in the X ORF tree (Figure 3D), and to D7 in the preS/S, P and preC/C ORFs tree (Figure 3A-C). This characteristic was also observed when analyzing the different shared amino acid residues in these ORFs.

To further characterize the D10 strains, we compared intergroup nucleotide divergence (mean ± SD %) over the complete genome to 132 different HBV sequences representing subgenotypes D1–D7. The mean divergence of the novel strains was found to be 4.9 ± 0.4% from each D1 and D2, 5.4 ± 0.4% from D3, 4.0 ± 0.3% from each D4 and D7, 6.2 ± 0.4% from D5, and 4.8 ± 0.4% from D6. Therefore, our data further provide support for the current definition of subgenotype [30,37,53].

Furthermore, comparison of amino acid sequences of each of the four ORFs/genes (preS/S, P, preC/C and X) between HBV D10 strains and other subgenotypes (D1–D9) identified nine amino acid signature positions unique to D10 strains; five of them were exclusively unique to D10, and four of them were shared with one or two strains of other D subgenotypes (Table 2). Of the five unique amino acid sites, three were found in the P gene and one each in the preS/S and X genes. Briefly, all the D10 strains had $Pro^{78}$ in the spacer, $Tyr^{124}$ in reverse transcriptase and $Gln/Asn^{136}$ in terminal protein of the P gene, $Ser^{33}$ in the preS2 gene, and $Pro^{47}$ in the X gene (Table 2). The four amino acid sites that are unique to D10 but shared with a few other D strains includes $Thr^{28}$ in the preS1 gene and $Leu^{89}$ in terminal protein, $His/Tyr^{31}$ in the spacer and $Ala^{317}$ in the reverse transcriptase of the P gene (Table 2). No D10- specific amino acid residues were found in the core and small S regions. Amino acid substitutions present within each subgenotype are shown in Table 2. The subgenotype-specific amino acid residues are shown in italic and the dominant ones in bold.

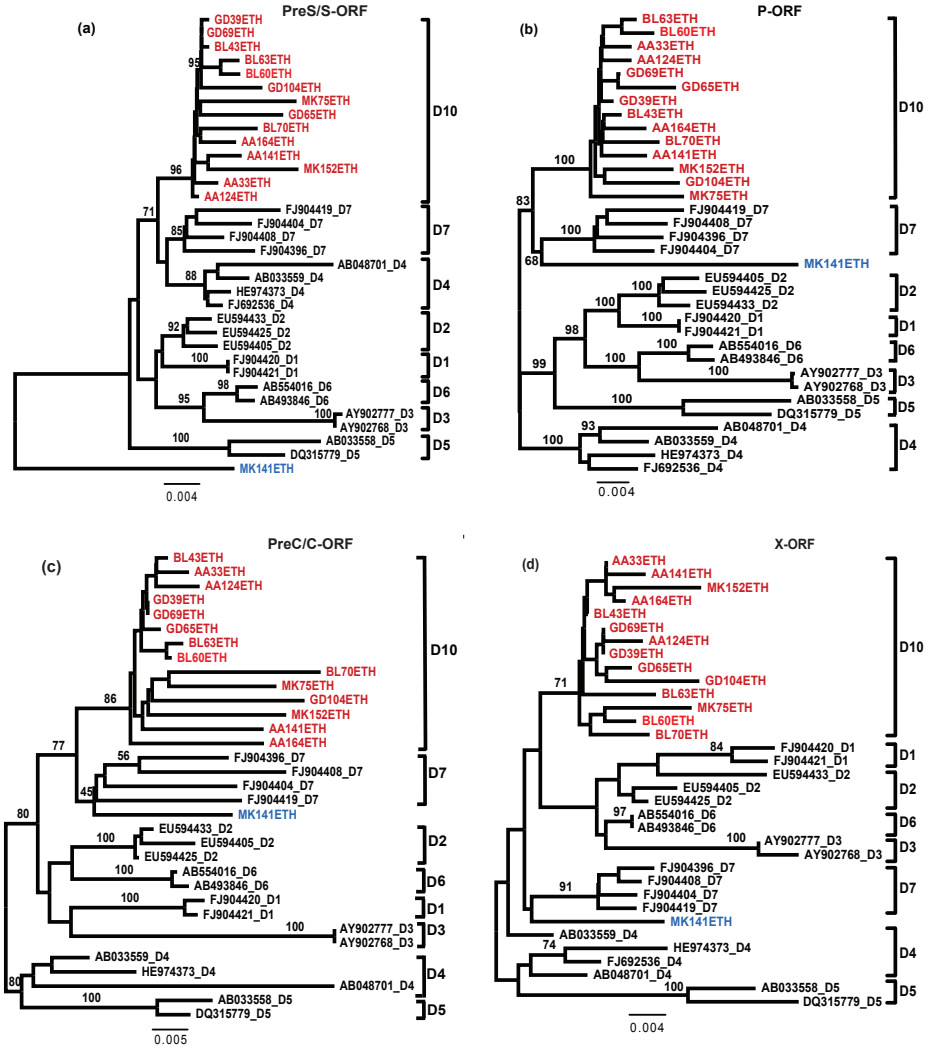



**Figure 3. Neighbour-joining trees based on phylogenetic analysis of the four open reading frames (ORFs) of 34 HBV strains belonging to genotype D;** (**a**) preS/S, positions 2848-3182, 1-835, (**b**) polymerase, positions 2307-3182, 1-1623 (**c**) preC/C, positions 1814-2452, and (**d**) X , positions 1374-1838. Reference sequences of seven subgenotypes of HBV genotype D (D1–D7) retrieved from GenBank were used for the construction of the phylogenetic tree. The tree was built using the Kimura two-parameter model and neighbor-joining method, and bootstrap resampling was carried out 1000 times and is indicated on the respective branches as a percentage. Ethiopian D10 strains are shown in red, MK141ETH in blue, and reference sequences are indicted by GenBank accession numbers, followed by the letter of the corresponding subgenotype.

**Table 2. Comparison of amino acid residues in the preS/S, polymerase (P), preC/C and X genes of D10 strains compared with the previously reported residues at the same position in different sub genotypes (D1-D9).**

| Gene[a] | Amino acids position | D subgenotypes[b] | | | | | | | | | | % of major aa |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | D1 | D2 | D3 | D4 | D5 | D6 | D7 | D8 | D9 | D10 | |
| **PreS/S** | | | | | | | | | | | | |
| pre-S1 | 28 | A | A | A | N | R | A | N/T [c] | R | A | **T** | 100 |
| pre-S1 | 74 | L | I | L/I | I/M,L | L | L | I/M,L | I | L | **I** | 100 |
| pre-S1 | 103 | N/D | N | N | N | D | N | N/D,V | D | N | **N**/D | 60 |
| pre-S2 | 33 | N | N | N | N | N | N | N | N | N | *S* | 100 |
| pre-S2 | 39 | V | A | A | A/V | A | A | A/V | A | A | **A** | 100 |
| pre-S2 | 41 | P/H | P | P/H | H | P | P | P | P | P | **H** | 100 |
| pre-S2 | 42 | I | I | L | L/I | I | L | I | I | L/I | **I** | 100 |
| S | 118 | T | *A/V* | T | T | T | T | T | T | T | **T** | 100 |
| **P gene** | | | | | | | | | | | | |
| tp | 16 | D | D | D | D | D | D | E | E | D | **E** | 100 |
| tp | 89 | I | I | I | I | I | I | I/L | I | I | **L**/I | 86.7 |
| tp | 90 | K | K | K | E/K | N/Y | K | K | K | K | **K**/N,Q | 86.7 |
| tp | 91 | K | K | K | K | K | K | R/K | K | K | **K** | 100 |
| tp | 101 | V | V | V | V | V | V | I/V | V | V | **V** | 100 |
| tp | 108 | Q | Q | Q | K | Q | Q | K/Q | K | Q | **K** | 100 |
| tp | 119 | V | V | V | V/F | V/F | V | F/R | V/G | V | **F**/V | 93.3 |
| tp | 136 | H | H | H | H | H | H | H | H | H | *Q/N* | 93.3 |
| tp | 156 | I | I | I | *V/F* | I | I | I | I | I | **I** | 100 |
| tp | 162 | T | T | T | S | T | T | T/S | S/T | T | **S** | 100 |
| sp | 31 | R | R/S | R | Q | Q | R | Q/H | Q | R | **H**/Y | 86.7 |
| sp | 32 | K | K | K | Q | K | K | Q/K | K | K | **K**/T | 93.3 |
| sp | 58 | G | G | G | R/G | G/R | G | G/R | G | G | **G** | 100 |
| sp | 78 | T | T | T | T/A | T | T | T/A,S,N | T | T | *P* | 100 |

**Table 2. Comparison of amino acid residues in the preS/S, polymerase (P), preC/C and X genes of D10 strains compared with the previously reported residues at the same position in different sub genotypes (D1-D9). (continued)**

| Gene[a] | Amino acids position | D subgenotypes[b] | | | | | | | | | | % of major aa |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | D1 | D2 | D3 | D4 | D5 | D6 | D7 | D8 | D9 | D10 | |
| sp | 95 | K | K | K | T | K | K | K/T | K | K | **K** | 100 |
| rt | 124 | H | H | H | H | H | H | H | H | H/N | *Y* | 100 |
| rt | 126 | H | *R* | H | H | H | H | H | H | H | **H** | 100 |
| rt | 149 | Q | Q | Q | K | Q | Q | K | K | Q | **K** | 100 |
| rt | 129 | M | M | M | L/M | M | M | M | M | M | **M** | 100 |
| rt | 153 | R | R | R | W/R | R | R | W | W | R | **W**/R | 86.7 |
| rt | 237 | P | P | P | P | P | P | T | T | P | **T** | 100 |
| rt | 267 | Q | Q | Q/H | L/H | H | Q | H | H | Q | **L**/H,Q | 86.7 |
| rt | 317 | S | S | S | S | S/A | S | S/A | S | S | **A** | 100 |
| rh | 20 | M | M | M | M/L | M | M | **I** | **I** | M | **M**/I | 93.3 |
| rh | 107 | L | L | L | *P* | L | L | I | I | L | **L** | 100 |
| rh | 138 | D | D | D | D/V | D | D | V/D | D | D | **D** | 100 |
| **C gene** | | | | | | | | | | | | |
| C | 74 | G | V | V | T/A | T | V | V/A,T | V | S | **V**/G | 93.3 |
| C | 116 | I | I | I | L | L | I | I/V,L | L | L | **I**/V,L | 73.3 |
| **X gene** | | | | | | | | | | | | |
| | 26 | R | R | R | R | R | R | C | C | R/C | **C** | 100 |
| | 36 | T | T/P | T | T/A,S | S | T | *D* | *D* | T | **T** | 100 |
| | 47 | T | T/S | T | A/T | T | T | A/T | T | T/S | *P* | 100 |
| | 88 | F | F | I | I | I/F | - | I | I | I | **F/C,S,**I | 73.3 |

[a]tp, terminal protein; sp, spacer; rt, reverse transcriptase, rh ribonuclease (RNase) H.

[b]Number of the subgenotypes included in the analysis: D1 = 31, D2 = 33, D3 = 5, D4 = 22, D5 = 11; D6 = 6, D7 = 24, D8 = 3, D9 = 5 and D10 = 15.

[c]The relative size of the font indicates the frequency of the amino acid for each subgenotype at that position. The amino acid residues that are unique to an individual subgenotype are italicized, whereas those that were found predominantly in D10 but shared by strains belonging to the other subgenotypes are shown in bold. The major amino acid of the D10 strain & its relative frequency are indicated in %.

**Mutations in the BCP/Precore region in Ethiopian HBV strains**

Ethiopian HBV strains in this study were analyzed for mutations in the basal core promoter (BCP) and precore (PC) region. The most common BCP mutation, a double A1762T/G1764A nucleotide substitution, was found in five (5/18; 27.8%) genotype D and one (1/8; 12.5%) genotype A strain. In addition, six strains, four from genotype A and one each from genotype D and the recombinant strain (AA46ETH) had a single G1764A mutation. One genotype A strain and two genotype D strains had triple BCP mutations, A1762T, G1764A and T153C. Overall, 63% of genotype A strains and 33% of genotype D strains had either double or single BCP A1762T/G1764A mutations. The PC premature stop codon G1896A was identified in thirteen of eighteen (72.2%) investigated genotype D strains and in both recombinant strains (MK141ETH and AA46ETH). None of the genotype A strains carried the precore stop codon mutation. Another PC mutation, G1899A, was found in nine (50%) genotype D strains and the two recombinant strains but not in genotype A strains. Moreover, two genotype D strains (AA141ETH and BL70ETH) had PC start codon mutations, and 1 genotype A strain (AA81ETH) had two nucleotide insertions at nucleotide 1839. Overall, genotype D strains exhibited a high frequency of BCP/PC mutations, with 15 strains out of 18 (83.3%) disclosed these mutations. Notably, 80% (12/15) of the D10 strains had at least one mutation in the BCP or PC region.

**Pre-S deletion mutations in Ethiopian HBV strains**

The prevalence of the pre-S mutations was also assessed in this study. Of the 29 HBV strains investigated, four strains (13.8%) revealed a pre-S deletions, ranging from 6 nt to 51 nt (Table 1). Three of these strains are from genotype D and one strain is from genotype A. All of the pre-S deletions occurred in the N- terminal of the pre-S2 region of the HBV genome. In addition, one genotype D strain (AA33ETH) had a pre-S2 start codon mutation.

**Recombination analysis**

All 29 sequences from Ethiopia were evaluated for potential recombination events using RDP4, SimPlot and jpHMM methods. Two strains were found to be a putative recombinant. Strain MK141ETH was identified as a triple recombinant between genotype D, E and A by jpHMM (Figure S1), and classified as a genotype D/E recombinant by RDP4 and SimPlot. As shown in Figure 4B, MK141ETH had genotype E sequence in the Pol N-terminal and preS2 regions from position 2702 to 2978 (beginning breakpoint 99% confidence interval (Cl): 2680-2755; end breaking point 99% Cl: 2793-2995), and genotype D (subgenotype D7) sequence in rest of the genome. Strain AA46ETH, which belongs to subgenotype

A1, showed a genotype D sequence in the X and pre-core region from position 1558-1916 (beginning breakpoint 99% Cl: 1391-1633; end breaking point 99% Cl: 1859-1983) (Figure 4A) as analyzed by RDP4 and jpHMM (Figure S1).

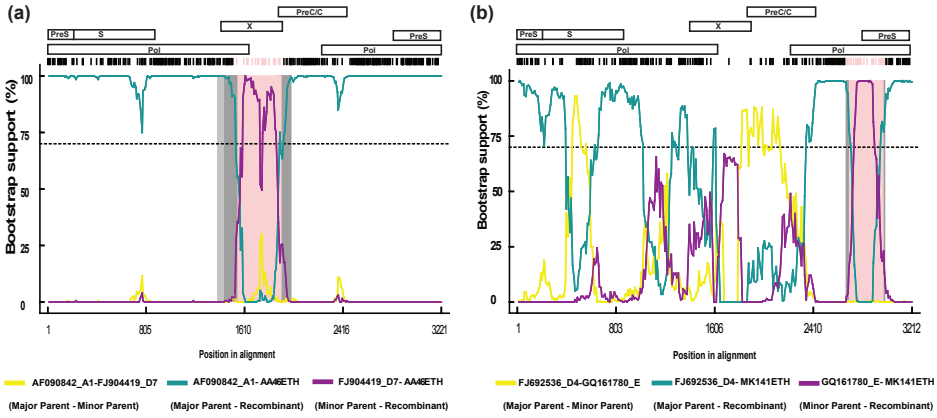



**Figure 4. Bootscan analysis of two recombinant, major parent, and the minor parent sequences (a and b).** The analysis was based on a pairwise distance model with a window size of 200, step size of 20, and 1000 bootstrap replicates generated by the RDP4 program. The left and right boundaries of the pink area indicate approximated recombination breakpoint positions. The recombinant area is shown in pink.

## DISCUSSION

This is the first study describing the full-genome molecular characterization of HBV strains circulating in different geographic regions of Ethiopia. Here, we investigated 29 complete HBV genome sequences, ten sequences belonging to genotype A and 19 to genotype D. The most important finding of this study was the identification of novel HBV strains that clearly form a new separate cluster, proposed subgenotype D10, within genotype D. Firstly, these strains were obtained from unrelated blood donors from different regions of Ethiopia. Secondly, complete genome and separate ORF phylogeny revealed that they formed a distinct monophyletic cluster strongly supported by high bootstrap value (100%). Thirdly, the mean inter-subgenotype nucleotide divergence compared with reference sequences of all existing subgenotypes D was > 4 % and < 7.5%. Lastly, they had nine unique amino acid substitution positions, five of those were exclusively unique to D10.

In our previous study, we reported a unique distribution of genotype D viruses in Ethiopia, with 55 % (44/80) of strains belonging to subgenotype D2 and 30% (24/80) unassigned new strains, based on the small S-gene sequence variations

[169]. In the present study, we performed a complete genome sequence analysis on fourteen unassigned new strains, three strains from D2, and one strain each from D4 and a suspected recombinant. As expected, the three strains were classified into D2, confirming our earlier findings. It is noteworthy that D2 is a major subgenotype in Europe but less frequent in Africa [157]. Interestingly, strain MK141ETH that was suspected as D/A recombinant in the previous study was found to be a triple recombinant (D/A/E) in this study. It is well known that recombination between HBV genotypes is a common event in countries where multiple genotypes are prevalent [38,40,49,50]. However, strains belonging to genotype E has not been reported so far in Ethiopia. Thus, this finding highlights a need for more sequences from this region. On the other hand, strain MK74ETH previously classified as D4 was classified as D10 at complete genome level. It is noteworthy to say, subgenotype classification may not be applicable to some HBV strains on the basis of the small S-gene sequence alone [53]. Interestingly, the novel D10 strains distinctly clustered not only on the complete genome level but also in each of the four open reading frames (preS/S, P, preC/C and X). This consistent result provides concrete evidence that D10 strains have homologous sequences throughout the genome and is free of hybrid sequences and recombination. This result was further confirmed by HBV recombinant analysis. In addition, the mean inter-subgenotype nucleotide divergence over the complete genome sequence between D10 and the recognized D subgenotypes ranged from 4 % (with D4 and D7) to 6.2% (with D5), which is the minimum requirement for definition of a new subgenotype [30,37,53].

Most recently, Pourkarim and colleagues [53] suggested that to introduce a novel subgenotype, strains harboring specific amino acid motifs should be identified. Fascinatingly, besides amino acid residues previously reported [30,56,157], D10 strains disclosed nine unique amino acid signatures, five exclusively unique to D10, and four only shared with very few other D strains. The novel D10 is easily distinguished from all subgenotypes, D1-D9, by $Ser^{33}$ in the preS2 gene, $Pro^{47}$ in the X gene, and $Pro^{78}$, $Tyr^{124}$ and $Gln/Asn^{136}$ in the P gene. Remarkably, many other amino acids residues specific to D10 but shared with few D4 or D7 strains, or common to all the three subgenotypes were identified. As seen in this study, the novel D10 is distantly related more to D4 and D7 than other subgenotypes from some amino acid signatures they shared, which was also revealed in the phylogenetic tree. However, the three subgenotypes prevail in geographically and historically unrelated regions, D4 in Oceania [30] and D7 in Maghreb [39,56]. Hence, further investigation is needed to determine the molecular evolutionary relatedness between the three subgenotypes.

Subgenotype A1, the main African A subgenotype, is the predominant strain in Ethiopia [169], alike most sub-Saharan African countries [27,133,150]. Surprisingly, the majority of Ethiopian A1 strains in this study clustered in the Asian-American clade together with Somalian strains than to the African clade. They are more closely related to strains from Japan, Haiti, Brazil, Somalia, and the Philippines. This result might strengthen the previous hypothesis of the dispersal of A1 into these regions from eastern Africa possibly through migration or slave trade [154]. A study in Haiti suggests that A1 was introduced from eastern Africa to Haiti during the slave trade around the turn of the 18th century [179]. Another study in Brazil revealed the close relatedness of the Brazilian, Asian and Somalian isolates, and suggests that the Brazilian A1 strains originate from southeast Africa at the middle of the 19th century [180]. Moreover, in the recent subgenotype A1 of HBV-tracing human migrations in and out of Africa, the co-clustering of Somalian and Latin American strains and the plausible sites of A1 origin and dispersal routes to Asia and Latin America are presented [154]. Furthermore, Ethiopia was the main route of human migration out of Africa and also the origin of modern humans as abundant key fossil evidence for human origins and evolution is found in modern-day Ethiopia [171-175]. From these shreds of evidence we speculate that Ethiopian strains might be the oldest lineage of the Asian-American clade, but should be studied further using modern phylogenetic analyses.

Another important observation in the present study was the high frequency of the classical precore and basal core mutations, and the pre-S deletion mutations. The majority (75%) of genotype D strains in this study harbored the precore premature stop codon G1896A mutation. This is in agreement with some studies [181,182]. In addition, 63% genotype A strains and 33% genotype D strains had the BCP mutation. Previous studies have indicated that BCP/PC mutations are the strongest viral factor associated with advanced liver disease progression in HBV carriers [183-185]. Moreover, it has been described that mutations in the BCP/PC region are correlated with HBV genotypes [41,184]. Thus, it is noteworthy to suggest that the clinical impact of the novel D10 strain in chronic HBV carriers needs further study. Besides, we detected one strain (AA81ETH) with two nucleotide insertion at position 1839 of the core region that was confirmed in 3 independent experiments. This insertion resulted in a frameshift and caused a premature stop codon that resulted in a truncated form of the core protein. We have also noted that 13.8% of strains in the present study harbored a pre-S deletion and one strain had a pre-S2 start codon mutation. These mutations were reported in a number of studies and suggested as independent risk factors for the development of hepatocellular carcinoma [186,187]. As the aforementioned mutations were observed

among healthy blood donors, our finding has thus an impact on the establishment of HBV patient management in Ethiopia.

In conclusion, this is the first report of complete HBV genome sequences from Ethiopia, where there have been scant molecular data on HBV. We identified a novel subgenotype D10 with a well-defined pattern, including unique amino acid signatures. This study highlights that HBV is still evolving genetically, and further studies in HBV endemic areas where molecular data are lacking are of paramount importance.

**Acknowledgements**

This study was supported by Nuffic grant from the Netherlands Fellowship Programmes (grant no. CF8605).

**Conflict of interest**

The authors declare no conflict of interest.

## SUPPORTING INFORMATION

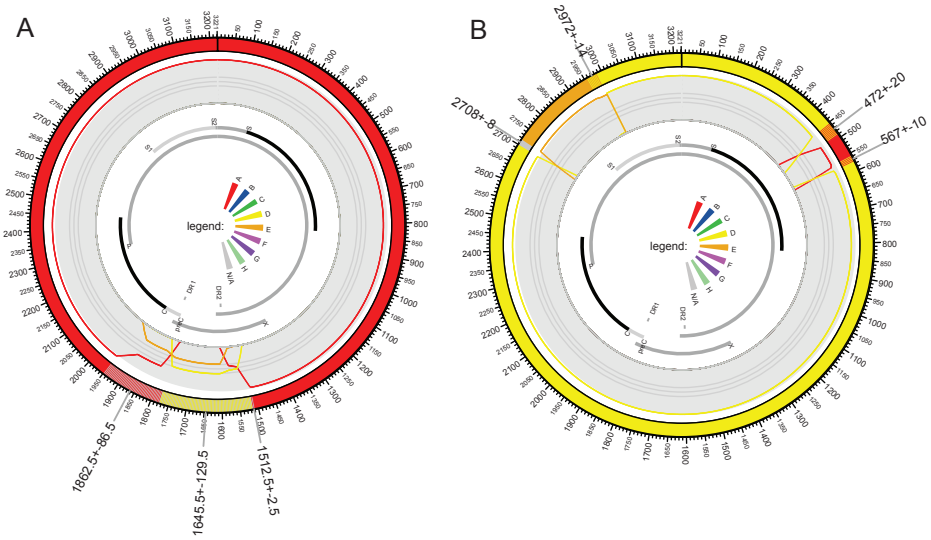



**Figure S1. Schematic diagram of jpHMM results showing the circular HBV genome recombinant AA46 ETH (A) and MK141ETH (B).**

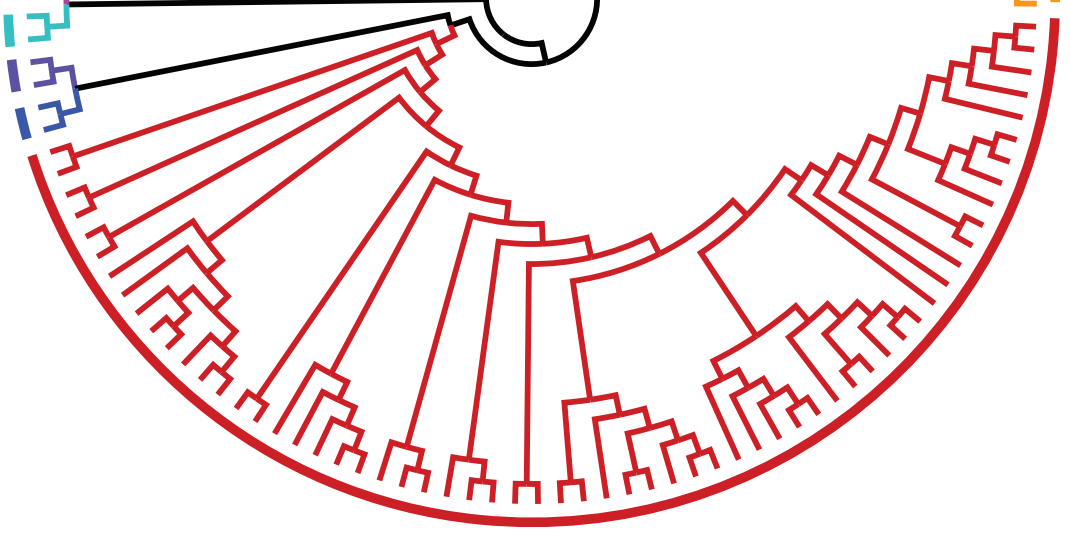
**Table S1. Primers used for complete genome DNA amplification and sequencing**

| Primers† | Nucleotide sequence (5' → 3') | Polarity | Position | Domain |
|---|---|---|---|---|
| **Sanger sequence** | | | | |
| HBPr1 | GGGTCACCATATTCTTGGG | Sense | 2850-2868 | HBPol |
| HBPr3 | CCACTGCATGGCCTGAGGATG | Antisense | 3226-3246 | preS1/S2/HBPol |
| HBPr33 | CTGAGGGCTCCACCCCA | Antisense | 3104-3120 | preS1/HBPol |
| HBPr94 | GGTAWAAAGGGACTCAMGATG | Antisense | 775-795 | HBPol/HBsAg |
| HBPr108 | TTTTTCACCTCTGCCTAATC | Sense | 1821-1840 | HBX/preCore |
| HBPr109 | AAAAAGTTGCATGGTGCTGG | Antisense | 1806-1825 | HBX/preCore |
| HBPr110 | CCTCTGCCGATCCATACTGCGGAAC | Sense | 1255-1279 | HBPol |
| HBPr111 | CTGCGAGGCGAGGGAGTTCTTCTTC | Antisense | 2406-2430 | Core/HBPol |
| HBPr113 | CCGGCAGATGAGAAGGCACAGACGG | Antisense | 1549-1574 | HBX/HBPol |
| HBPr134 | TGCTGCTATGCCTCATCTTC | Sense | 414-433 | HBPol/HBsAg |
| HBPr303 | CCCACCTTATGAGTCCAAGG | Antisense | 2493-2512 | HBPol |
| HBPr440 | TATGGATGATGTGGTATTGGG | Sense | 738-758 | HBPol/HBsAg |
| HBPr446 | GGAGTGTGGATTCGCACTCC | Sense | 2303-2323 | Core |
| S1 | GTATGTTGCCCGTTTGTCCTC | Sense | 459-479 | HBPol/HBsAg |
| C1 | TAGGAGGCTGTAGGCATAAATTGGT | Sense | 1774-1798 | HBX/preCore |
| C20 | GCTTCACCTCTGCACGT | Sense | 1588-1604 | HBX/HBPol |
| HT26-2 | CCTGCTGGTGGCTCCAGTTC | Sense | 56-75 | HBPol/HBsAg |
| HT26-3 | ATAAAACGCCGCAGACACATCCAGCGA | Antisense | 370-396 | HBPol/HBsAg |
| HT26-4 | TGAGAGAAGTCCACCACGAGTCTAGA | Antisense | 247-272 | HBPol/HBsAg |
| HT26-6 | ACTTTCCAATCAATDGG | Antisense | 970-986 | HBPol |
| Ymdd2triple mix: | | | | |
| Ymdd2a | ACCCCATCTTTTTGTTTT | Antisense | 844-861 | HBPol |
| Ymdd2genoC | ACCCCAACGTTTGGTTTTATTAGG | Antisense | 838-861 | HBPol |
| Ymdd2 | ACCCCATCTTTTTGTTTTGTTAGG | Antisense | 838-861 | HBPol |
| **454 sequence** | | | | |
| VS806 | CTGCTGGTGGCTCCAGTT | Sense | 57-74 | HBPol/HBsAg |
| VS807 | CCCTACGAACCACTGAACAAA | Antisense | 689-709 | HBPol/HBsAg |
| VS808 | TGCACGACTCCTGCTCAA | Sense | 524-541 | HBPol/HBsAg |
| VS809 | GGTTGCGTCAGCAAACACT | Antisense | 1179-1197 | HBPol |
| VS810 | ACTTTCTCGCCAACTTACAAGG | Sense | 1093-1114 | HBPol |
| VS811 | CTCTTATGTAAGACCTTGGGC | Antisense | 1640-1160 | HBX/HBPol |
| VS812 | GCGCACCTCTCTTTACGC | Sense | 1524-1541 | HBX/HBPol |
| VS813 | CCCAGGTAGCTAGAGTCAT | Antisense | 2095-2114 | Core |
| VS814 | TGCCTTCTGACTTCTTTCCTTC | Sense | 1956-1977 | Core |
| VS815 | TTTCCCACCTTATGAGTCCAAG | Antisense | 2464-2485 | HBPol |
| VS816 | AGACCACCAAATGCCCCTAT | Sense | 2297-2317 | Core |
| VS817 | GGATTGAAGTCCCAATCTGG | Antisense | 2974-2993 | preS1/HBPol |
| VS818 | GGGTCACCATATTCTTGGGA | Sense | 2820-2839 | HBPol |
| VS819 | TAACACGAGCAGGGGTCCTA | Antisense | 180-199 | HBPol/HBsAg |

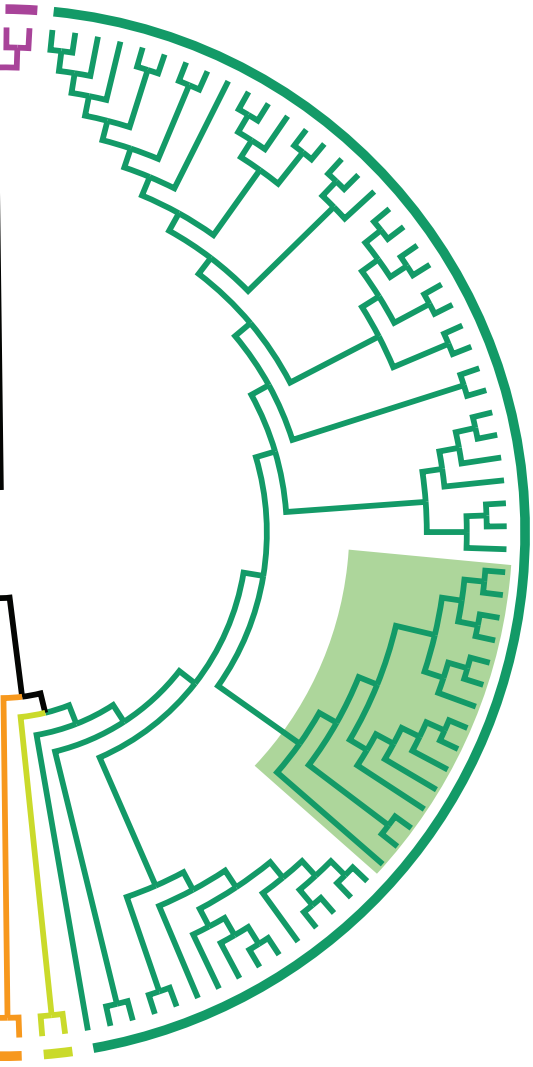
†HBPr. primers are from (Stuyver & others, 2000), the remaining are in-house designed ones.

**Table S2. List of GenBank accession numbers of reference sequences used in this study**

| Genotype | GenBank accession number |
|---|---|
| A1 | AB937796, AB937791,AB453988, AB453987,FJ692590, FJ692557, FJ692574,FJ692578,FJ692588, FJ692563, FJ692571, FJ692560,FJ692564, FJ692565, FJ692569, FJ692559, FJ692566, AB241115, AY934774, EU410082, AB241114, AB116094, M57663, AB116093, AY934769, AY934768, AY934770, AY934771, AY934767, KJ854705, KJ854692, KJ854702, KJ854687, FJ692591, KJ854704, KJ854697, KJ854707, KJ854703, KJ854696, KJ854694, KJ854706, KJ854700, KJ854688, KJ854686, KJ854690, JX154581, AM494718, JX154580, AB076678, AY934773, HM535200, FM199979, KT347088, AY233288, FM199980, JN182328, FM199981, FM199977, DQ020002, AY934772, FM199974, AB076679, AY934766, DQ020003, KJ854689 |
| A2 | AF297622, AY233286 |
| A3 | AB194950, AM184126 |
| A6 | GQ331048, GQ331047 |
| B | D00331, D00329 |
| C | D00630, D12980 |
| D1 | FJ904432, FJ904421, FJ904420, GQ477459, GU456653, JF754596, FJ904415, AF151735, GU456655, FJ899792, AJ344116, GU456646, GQ183483, GU456640, GU456647, AB188244, JF754602, GU456684, GU456639, JF754632, JF754624, AB583680, JF754601, AF280817, AB104709, AB104711, AF121242, JF754634, FJ904424, GQ477458, GU456651 |
| D2 | AB078033, AB078031, AB078032, AB109477, AB109478, EU594425, EU594430, EU594405, EU594433, EU594410, EU594402, EU594400, GQ477457, EU594407, EU594428, EU594409, EU594432, EU594398, AF043594, GQ477453, JX096957, EU594399, KF679996, KR905424, JF754597, AY090453, JN642143, JN642160, EU594401, EU594406, EU594431, EU594416, AB110075 |
| D3 | AY902768, AY902769, AY902777, AY902773, AY902770, AJ344117, DQ315776, HQ236014, AY233296, HQ236015, GQ922001, HQ236016, AY233294, EU594434, GQ922000, KP090181, KP322602, AB583679, EU594382 |
| D4 | FJ692532, FJ692533, FJ692536, HE974373, AB033559, GQ922003, AB048701, KJ470884, KJ470890, KJ470886, KJ470888, KJ470891, KJ470887, KM606744, KF192831, KF192841, KF192837, KF192839, KF192830, KJ470889, KF192834, KF192838, KM606752, KF192840, KF192836 |
| D5 | DQ315779, GQ205384, GQ205377, GQ205387, GQ205378, GQ205389, GQ205379, GQ205382, GQ205385, GQ205388, AB033558 |
| D6 | AB493846, AB554016, AB554023, AB554024, AB493848, AB493845 |
| D7 | FJ904410, FJ904409, FJ904447, FJ904407, FJ904408, FJ904403, FJ904394, FJ904404, FJ904419, FJ904395, FJ904398, FJ904414, FJ904405, FJ904396, FJ904435, FJ904430, FJ904437, FJ904417, FJ904416, FJ904413, FJ904406, FJ904442, FJ904439, FJ904436, FJ904440, FJ904441, FJ904400, KM606751, KM606750 |
| D8 | FN594769, FN594770, FN594771 |
| D9 | JN664919, JN664920, JN664921, JN664922, JN664942 |



# Chapter 5



## Genetic diversity of hepatitis C virus in Ethiopia

Gadissa Bedada Hundie, V. Stalin Raj, Daniel GebreMichael, Suzan D. Pas, Bart L. Haagmans

*PLOS ONE. 2017 Jun 1;12(6):e0179064.*

## ABSTRACT

Hepatitis C virus (HCV) is genetically highly divergent and classified in seven major genotypes and approximately hundred subtypes. These genotypes/subtypes have different geographic distribution and response to antiviral therapy. In Ethiopia, however, little is known about their molecular epidemiology and genetic diversity. The aim of this study was to investigate the distribution and genetic diversity of HCV genotypes/subtypes in Ethiopia, using 49 HCV RNA positive samples. HCV genotypes and subtypes were determined based on the sequences of the core and the nonstructural protein 5B (NS5B) genomic regions. Phylogenetic analysis revealed that the predominant was genotype 4 (77.6%) followed by 2 (12.2%), 1 (8.2%), and 5 (2.0%). Seven subtypes were identified (1b, 1c, 2c, 4d, 4l, 4r and 4v), with 4d (34.7%), 4r (34.7%) and 2c (12.2%) as the most frequent subtypes. Consistent with the presence of these subtypes was the identification of a potential recombinant virus. One strain was typed as genotype 2c in the NS5B region sequence and genotype 4d in the core region. In conclusion, genotype 4 HCV, subtypes 4d and 4r, are the most prevalent in Ethiopia. This genotype is considered to be difficult to treat, thus, our finding has an important impact on the development of treatment strategies and patient management in Ethiopia.

## INTRODUCTION

Hepatitis C virus (HCV) is an important human pathogen that causes substantial morbidity and mortality worldwide. The most recent study estimated that more than 185 million people are chronically infected with HCV and 3–4 million new infections occur each year [75]. HCV is one of the leading causes of end-stage liver disease, cirrhosis and hepatocellular carcinoma [188], resulting in over 700,000 deaths annually [6]. In addition, HCV is the most common cause of death in HIV-positive patients on highly active antiretroviral therapy [80]. Consequently, the burden of HCV-related morbidity and mortality will likely continue to increase over the next twenty-five years due to the existing pool of chronic HCV infections in low-income countries [189].

HCV is an enveloped, positive-sense, single-strand RNA virus belonging to the genus *Hepacivirus* in the family *Flaviviridae*. The viral genome is ~9.6 kb in length and contains a single open reading frame that encodes a polyprotein of about 3,000 amino acids [190]. HCV shows extreme genetic diversity and classified into seven main genotypes (1–7), further classified into 67 confirmed, 20 provisional, and 21 unassigned subtypes [94]. The HCV genotypes differ from each other by approximately 30-35%, and the subtypes differ from each other by at least 15% over complete genome [94,95]. Different HCV genotypes and subtypes display distinct geographic distribution patterns and levels of genetic diversity [190]. Overall, genotypes 1–3 are globally distributed, causing the majority of cases in the world, whereas genotypes 4–7 are more geographically restricted [95]. Genotype 4 is found mainly in Central Africa and the Middle East [191,192], genotype 5 in South Africa [96], and genotype 6 in Southeast Asia [107,193]. The recently identified genotype 7 has been isolated from a Congolese immigrant in Canada [103]. This genetic diversity and geographic variation of HCV have an important impact on disease epidemiology and clinical practice because it is one of the most important predictors of response to anti-viral therapy [96,190]. Therefore, for the development of treatment strategies and patient management, an in-depth understanding of the prevalent genotype and subtypes in different geographic regions, including Ethiopia is highly important.

Ethiopia is a large, geographically diverse nation, home to over 100 million inhabitants. Available data shows that its adult population has a low to moderate prevalence of HCV infections (0.52 to 5.8%) [102,194,195]. In Africa, distribution of HCV genotype exhibits two epidemiological patterns: one characterized by high genetic diversity, distributed in West Africa, with genotypes 1 and 2 [106,107,196], and the other in Central and Norther Africa, with genotype 4 [197]. However, little is known about the HCV genotype distribution in Easter Africa in general and in

Ethiopia in particular. The only study reporting on HCV genotypes in Ethiopia was limited to few patients attending a voluntary counseling and testing center in Addis Ababa. A total of 18 HCV RNA positive samples were analysed and a dominance of genotype 4 (50%, nine cases) followed by 2 (6 cases), 5 (2 cases) and 1 (1 case) was found [198]. However, no country-wide studies have been published and information on the HCV molecular diversity throughout the country is lacking. The aim of the present study was, therefore, to investigate the prevalence and genetic diversity of HCV genotypes and subtypes in different geographic regions of Ethiopia.

## MATERIALS AND METHODS

### Study Area and Population

This study was conducted using serum samples collected between March 2013 and April 2014 from voluntary healthy blood donors at blood bank centers in five geographic regions: Addis Ababa (the capital, central) Adama (Oromia, central-east) Gondar (Amhara, north-west) Mekelle (Tigray, north) and Jimma (Oromia, south-west). A total of 56 885 sera were collected, of which 294 were tested anti-HCV antibody positive in Ethiopia [195]. Of these, 98 sera were stored at −80°C until use and shipped to the Netherlands for further molecular study. The study was approved by the National Research Ethics Review Committee at the Federal Ministry of Science and Technology, Addis Ababa, Ethiopia. Informed written consent was given by all participants.

### Serological assays

The serological markers for HBsAg and anti-HCV antibodies were tested using commercially available enzyme-linked immunosorbent assay (ELISA) kits (DIALB Diagnostics GmbH, Vienna, Austria). Serum was also screened for anti-HIV-1 and -HIV-2 antibodies by ELISA (Vironostika HIV Ag/Ab, Bio-Merieux, Boxtel, The Netherlands) and anti-*Treponema Pallidum* antibodies by Rapid Plasma Reagin Test (RPR) (DIALB Diagnostics GmbH, Vienna, Austria). All tests were carried out in accordance with the manufacturers' instructions. Anti-HCV antibody-positive serum samples from donors co-infected either with HBV or HIV were excluded from the present study. As the study subjects are healthy blood donors, clinical chemistry parameters like AST and ALT are not available.

### HCV RNA extraction, RT-PCR amplification, and sequencing

HCV RNA was extracted from 140µL serum using QIAamp Viral RNA mini kit (QIAGEN, Hilden, Germany) or High Pure Viral Nucleic Acid kit (Roche). Complementary DNA (cDNA) was synthesized using random Hexamer primers (Promega, Madison, WI, USA) and SuperScriptIII (SSIII) First-Strand Synthesis System (Invitrogen) according to the manufacturers' instruction. Briefly, the RT reaction mixture included 10µl of viral RNA, 1µl of random hexamers, 1µl of 40U/µl RNasin, 4µl of 5× First Strand Buffer, 2µl of 0.1M DTT, 1µl of dNTP mix (20 mM) and 1µl (200 U/µl) SSIII reverse transcriptase; in a final volume of 20µl. A nested PCR technique was used to amplify DNA fragments in the core and the nonstructural protein 5B (NS5B) genes. The first round of PCR was performed with 5µl cDNA in a total reaction mixture of 50µl containing 10µl of 5× HotStar HiFidelity PCR Buffer, 1µl HotStar HiFidelity DNA Polymerase (QIAGEN) and 1µl each forward and reverse outer primers (20 pmol/µl) described below. The same conditions were used in the second round PCR except 2µl of first round PCR product and 1µl each forward and reverse inner primers used here. The core gene fragment (287 -751 nt, according to the position of reference HCV isolate H77 GenBank accession number AF009606) was amplified by nested PCR with outer primers s17, 410 and inner primers 953, 954, 951, as described previously [199] with some modifications. The NS5B gene (8244 to 8656 nt, according to H77) was amplified by nested PCR with a combinations of primers Pr1, Pr2, Pr3, Pr4, 122, 1204, 123 and 1203, as described previously [199,200]. All PCR reactions were conducted with the following thermal profile 95°C for 5 min, then 40 cycles of 95°C for 1 min, 48°C/50°C for 1 min and 72°C for 1 min, with a final elongation step at 72°C for 10 min. The amplified products were gel-purified by MinElute Gel Extraction kit (QIAGEN, Hilden, Germany) and bi-directionally sequenced using Big Dye Terminator v3.1 kits and an ABI prism 3130xl auto-sequencer (Applied Biosystems, Foster City, California, USA).

### HCV genotyping and phylogenetic analyses

Sequences were assembled and edited with SeqMan Pro (DNASTAR Lasergene 10). Comparative analyses were performed using the CLUSTAL W multiple sequence alignment program, MEGA 6 software [143]. HCV genotypes and subtypes were determined by aligning to the reference sequences retrieved from GenBank or the Los Alamos (http://hcv.lanl.gov/content/sequence/NEWALIGN/align.html) databases and geno2pheno software (http://www.geno2pheno.org/). Bayesian inference (BI) analyses were conducted using MrBayes v3.2 [201]. For the construction of the Kimura two-parameter model and neighbor-joining method. Genetic distances were estimated using the Bayesian phylogeny, we selected the

best model of nucleotide substitution based on the lowest Bayesian information criterion (BIC) and Akaike information criterion (AIC) score using TOPALi v2.5 [202]. The best model selected was a general time reversible (GTR) with a gamma-distributed rate variation across sites and a proportion of invariant sites (GTR+I+G). The Bayesian phylogenetic tree was constructed in MrBayes v3.2 with GTR+I+G model using 4 chain simultaneous run for 10 million generations, a sampling frequency of 100, and a 50% burn-in. The obtained trees were visualized in FigTree v1.4.2 [ http://tree.bio.ed.ac.uk/software]. Finally, mutation in the NS5B region, codons 230-347, were assessed by geno2pheno software (http://www.geno2pheno.org/).

### Nucleotide sequence accession numbers

The nucleotide sequences of HCV Core and NS5B partial regions determined in this study have been deposited in the GenBank sequence database, under the accession KY627917-KY628007.

## RESULTS

### Demographic characteristics

A total of 98 anti-HCV positive healthy blood donors sera (51 from Addis Ababa, 13 from Adama, 15 from Gondar, 3 from Jimma and 16 from Mekelle) were analyzed in the present study. Of these, 72 (73.5%) were males and 26 (26.5%) were females. The mean age was 31. 8 years (range: 18-60 years) (Table 1).

### HCV genotypes and phylogenetic analysis

HCV RNA was successfully amplified in 49 out of the 98 anti-HCV-positive samples. In 49 samples, even after repeated nested PCR with a combination of different primers, no genome fragment could be amplified, which could be due to storage conditions, a high rate of false positive anti-HCV antibody results or HCV clearance. Among the PCR positive samples, 42 could be genotyped both in the NS5B and core regions, 4 in the NS5B region and 3 in the core region only. Hence, 46 and 45 samples were successfully sequenced in the NS5B and core regions, respectively (Table 2).

**Table 1. Demographic characteristics of blood donors enrolled in the study and PCR result (HCV-RNA)**

| Variables | Regions | | | | | Total |
|---|---|---|---|---|---|---|
| | Addis Ababa | Adama | Gondar | Jimma | Mekelle | |
| **Sex** | | | | | | |
| **Male** | 35 (68.6%) | 11 (84.6%) | 13 (86.7%) | 2 (66.7) | 11 (68.8%) | 72 (73.5%) |
| **Female** | 16 (31.4%) | 2 (15.4) | 2 (13.3%) | 1 (33.3) | 5 (31.2%) | 26 (26.5%) |
| **Age, years** | | | | | | |
| **≤ 20** | 7 (13.7%) | 0 | 2 (13.3%) | 1 (33.3%) | 1 (6.2%) | 11 (11.2%) |
| **21-30** | 14 (27.5%) | 8 (61.5%) | 6 (40.0%) | 0 | 9 (56.3%) | 37 (37.6%) |
| **31-40** | 20 (39.2%) | 3 (23.1%) | 3 (20.0%) | 1 (33.3%) | 5 (31.3%) | 32 (32.7%) |
| **41-50** | 8 (15.7%) | 1 (7.7%) | 4 (26.7%) | 1 (33.3%) | 1 (6.2%) | 15 (15.3%) |
| **> 50** | 2 (3.9%) | 1 (7.7%) | 0 | 0 | 0 | 3 (3.1%) |
| **PCR** | | | | | | |
| **Positive** | 33 (64.7%) | 3 (23.1%) | 3 (20.0%) | 2 (66.7%) | 8 (50.0%) | 49 (50.0%) |
| **Negative** | 18 (35.3%) | 10 (76.9%) | 12 (80.0%) | 1 (33.3%) | 8 (50.0%) | 49 (50.0%) |



A phylogenetic tree of the NS5B region was constructed of 46 Ethiopian sequences obtained in this work and reference sequences retrieved from GenBank. Four HCV genotypes (1, 2, 4 and 5) with seven subtypes (1b, 1c, 2c, 4d, 4l, 4r and 4v) were identified. As shown in Figure 1, the predominant was genotype 4, found in 35 strains (76.1%), followed by 2 ($n$ = 6, 13%), 1 ($n$ = 4, 8.7%), and 5 ($n$ = 1, 2.2%). Of the 35 genotype 4 strains, 16 strains belong to subtype 4r, 15 strains to 4d, 3 strains to 4l and one isolate to 4v (Table 2). All genotype 2 strains belong to subtype 2c and 75% of genotype 1 strains to subtype 1b. All the genotypes and subtypes were confirmed by geno2pheno software (http://www.geno2pheno.org/). A Bayesian inference phylogenetic tree for the core region was also constructed using the 45 sequences (42 from strains used in NS5B and 3 additional strains) and reference sequences from the GenBank. The most commonly detected genotype was genotype 4 ($n$ = 38, 77.6%), with predominant subtypes 4d ($n$ = 18, 36.7%) and 4r ($n$ = 17, 34.7%), followed by 4l ($n$ = 2, 4.1%) and 4v ($n$ = 1, 2.0%). Genotype 1 ($n$ = 3, 6.1%) comprised the subtypes 1b ($n$ = 2, 4.1%) and 1c ($n$ = 1, 2.0%). Genotype 2c was represented by 3 strains (6.1%) while one strain was classified into genotype 5a (Figure 2). Genotyping was consistent when analyzed by either NS5B or core regions, except one strain (MK32ETH) which was 2c in the NS5B region but 4d in the core region, suggesting the identification of a novel recombinant HCV (Figure 1 and 2). Further full genome sequencing of this sample however failed.

**Table 2. Concordance of HCV genotypes/subtypes between Core and NS5B regions.**

| Isolates | Genotypes/subtypes | |
|---|---|---|
| | Core region | NS5B region |
| **AA328ETH** | 1b | 1b |
| **AA331ETH** | 1c | 1c |
| **AA332ETH** | 4d | 4d |
| **AA333ETH** | 4r | 4r |
| **AA345ETH** | 4d | 4d |
| **AA346ETH** | 4d | 4d |
| **AA347ETH** | 4d | 4d |
| **AA348ETH** | 4r | 4r |
| **AA350ETH** | 4d | 4d |
| **AA351ETH** | 4r | 4r |
| **AA352ETH** | 4d | 4d |
| **AA393ETH** | 4l | 4l |
| **AA394ETH** | 4r | 4r |
| **BL68ETH** | 4v | 4v |
| **BL78ETH** | NA[b] | 4l |
| **JM19ETH** | 4d | 4d |
| **MK07ETH** | 4r | 4r |
| **MK08ETH** | 4r | 4r |
| **MK32ETH**[a] | 4d | 2c |
| **MK144ETH** | 5a | 5a |
| **AA07ETH** | 4d | 4d |
| **AA44ETH** | 4r | 4r |
| **AA58ETH** | 2c | 2c |
| **AA108ETH** | 4d | 4d |
| **AA19ETH** | 4r | 4r |
| **AA148ETH** | 4r | 4r |
| **AA55ETH** | 4r | 4r |
| **AA57ETH** | 4l | 4l |
| **AA67ETH** | 4d | 4d |
| **AA71ETH** | NA | 1b |
| **AA214ETH** | 4d | 4d |
| **AA230ETH** | 4r | 4r |
| **AA232ETH** | 4d | 4d |
| **AA237ETH** | 4d | 4d |
| **AA174ETH** | 2c | 2c |

**Table 2. Concordance of HCV genotypes/subtypes between Core and NS5B regions. (continued)**

| Isolates | Genotypes/subtypes | |
|---|---|---|
| | Core region | NS5B region |
| **AA175ETH** | 4r | 4r |
| **AA188ETH** | 4r | 4r |
| **AA197ETH** | 4r | 4r |
| **AA239ETH** | 4d | 4d |
| **AA323ETH** | 4d | 4d |
| **GD07ETH** | 1b | 1b |
| **GD45ETH** | 2c | 2c |
| **GD93ETH** | NA | 2c |
| **JM05ETH** | 4d | 4d |
| **BL94ETH** | NA | 2c |
| **MK116ETH** | 4r | 4r |
| **MK58ETH** | 4r | NA |
| **MK159ETH** | 4d | NA |
| **MK114ETH** | 4d | NA |

[a]Potential recombinant virus

[b]NA; not available (the genomic region couldn't be amplified for the corresponding isolate)

Table 3 shows HCV genotype distribution in Ethiopia by age and gender. All genotype 2 strains were observed among males. Overall, no significant differences were observed between HCV genotypes with respect to donor's age and gender.

**Table 3. HCV genotypes distribution by age and gender № (%).**

| Variables | Genotype 1 № (%) | Genotype 2 № (%) | Genotype 4 № (%) | Genotype 5 № (%) | Total № (%) |
|---|---|---|---|---|---|
| **Age group** | | | | | |
| < 20 | 0 | 0 | 3 (7.9) | 0 | 3 (6.1) |
| 21-30 | 1 (25.0) | 1 (16.7) | 13 (34.2) | 1 (100) | 16 (32.7) |
| 31-40 | 2 (50.0) | 5 (83.3) | 14 (36.8) | 0 | 21 (42.9) |
| 41-50 | 1 (25.0) | 0 | 6 (15.8) | 0 | 7 (14.3) |
| >50 | 0 | 0 | 2 (5.3) | 0 | 2 (4.0) |
| **Gender** | | | | | |
| Male | 2 (50.0) | 6 (100) | 28 (73.7) | 0 | 36 (73.5) |
| Female | 2 (50.0) | 0 | 10 (26.3) | 1 (100) | 13 (26.5) |

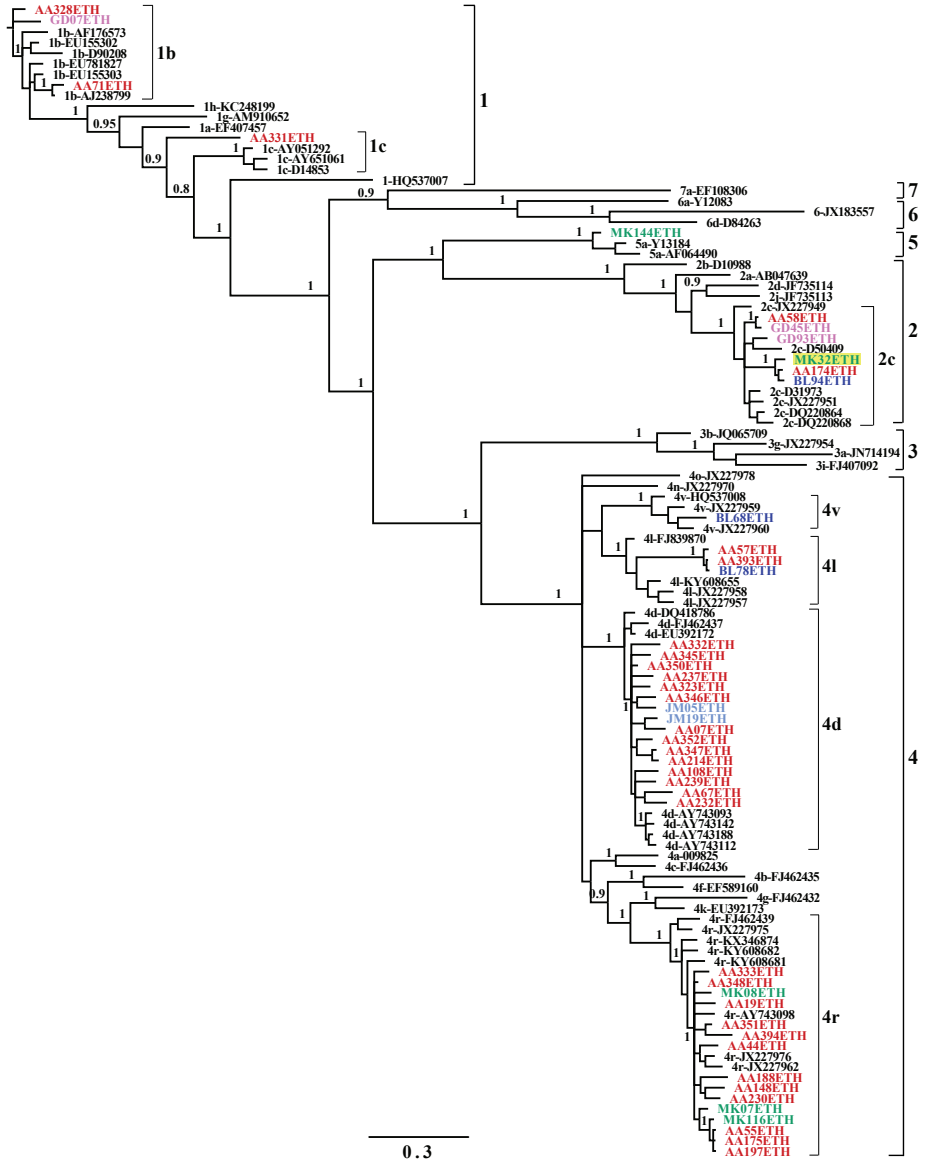



**Figure 1. Bayesian inference phylogenetic tree of the NS5B region.** A phylogenetic tree was constructed based on the NS5B region (356 nucleotides), corresponding to nucleotide numbering of 8288-8643 in the H77 genome, using the GTR+I+G model of evolution. Numbers on the branches are posterior probabilities from the Bayesian inference analysis. Reference sequences are labeled to the right of each branch in the order of subtype hyphen GenBank accession number. Ethiopian sequences are shown in color (color corresponding to their geographic origin: Red = Addis Ababa; Blue = Adama; Green = Mekelle; Pink = Gondar; Light blue = Jimma), a potential recombinant MK32ETH highlighted in yellow.

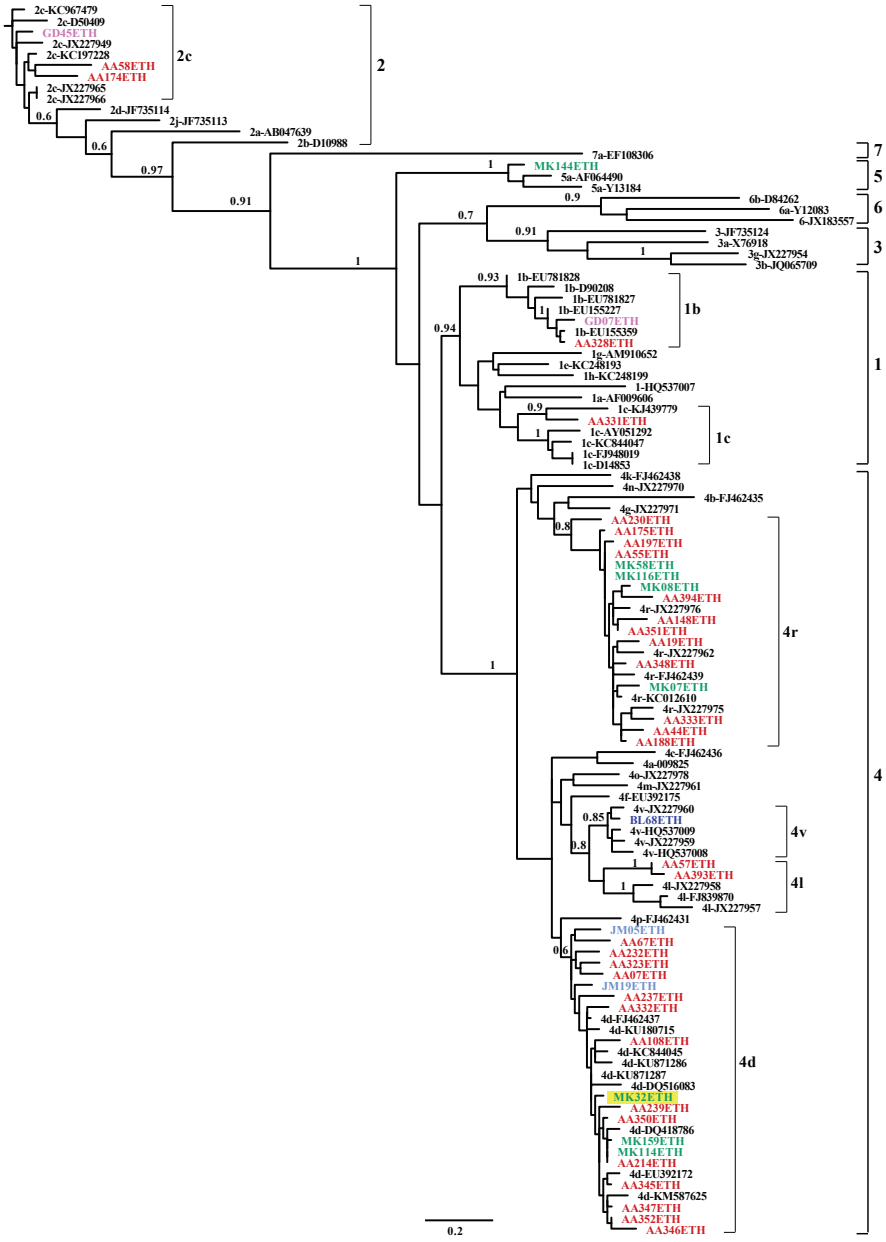



**Figure 2. Bayesian inference phylogenetic tree of the core region.** A phylogenetic tree was constructed based on the core region (418 nucleotides), corresponding to nucleotide numbering of 300-717 in the H77 genome, using the GTR+I+G model of evolution. Numbers on branches are posterior probabilities from the Bayesian inference analysis. Reference sequences are labeled to the right of each branch in order of subtype hyphen GenBank accession number. Ethiopian sequences are shown in color (color corresponding to their geographic origin: Red = Addis Ababa; Blue = Adama; Green = Mekelle; Pink = Gondar; Light blue = Jimma), a potential recombinant MK32ETH highlighted in yellow.

### Mutation in NS5B region

Mutations in the NS5B region spanning amino acid codons 230-347 were investigated using the geno2pheno software. We found that all the 46 Ethiopian sequences had mutations ranging from one to thirteen amino acid changes (S1 Table). Interestingly, the 316N variant associated with resistance to the direct acting antiviral Dasabuvir was detected in two subtypes 1b strains (AA328ETH and GD07ETH) (S1 Table).

## DISCUSSION

Knowledge of the HCV genotypes and subtypes has gained importance, because it plays a vital role in predicting the therapeutic and clinical outcome of the HCV infection [190,203]. Moreover, the global epidemiological data show that HCV genotypes and subtypes distributions vary in different geographic regions, even among regions of the same country [96]. However, there is very limited information on the molecular epidemiology and genetic diversity of HCV infections in Eastern Africa including Ethiopia. In this study we report the first nationwide molecular epidemiology and genetic diversity of HCV in Ethiopia, one of the geographically diverse and the second most populous country in Africa. We found a high level of HCV genetic diversity, with four major genotypes (1, 2, 4 and 5) and seven subtypes are circulating in the country.

Phylogenetic analysis revealed that the majority of Ethiopian HCV strains are belonging to genotype 4 (78%). This genotype is the most frequent cause of chronic hepatitis C in the Middle East, Egypt and Central Africa, accounting for more than 80% of HCV infections [204]. For instance, more than 90% of HCV infections in the Central African Republic, the Democratic Republic of Congo, Gabon and Egypt were attributed to genotype 4 [204,205]. In Eastern Africa, however, due to the paucity of data the prevalence of HCV genotype 4 is currently not well known [197]. In recent years, the epidemiology of genotype 4 has changed and this genotype has begun spreading beyond its strongholds in Africa and the Middle East to several Western countries, particularly in Europe, due to variations in population structure, immigration and injection drug use (IDU) [191,206]. For example, 10 to 24% of chronic HCV infections in southern Europe, particularly in France, Italy, Greece, and Spain and 11% of chronic HCV infections in The Netherlands were attributed to genotype 4 [191,207,208]. In addition, a high frequency of genotype 4 (42%) was found among Somali immigrants in Minnesota, USA [209]. Nowadays, a wide-ranging overview of genotype 4 epidemiology and diversity may prove useful for public health assessments not only in Africa but also outside

Africa, because it accounts for 20% of total global HCV infection [197,207,210]. In addition, genotype 4 is considered difficult to treat and has a poor sustained virological response rate of 43-63% to the standard pegylated IFN/ribavirin combination therapy [208], which is relatively higher than genotype 1 but lower than genotypes, 2, 3, 5, and 6 [210,211].

Overall, the presence of four different subtypes (4d, 4r, 4l, and 4v) of genotype 4 in Ethiopia is indicative of a greater genetic diversity compared to HCV genotype 4 viruses reported in the surrounding countries. In our study, subtypes 4d and 4r are the predominant subtypes, accounting for 89% of the HCV genotype 4 infections. These subtypes are of interest because subtype 4r is not commonly reported while 4d is mostly reported in Saudi Arabia and in the majority of European IDU population [191,204,206,212]. This result is thus different from studies in other African countries where subtypes 4a is the predominant subtype in Egypt [204,211], 4f in Cameroon [213], 4e in Gabon [192,205], 4k in the Central African Republic and Democratic Republic of the Congo [197,214], 4c in the Republic of Congo [215], and 4q/4v in Rwanda [197]. The high diversity and predominance of genotype 4 suggests that this genotype has been endemic for a longer time in Ethiopian population. A previous study also showed the dominance of this genotype in Ethiopia [198].

The risk factors for genotype 4 and its subtypes transmission are determined by the geographical distribution of this genotype. Intravenous drug use is the most common route of transmission for genotype 4 infection in Europe while unsafe medical practice cause most cases in endemic countries [211]. In Ethiopia, although we didn't assess the risk concerning genotype 4 in the present study, we suggest that sharing contaminated needles and razor blades during tattooing, body piercing, scarification, and circumcision may be the main mode of transmission, as these are common practices particularly in rural Ethiopia.

Genotype 2 was the second most prevalent HCV genotype in our study. HCV genotype 2 originated from Western Africa and disseminated to the globe through trans-Atlantic slave trade, colonial history and migrations [104,106,216]. In Africa, it is frequently prevalent in West Africa and its prevalence relative to other HCV genotypes declines from west to east [106,216]. Our result is in agreement with the aforementioned studies. HCV genotype 2 in West Africa exhibits high genetic diversity with several subtypes identified within this genotype [104,106]. The HCV genotype 2 from Ethiopia was found to be less diverse and belong to a single epidemic subtype 2c, suggesting that it might be introduced from western Africa to eastern Africa. This could be due to frequent travel of people from this geographical location to Ethiopia and vice versa.

Unlike hepatitis B virus [217], recombination between genotypes is not a common event in HCV. To date, ten HCV recombinants have been identified worldwide, of which only recombinant 2k/1b was detected multiple times and in different countries [94,99,100]. In the present study, we found that strain MK32ETH was characterized as genotype 2c in the NS5B gene (Figure 1) and genotype 4d in the core gene(Figure 2) as revealed by 3 different independent experiments. However, although we tried to obtain the complete genome as we did previously [218], using the Bull et al. method [219], as well as 454 deep sequencing, all methods were unsuccessful.

In conclusion, this study provides important data on HCV genotypes and subtypes in Ethiopia, where four main genotypes and seven subtypes were identified. HCV infection in Ethiopia is characterized by the predominance of HCV genotype 4 (4d and 4r), with high genetic diversity, suggesting that this genotype has been endemic for a long time in Ethiopia. However, further large-scale studies on the molecular epidemiology of HCV in Ethiopia are needed. As genotype 4 is difficult to treat, our findings have a major impact in developing treatment guidelines and patient management.

**Acknowledgments**

We thank our colleagues from the Ethiopian Blood Bank Centers (national and regional) for collecting the samples and data.

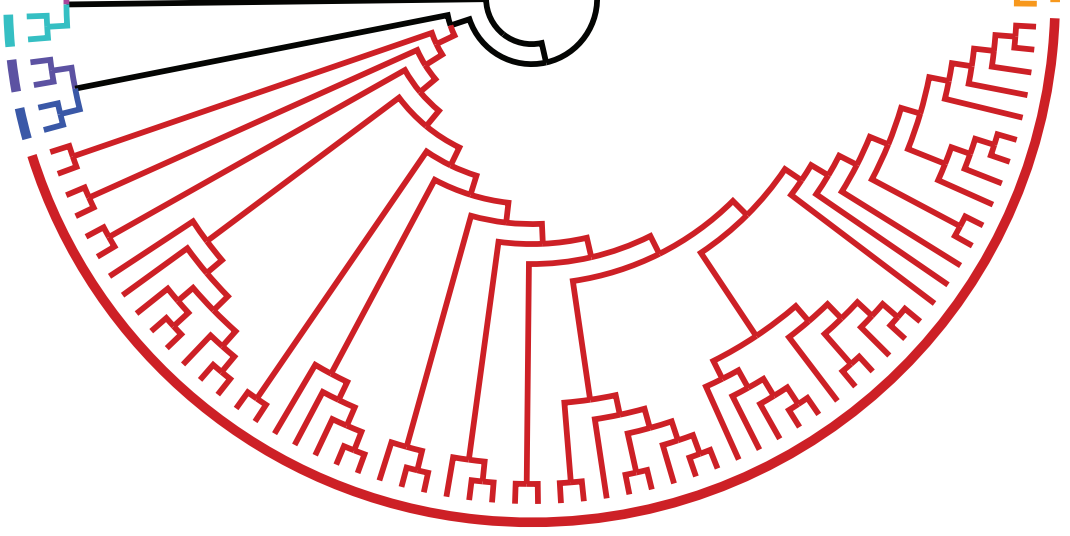
**Author Contributions**

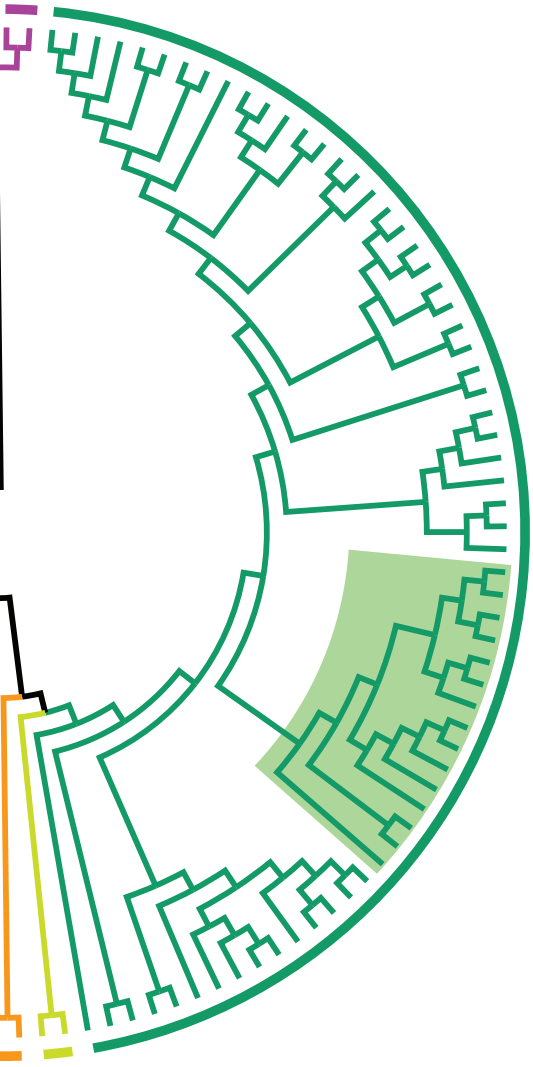
Conceptualization: GBH, BLH. Data curation: GBH, BLH, VSR. Formal analysis: GBH, BLH, VSR. Funding acquisition: GBH, BLH. Investigation: GBH, BLH, VSR. Methodology: GBH, BLH. Project administration: GBH, BLH, DG. Resources: SDP, DG. Supervision: BLH. Validation: GBH, BLH. Visualization: GBH, VSR. Writing – original draft: GBH. Writing – review & editing: GBH, VSR, SDP, BLH.

**S1 Table. Mutation in the NS5B region of HCV**

| Ethiopian Sequence | subtype | Mutation in NS5B region | Dasabuvir resistance-associated mutation | Sofosbuvir resistance-associated mutation |
|---|---|---|---|---|
| AA328ETH | 1b | C316N | 316N | none |
| AA331ETH | 1c | A238T, K254R, V262I, E272Q, L293I, T300A, S335A, A338G | none | none |
| AA332ETH | 4d | F285Y, S300T, I303T, D327N | NA* | none |
| AA333ETH | 4r | T254S, K270R, C316H, V321I | NA | none |
| AA345ETH | 4d | S254A, F285Y, S300N | NA | none |
| AA346ETH | 4d | S300T | NA | none |
| AA347ETH | 4d | A255S, S300NT, I303IT, D310DN, G328S | NA | none |
| AA348ETH | 4r | A252AV, C316H, V321I, E327AV | NA | none |
| AA350ETH | 4d | V252AV, S300T, I303IV, S347P | NA | none |
| AA351ETH | 4r | K270R, C316H, V321I, E327AE | NA | none |
| AA352ETH | 4d | S300T, I303T, D327E | NA | none |
| AA393ETH | 4l | E237G, E248D, A249T, K270R, F285L, M300T, G327D, G333T | NA | none |
| AA394ETH | 4r | K270R, C316H, V321I, E327A, A333E | NA | none |
| BL68ETH | 4v | D244N, K251R, T254D, V284T | NA | none |
| BL78ETH | 4l | E237G, E248D, A249T, K270R, F285L, M300T, G327D, G333T | NA | none |
| JM19ETH | 4d | S300T, I303T, D327G | NA | none |
| MK07ETH | 4r | C316H, V321I, E327A | NA | none |
| MK08ETH | 4r | C316H, V321I | NA | none |
| MK32ETH | 2c | T235N, N291S, K300R, V329A, R337K | NA | none |
| MK144ETH | 5a | A251V, K307R | NA | none |
| AA07ETH | 4d | D244N, I276T, S300T, I303T, D327G | NA | none |
| AA44ETH | 4r | C316H, V321I, E327A | NA | none |
| AA58ETH | 2c | K300R, V329A | NA | none |
| AA108ETH | 4d | S300T, I303V, R304K, A335S, Y346N | NA | none |
| AA19ETH | 4r | A252V, T303I, K309R, C316H, V321I, E327A | NA | none |
| AA148ETH | 4r | K270R, C316H, V321I, E327A, Y346N | NA | none |
| AA55ETH | 4r | C316H, V321I, E327A, Y346N | NA | none |
| AA57ETH | 4l | E237G, E248D, A249T, K270R, F285L, M300T, G327D, G333T, Y346N | NA | none |
| AA67ETH | 4d | K270R, I276T, F285Y, S300T, Y346N | NA | none |
| AA71ETH | 1b | S300A, D318H, D319H, A333E, R337W, V338A, Y346N | none | none |
| AA214ETH | 4d | S300T, I303T, G328S, Y346N | NA | none |
| AA230ETH | 4r | C316H, V321I, E327A, Y346N | NA | none |
| AA232ETH | 4d | I276T, S300T, Y346N | NA | none |
| AA237ETH | 4d | S300T, D310N, V322I, Y346N | NA | none |
| AA174ETH | 2c | T235N, T267R, K300R, V329A, R337K, Y346N | NA | none |
| AA175ETH | 4r | C316H, V321I, E327A, Y346N | NA | none |
| AA188ETH | 4r | 247T, K270KR, C316H, V321I, E327A, D332N, R334E, A335P, T340P, R345G, Y346N | NA | none |
| AA197ETH | 4r | C316H, V321I, E327A, Y346N | NA | none |
| AA239ETH | 4d | K270R, I276T, S300T, I303T, Y346N | NA | none |
| AA323ETH | 4d | S300T, I303T, D327G, M343I, Y346N | NA | none |
| GD07ETH | 1b | A252V, R254K, Q309R, C316N | 316N | none |
| GD45ETH | 2c | A299D, K300R, V309IV, V329A | NA | none |
| GD93ETH | 2c | L241Q | NA | none |
| JM05ETH | 4d | S269G, T286P, L293M, S300T | NA | none |
| BL94ETH | 2c | T235N, T267R, K300R, R337K | NA | none |
| MK116ETH | 4r | C316H, V321I, E327A | NA | none |

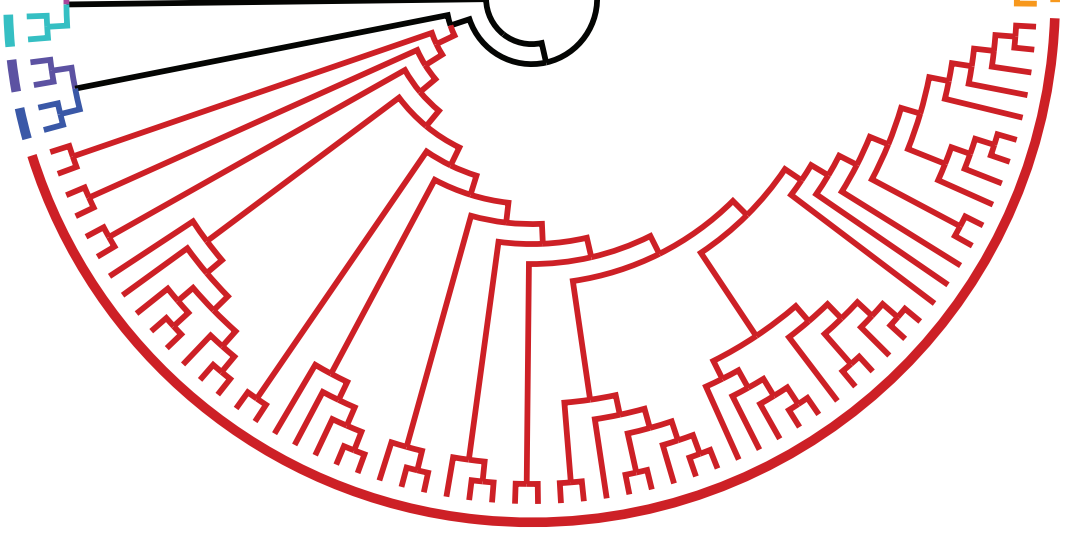
*NA; not available as Dasabuvir not licensed for this subtype



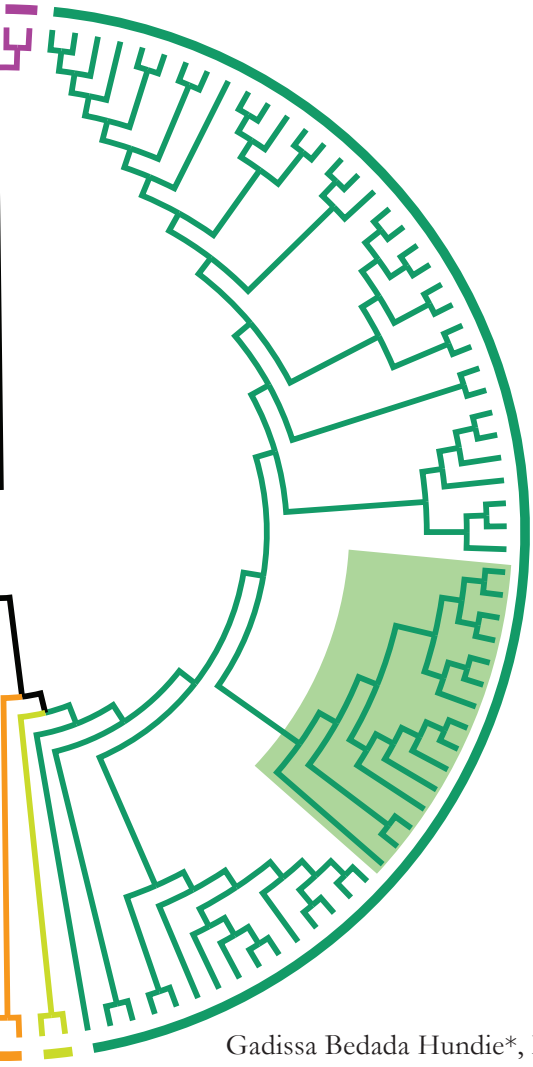


# Part II

## HBV and HCV quasispecies variants



# Chapter 6



## Hepatitis B virus quasispecies variants associated with HBeAg status and viral load as determined by full-length genome deep sequencing

Gadissa Bedada Hundie*, David van de Vijver*, Ad J.P. van der Meer, Margo J.H. van Campenhout, Steffen Wildum, Fernando Garcia-Alcalde, Alan J. Mueller-Breckenridge, Damien Chopy, Charles A. Boucher, Bart L. Haagmans

**These authors contributed equally to this study*

*(Manuscript in preparation)*

## ABSTRACT

Hepatitis B virus (HBV) quasispecies genetic diversity may influence the clinical course and treatment outcomes of chronic HBV infection. Therefore, we applied a full-genome deep-sequencing approach to identify potential association of baseline viral variants with HBeAg status and viral loads of HBV patients. A total of 186 plasma samples obtained from treatment-naive chronic hepatitis B patients (HBeAg-positive, $n$ = 99; HBeAg-negative, $n$ = 87) were analyzed by ultra-deep Illumina sequencing. Deep-sequencing analysis revealed a total of 4615 variants across the full-genome of 5 different genotypes (A to E). Although genotype A and B viruses showed a relatively higher range of variants, insignificant variant composition was seen between genotypes C and D, but variant diversity was significantly higher in viruses from HBeAg-negative compared to HBeAg-positive patients. We identified 30 HBV genome variants that can predict HBeAg status, with G1896A being the most important predictor, followed by G1899A, A1762T, G1764A, A1934T, and T1753C. At least 15 variants including G1896A, G1899A, A1762T, G1764A were found to be associated with HBeAg-negative status. We also identified 22 HBV variants that had a significant association ($p$ <0.05) with HBV DNA load; the majority of variants that are associated with low DNA loads also were found to be associated with HBeAg-negative status, indicating a correlation between HBeAg and viral load. In conclusion, we identified several viral variants that are significantly associated with HBeAg and viral load as well as variants that can predict the HBeAg status of the patient.

## INTRODUCTION

Hepatitis B virus (HBV) is a 3.2 kb relaxed circular, partially double-stranded DNA virus, with a unique genomic organization and replication mechanism. HBV replicates through an error-prone reverse transcriptase that lacks proofreading capacity. The lack of proofreading activity of the HBV reverse transcriptase combined with high viral replication rates ($10^{12}$ virions per day) results in the generation of quasispecies — defined as a population of genetically distinct but closely related variants [65,97,220]. This HBV quasispecies contains a spectrum of mutant variants that possess different fitness levels, which allow them for a rapid adaptation to a highly mutagenic environment when the virus is subject to selective pressures such as host immunity, vaccine or antiviral agents [221,222]. Such HBV variants have important implications in the pathogenesis, persistence, disease progression, clinical as well as therapeutic outcomes [223-226].

Many studies have disclosed some of these clinically important variants in different regions of the HBV genome [63,227-229]. Such variants include the G1896A (W28*) mutation in the precore (PC), A1762T, G1764A, and A1762T/G1764A double mutations in the basal core promoter (BCP) that have been shown to be associated with different clinical courses [65,226]. These mutations either completely abolish or decrease hepatitis B virus e antigen (HBeAg) production up to 70%, but enhance viral replication [65]. As noted, the presence of HBeAg in serum is a marker of high levels of virus replication while its seroconversion is an established therapeutic endpoint or clearance of the virus. Several studies revealed this correlation between HBeAg status and HBV viral load in chronic hepatitisB patients [230-233]. Besides PC and BCP variants, various mutant variants that are related to clinical characteristics have been described in the pre-surface (preS1/S2) [234,235], surface, polymerase and X regions [228] of the HBV genome. Hence, early detection of these quasispecies variants with low frequency is very important.

Nevertheless, most of the previous studies regarding the genetic diversity of HBV quasispecies variants were performed mainly using population-based sequencing which has low sensitivity for the detection of mutant variants with low frequencies (< 20%) [236]. In addition, those studies usually focused either on certain known specific HBV mutants or genomic regions, which therefore may have overlooked the potential influence of other novel viral variants. However, the next generation sequencing (NGS) enables the detection of viral variants with much greater sensitivity at frequencies as low as 0.5-1% in the quasispecies pool [221,237,238]. Nowadays, this method has been applied in HBV studies to analyze minor antiviral resistance mutations [239,240], quasispecies heterogeneity in reverse transcriptase and pre-S region [241,242], and mutations in BCP/PC regions [243]. Thus

far, these studies mostly focused on targeted HBV genes in specific clinical groups, still lacking full-genome investigation and evaluation of baseline quasispecies diversity in association with a clinical characteristic such as HBeAg status, viral DNA levels and alike. In light of this, it would be important to investigate a comprehensive full-gnome evaluation of baseline HBV quasispecies diversity using NGS to uncover novel viral variants that have a potential association with the patient's clinical characteristics.

The aim of this study was to investigate the association of baseline HBV quasispecies diversity with patient clinical characteristics such as HBeAg status and viral load in chronic hepatitis B patients using a full-length HBV genome Illumina deep-sequencing.

## MATERIALS AND METHODS

### Patients and samples

A total of 186 plasma samples consecutively collected from chronic HBV patients during the period from 1985 to 2012 and stored at the Erasmus University Medical Center in Rotterdam, the Netherlands were used in the present study. This study was conducted in accordance with the guidelines of the Declaration of Helsinki and the principles of Good Clinical Practice. The study was approved by the ethical review board of the Erasmus Medical Center, Rotterdam, The Netherlands. Due to retrospective nature of the study, written informed consent was not needed from each patient.

### Data acquisition

Data on demographics (sex, age, race, ethnicity, height, weight, route of HBV transmission) and clinical data were obtained from the electronic medical charts at the Erasmus University Medical Center in Rotterdam, the Netherlands. Data on liver biochemistry (alanine aminotransferase (ALT), aspartate aminotransferase (AST), gamma-glutyltransferase (γ-GT), bilirubin, albumin), and virological results (serum qualitative HBsAg, anti-HBs, HBeAg, anti-HBeAg, HBV viral load) at the baseline were obtained from the clinical laboratory and the department of Virology at the Erasmus Medical Center.

### DNA isolation and PCR amplification

HBV DNA was extracted from 200 µl plasma using the QIAamp MinElute Virus kit (Qiagen) or the Roche MagNA Pure LC instrument according to the manufacturer's instructions. Samples with a viral load >100000 IU/ml were

isolated with Roche MagNA Pure (LC) instrument, eluted in 50 µl elution buffer. Whereas samples with a viral load <100000 IU/ml or samples that remained negative with Roche MagNA Pure (LC) instrument were isolated with the QIAamp MinElute Virus kit (Qiagen, Hilden, Germany), eluted in 20 µl elution buffer. A well-characterized sample containing a known viral load was used as positive control. The full-length HBV genome was amplified either by single-step PCR or nested PCR method using the Expand high-fidelity PCR kit (Roche Molecular Systems) and combinations of different sets of primers developed by DDL Diagnostic Laboratory (Rijswijk, Netherlands). Briefly, the first PCR contains a primer set (forward and reverse primers) that amplifies the complete genome. If this PCR is negative, a nested PCR is deployed using semi-nested primers to amplify the entire genome. If this PCR was also negative then 2 partial PCRs were used to amplify the genome in two parts using sets of 2 outer and 2 nested primers. The generated amplimers are then fragmented for the ultra-deep sequence (UDS) analysis using Illumina technology.

**Deep sequencing and data analysis**

The deep sequencing of HBV samples was conducted by DDL Diagnostic Laboratory (Rijswijk, Netherlands). The PCR products that showed a clear band of the expected size after gel-electrophoresis were purified and cleaned up by Agencourt AMPure XP PCR purification system (Beckman Coulter) to remove primer-dimers and small nonspecific PCR products. The purified PCR products were quantified using the Quant-iT PicoGreen dsDNA kit (Life Technologies), followed by dilution and pooling of targets. The diluted PCR products (DNA strands) were fragmented and tagged using the Nextera XT sample preparation kit (Illumina, San Diego, California, USA) according to the manufacturer's instructions. Index primers were added by limited cycle PCR using the Nextera XT Index kit, and a beads-based normalization of each library was performed using beads with maximum binding capacity (Nextera XT sample preparation kit) according to the manufacturer's instructions to ensure equal library representation in the pooled sample. Multiplexed paired-end sequencing was performed on the Illumina MiSeq platform using the MiSeq v2 sequencing kit with 300 cycles. The outputs generated as demultiplexed FASTQ reads were quality trimmed and mapped to a genotype-specific reference sequence (GenBank accession numbers X02763, AB219428, GQ924620, AF121240 and AB106564, of genotype A to E respectively) in DDL's Athena pipeline for Quality Based Variant Detection. Sequences were trimmed to remove nucleotides with a quality score (Q) lower than 30. The alignment and trimming procedure were performed using Bowtie2 v2.2.2 (http://bowtie-bio.sourceforge.net) with the script written in the Python

programming language (Python 2.6.6). The minimum read length was ≥ 50 nucleotides and the minimum coverage was 5,000 sequencing reads. The quality-based variant detection is performed with DDL's Athena pipeline proprietary software (DDL Diagnostic Laboratory) which was specifically designed for detection of mutations within HBV populations. The threshold for detection of the variants by deep sequencing was set at 1% and the variant detection cut off coverage at ≥ 500.

### Statistical analysis

Statistical analyses were performed using the R program (version 3.4.1). All analyses conducted were based on variant calls that passed quality control criteria. Linear regression was used to assess the association of quasispecies variants with HBeAg status and viral load of the patient. Statistical significance was adjusted for multiple hypotheses testing using the Benjamini-Hochberg method using a false discovery rate of 0.05. Random Forest analysis was used to predict HBeAg status. Genetic diversity was measured using Shannon entropy [244].

## RESULTS

### Characteristics of the study cohort

The baseline characteristics of the 186 chronic hepatitis B patients used in this study are summarized in Table 1. The patients included 128 males and 58 females, with a mean age of 35 ± 13.5 years (mean ± sd). Eighty-seven of the patients were HBeAg negative (HBeAg-) and 99 were HBeAg positive (HBeAg+), with a mean serum HBV DNA level of 6.7 log IU/ml (range, 3.5 log IU/ml to 9.85 log IU/ml). In total, 186 HBV samples were analyzed for viral quasispecies diversity, of these 58 were of genotype A, 19 of genotype B, 44 of genotype C, 63 of genotype D and 2 of genotype E (Table 1).

### HBV quasispecies variants diversity

The HBV full-genome sequences were analyzed for quasispecies variants diversity. A total of 4,615 single nucleotide variants were found across the full-length HBV genome of all sequences. These variants were present at a threshold of ≥1% and found at least in one patient. There is sufficient evidence for differences in variant diversity between genotypes (A to D) (Figure 1). Genotype A samples had on average 171 variants; the majority of variants appeared in fewer than 10 patients with most variants (n = 987) only found once across all patients. In contrast, samples consisting of genotype B had on average 207 variants. There was a trend

($p$ =0.054) towards a significant difference between genotype A and B in the mean number of variants per patient. There were a similar number of variants found in each patient for genotype C (average 153 variants) and genotype D (average 132 variants). Genotype D had significantly fewer variants per patient than either genotype A or B. Only n = 208 variants were found to be shared across genotypes A-D. For all genotypes, HBV variant diversity across the genome was higher in HBeAg negative patients than HBeAg positive patients (Figure 2).

**Table 1. Baseline characteristics of Patients**

| Characteristics | Value, number (%) |
|---|---|
| Number of patients | 186 |
| Male | 128 (68.8%) |
| Female | 58 (31.2%) |
| Age, mean (±sd) | 35 (13.5) |
| Route of transmission | |
| Blood transfusion | 1 |
| Sexual contact | 15 |
| Vertical transmission | 161 |
| Unknown | 5 |
| Parental | 3 |
| Needles, other | 1 |
| HBV genotype | |
| A | 58 |
| B | 19 |
| C | 44 |
| D | 63 |
| E | 2 |
| HBeAg (+) | 99 (53.2%) |
| HBeAg (-) | 87 (46.8%) |
| $Log_{10}$ HBV DNA load IU/ml, mean (±sd) | 6.7 (2.0) |

Table 2 summarizes the 20 most frequent found variants at the nucleotide level and their ranks corresponding to a number of variants exceeding 10% or 50%. The BCP variants G1764A and A1762T and PC stop codon variant G1896A are the most frequently observed ones. Similarly, the most frequently observed variants at the amino acid level were analyzed across the different regions of the HBV genome (Table 3). Almost 50% of the variants were seen in the polymerase region. The HBV variants harboring the W28* (G1896A) precore stop codon, K130M (A1762T), and V131I (G1764A) were among the most frequently observed variants.

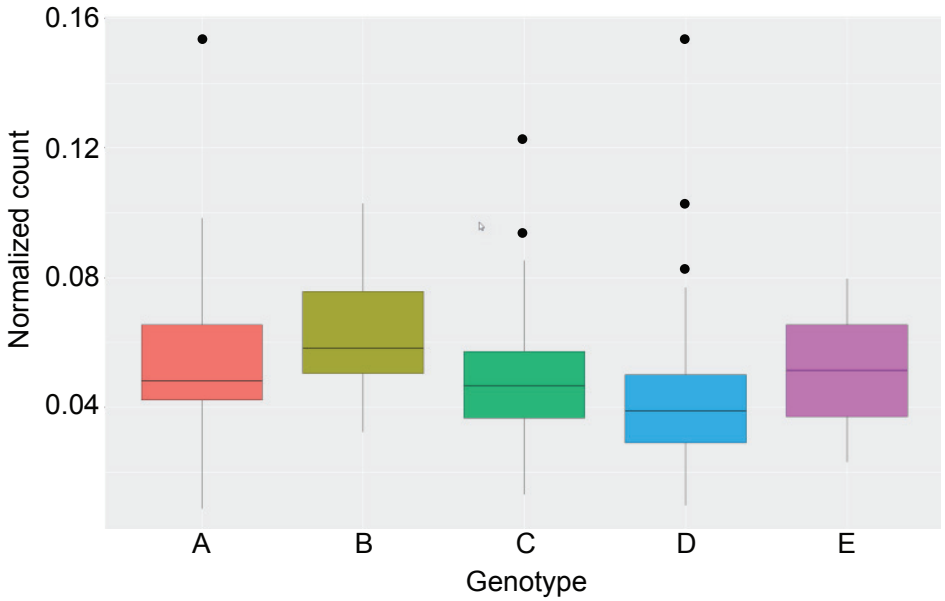



**Figure 1. Normalized Shannon entropy variant diversity count per HBV genome position according to genotype.** The box plot shows median values, upper and lower quartiles and the largest and smallest observations.

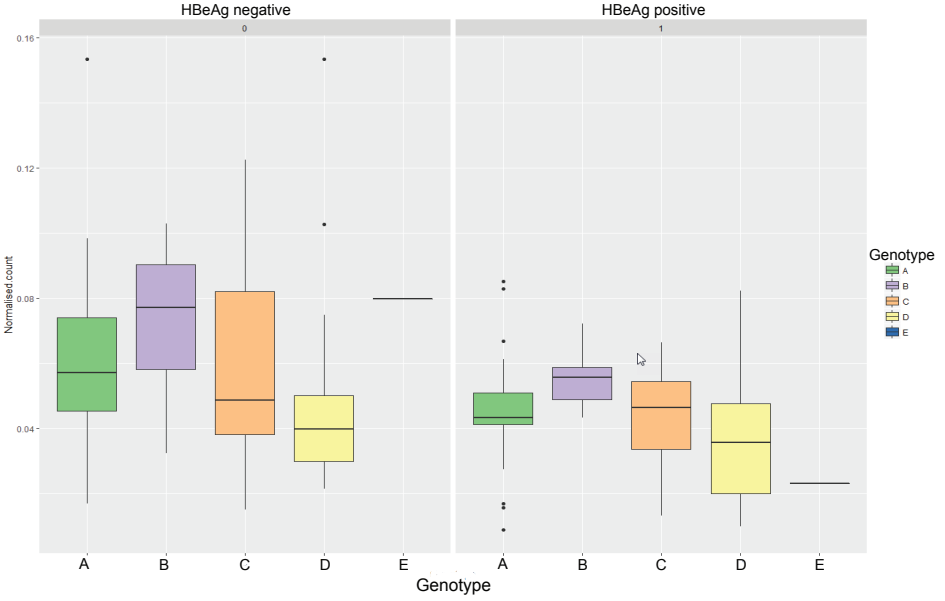



**Figure 2. Normalized Shannon entropy variant diversity count per HBV genome position according to HBeAg status.** The box plot shows median values, upper and lower quartiles and the largest and smallest observations.

### Prediction of HBeAg status using different HBV genome variants

Predictive analysis using machine-learning methods allows fast data processing and real-time predictions. In this study, we used the Random Forest model to predict the HBeAg status from HBV quasispecies variants. Using this model, we 30 identified variants that can predict the HBeAg status with an accuracy of 92% (Figure 3). The majority (66.7%) of these variants were located in the basal core promoter/core region followed by surface region (23.3%). G1896A is the most important predictor, followed by G1899A, A1762T, G1764A, A1934T and T1753C.

**Table 2. The 20 most important variants at nucleotide levels and their ranks based on number of variants exceeding 10% or 50%.**

| mutation | Variants >10%, No. (rank) | Variants >50%, No. (rank) |
|---|---|---|
| G1764A | 83 (1) | 75 (2) |
| A2011G | 82 (2) | 81 (1) |
| C2570T | 76 (3) | 75 (2) |
| A1762T | 74 (4) | 66 (5) |
| A2095G | 69 (5) | 68 (4) |
| G1896A | 67 (6) | 54 (31) |
| A562C | 63 (7) | 63 (6) |
| T843C | 63 (7) | 63 (6) |
| C2540T | 63 (7) | 62 (8) |
| A1173G | 62 (10) | 61 (11) |
| T2009C | 62 (10) | 62 (8) |
| T2539C | 62 (10) | 62 (8) |
| G842A | 61 (13) | 61 (11) |
| G2439A | 61 (13) | 61 (11) |
| C2534A | 61 (13) | 60 (14) |
| T3099C | 61 (13) | 59 (16) |
| G2797A | 60 (17) | 60 (14) |
| A1026T | 59 (18) | 59 (16) |
| C1239T | 59 (18) | 59 (16) |
| C1306A | 59 (18) | 57 (20) |

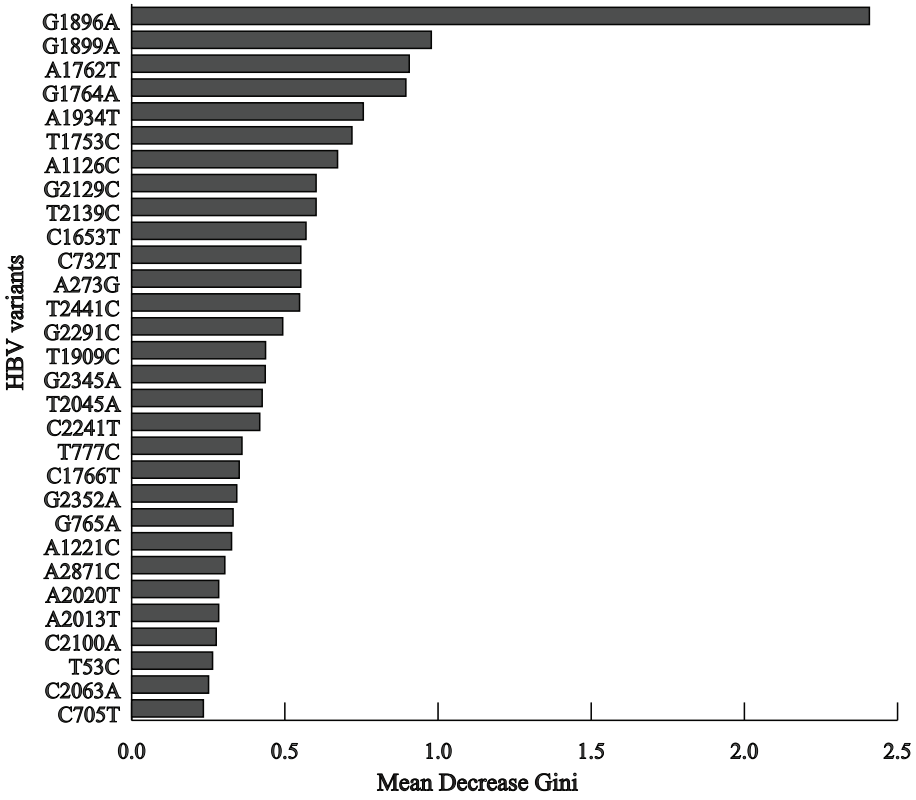



**Figure 3. Relative importance of HBV variants to prediction of HBeAg status.** The plot indicates each variants contribution to the mean decrease in the Gini index.

**Table 3. The 20 most important variants at amino acid level and their ranks based on number of variants exceeding 10% or 50%.**

| *Region* | *Substitution* | *>10%, number (rank)* | *>50%, number (rank)* |
|---|---|---|---|
| Pre core | W28* | 69 **(1)** | 57 (13) |
| X | V131I | 66 **(2)** | 57 (13) |
| Polymerase | F78S | 61 **(3)** | 61 **(1)** |
| Polymerase | M480L | 61 **(3)** | 61 **(1)** |
| Core | G180E | 60 (5) | 60 **(3)** |
| Polymerase | D45N | 60 (5) | 60 **(3)** |
| Polymerase | H36N | 59 (7) | 59 **(5)** |
| Polymerase | R164H | 59 (7) | 59 **(5)** |
| Polymerase | S240G | 59 (7) | 59 **(5)** |
| Polymerase | K248R | 59 (7) | 59 **(5)** |

**Table 3. The 20 most important variants at amino acid level and their ranks based on number of variants exceeding 10% or 50%. (continued)**

| *Region* | *Substitution* | *>10%, number (rank)* | *>50%, number (rank)* |
|---|---|---|---|
| Pre S1 | R66G | 59 (7) | 59 **(5)** |
| Core | V93M | 58 (12) | 58 (10) |
| Core | P171Q | 58 (12) | 58 (10) |
| Polymerase | F265L | 58 (14) | 56 (15) |
| Pre S1 | D51H | 58 (14) | 58 (10) |
| X | K130M | 58 (14) | 51 (26) |
| Polymerase | N312S | 57 (17) | 54 (19) |
| PS2 | T11A | 57 (17) | 54 (19) |
| Polymerase | R230H | 56 (19) | 56 (15) |
| Polymerase | S402T | 56 (19) | 56 (15) |
| Pre S1 | V48I | 56 (19) | 56 (15) |

## Association of quasispecies variants with patients HBeAg status

Linear regression was used to analyze the association with the proportion variation found for a particular variant and the HBeAg status in the presence of multiple hypothesis testing. As depicted in Table 4 and Table S1 (not shown), the coefficient (slope) shows the gradient of the linear regression, and mean is the average of all patients for a particular nucleotide substitution. A positive value for the slope means that the mean variation in HBeAg+ is greater than those with HBeAg- while a negative slope means that the variation in HBeAg- is greater than HBeAg+. Based on this analysis, we found that fifteen mutations including the PC (G1896A, G1899A) and BCP (A1762T, G1764A, T1753C, C1653T) variants are significantly associated with HBeAg- status (Table 4). Similarly, seventeen variants were found to be associated with HBeAg+ (Table S1, not shown). The indicated results only include mutations for which the mean variation across all patients was at least 10%.

**Table 4. Linear regression analysis of variants associated with HBeAg- in the presence of multiple hypothesis testing. Slope shows the gradient of the linear regression (with a gradient <-0.2), mean is the average variation of all patients for a particular nucleotide substitution.**

| Mutation | Coefficient | Mean | *P*-value |
|---|---|---|---|
| G1896A* | -0.685 | 0.304 | <0.001 |
| A1934T* | -0.622 | 0.147 | <0.001 |
| G1899A* | -0.575 | 0.156 | <0.001 |
| G2129C* | -0.539 | 0.109 | <0.001 |
| T1753C* | -0.483 | 0.131 | <0.001 |
| T2139C* | -0.438 | 0.165 | <0.001 |
| C1653T* | -0.437 | 0.123 | <0.001 |
| G1764A* | -0.374 | 0.411 | <0.001 |
| G774A* | -0.358 | 0.176 | <0.001 |
| C2245T* | -0.309 | 0.102 | 0.012 |
| A1762T* | -0.285 | 0.359 | <0.001 |
| T2660G | -0.283 | 0.101 | 0.029 |
| T147C* | -0.381 | 0.131 | 0.011 |
| A111C* | -0.269 | 0.155 | 0.009 |
| G1727A* | -0.249 | 0.198 | 0.009 |
| T1912C | -0.236 | 0.130 | 0.031 |
| T2167C | -0.221 | 0.234 | 0.012 |
| T2507A | -0.219 | 0.195 | 0.019 |
| C1773T | -0.218 | 0.145 | 0.043 |
| C820T* | -0.213 | 0.268 | 0.011 |

* Statistically significant in the presence of multiple hypothesis testing based on the Benjamini-Hochberg method using a false discovery rate of 0.05.

### Variants associated with HBV DNA load

Quasispecies variants were analyzed for possible association with HBV viral load of the patient. There were n = 22 variants that had a significant association ($p$ <0.05) with HBV DNA load when 10e4 IU/ml (<20,000 IU/mL was used as a threshold) used as a threshold to define high *versus* low status (Figure 4); 10e4 represents the lower end of the range for the definition of chronic hepatitis in HBeAg positive samples in the EASL guidelines. Some of the variants that have a significant association with low viral loads include the PC (G1896A, G1899A) and BCP (A1762T, G1764A, T1753C, C1653T) variants. Interestingly, those variants that are highly associated with HBeAg-negative are exhibiting low viral load, indicating a correlation between HBeAg and low viral loads.

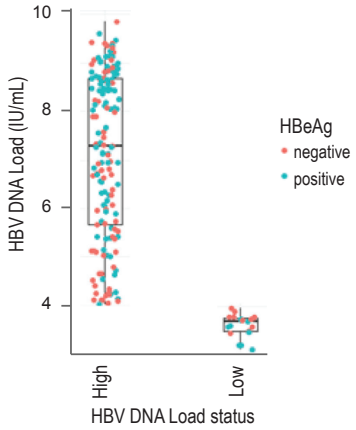



**Figure 4: HBV DNA load (log10 IU/ml) using log10 4 IU/mL as a threshold for high or low HBV DNA load status.** There was a significant difference between the mean HBV DNA load for high and low status ($p < 0.001$).

## DISCUSSION

Understanding the diversity of HBV genome, and identifying quasispecies variants that are associated with patients clinical characteristics are important for the management of chronic HBV infections. In light of this, we investigated HBV quasispecies variants and their associations with relevance to clinical characteristics of the patients using a full-genome deep-sequencing approach. We identified a total of 4615 different variants that appear at least in one patient at a threshold of 1% across the whole HBV genome. There is sufficient evidence for a difference in variant diversity among different genotypes (A-D), with genotype A and B show a greater range of variants across patients. Overall, genotype A shows n=989 unique variants. Of the 4615 variants that were identified across the whole data set only n =208 were found in genotypes A-D. The greatest number of shared variants was between genotypes A and D (n=759). Interestingly, for all genotypes, HBV variant diversity across the genome was higher in viruses from HBeAg- negative patients than HBeAg-positive patients.

It is well-known that the presence of HBeAg in serum is a marker of high levels of virus replication and infectivity while its seroconversion is used as a critical endpoint of treatment or clearance of the virus [245]. However, different mutations in the HBV genome may cause HBeAg-negative chronic hepatitis, in which replicative infection continues and HBV viral load remains detectable. In this study, we identified more than 15 HBV variants that are significantly associated with HBeAg-negative patients. As expected, the uppermost variant was G1896A

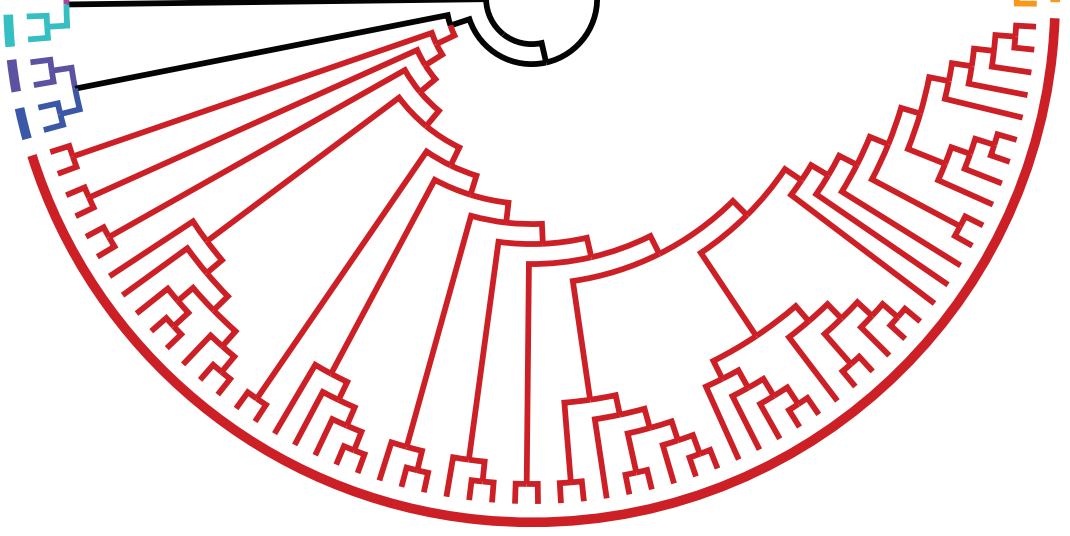
followed by variants including G1899A, A1762T, G1764A, A1934T, and T1753C. The G1896A variant is a G-to-A mutation at nucleotide1896, which creates a premature stop codon and results in complete loss of HBeAg synthesis. Some reports indicated that patients who undergo HBeAg seroconversion are more likely to have improved long-term outcomes, including disease remission, lower cirrhosis and HCC incidence, and increased survival rates [246,247]. Nevertheless, this is not the case in HBeAg negative patients with the G1896A PC mutation. For instance, it has been shown that patients harboring this mutant continue to synthesize HBV DNA at sufficient levels that eventually lead to the progression of liver disease to a more advanced stage [185,248]. Similarly, the BCP A1762T/G1764A mutant variants increase the risk of liver cirrhosis and HCC development [65,71]. BCP variants do not result in absolute HBeAg negativity compared to the PC variant, rather cause a substantial decrease in HBeAg production (~70%) and an apparent increase in viral replication [65]. As the loss of HBeAg alone can't be indicative of viral clearance or disease remission, it is of utmost importance to screen viral variants that are associated with HBeAg-negative and have a clinical impact.

The prediction of HBeAg status using different groups of HBV genome variants was assessed in the present study. To our knowledge, this is the first such study although HBV variants were previously used to predict the significant fibrosis in HBV patients, which unveiled the risk of having significant fibrosis was independently associated with the presence of PC and BCP variants and suggested that detection of these variants may be helpful for monitoring of fibrosis [249]. Importantly, we identified at least thirty viral variants that can predict the HBeAg status, with G1896A being the most powerful predictor. The next most important predictors are variants of the basal core promoter and precore regions including G1899A, A1762T, G1764A, A1934T, and T1753C. Interestingly, all of these variant results in a nonsynonymous substitution and with the exception of A1934T, the variant we found in association with both HBeAg and low viral load, their clinical importance is well studied [71,185]. Our results show that in the absence of the well-known variants, G1899A, A1762T, G1764A, other variants could contribute to HBeAg reduction and therefore helpful in monitoring patients in the clinics.

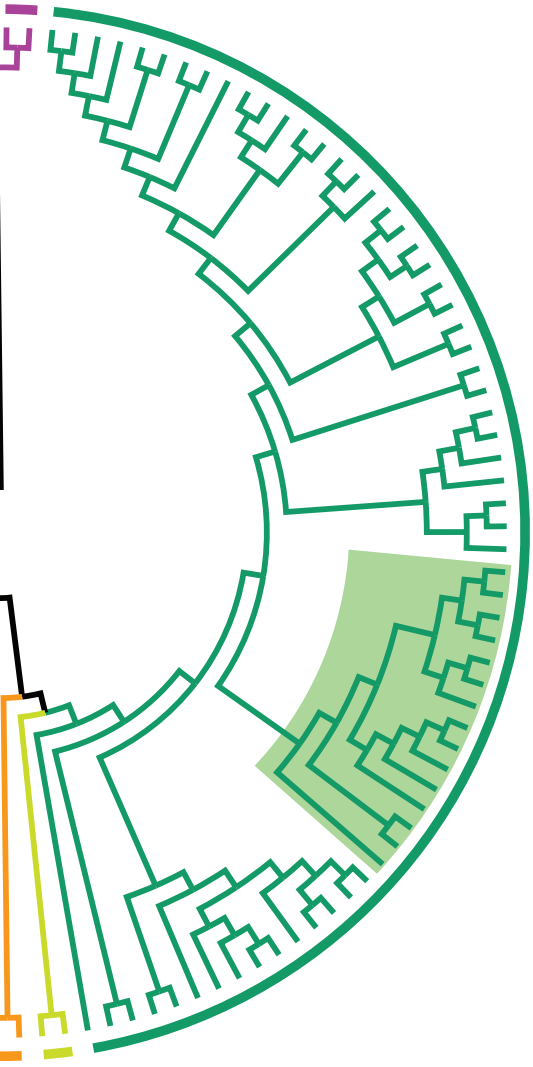
Lastly, we evaluated HBV variants that are associated with baseline HBV DNA levels of the patient. We found twenty-two variants that were significantly associated with low viral load. The well-known PC (G1896A) and BCP (G1764A, A1762T) variants were highly associated with low viral loads when 10e4 IU/ml used as a threshold to define high *versus* low status. Although there have been reports regarding the association of BCP and PC mutants with HBV viral

loads, the association remains controversial: no correlation [243], associated with high viral loads [250], associated with low viral loads [251,252]. In addition, other variants (G1899A, A1934T, G2129C, T2139C, T1753C, C1653T, C2093T) of the precore and core region of the HBV genome were also found to be associated with low viral loads. Interestingly, these variants are significantly associated with HBeAg-negative, indicating a correlation between HBeAg and viral load.

In conclusion, with a full-genome deep-sequencing approach we identified several viral variants in different HBV genomic regions to be significantly associated with baseline levels of viral loads and HBeAg status. It is interesting to note that 30 variants, including the well-characterized G1896A, A1762T, and G1764A, predicated HBeAg status with 92% accuracy as seen in this study. Hence, having a high-quality HBV genome deep-sequencing data can provide novel insight into HBV quasispecies diversity and its relationship to patient clinical characteristics.



# Chapter 7



## Identification of HCV resistant variants against direct acting antivirals in plasma and liver of treatment naïve patients

V. Stalin Raj, Gadissa Bedada Hundie, Anita C. Schürch, Saskia L. Smits, Suzan D. Pas, Sophie Le Pogam, Harry L.A. Janssen, Rob J. de Knegt, Albert D.M.E. Osterhaus, Isabel Najera, Charles A. Boucher, Bart L. Haagmans

*Scientific Reports 2017; 7: 4688 | DOI:10.1038/s41598-017-04931-y.*

## ABSTRACT

Current standard-of-care treatment of chronically infected hepatitis C virus (HCV) patients involves direct-acting antivirals (DAA). However, concerns exist regarding the emergence of drug -resistant variants and subsequent treatment failure. In this study, we investigate potential natural drug-resistance mutations in the NS5B gene of HCV genotype 1b from treatment-naïve patients. Population-based sequencing and 454 deep sequencing of NS5B gene were performed on plasma and liver samples obtained from 18 treatment- naïve patients. The quasispecies distribution in plasma and liver samples showed a remarkable overlap in each patient. Although unique sequences in plasma or liver were observed, in the majority of cases the most dominant sequences were shown to be identical in both compartments. Neither in plasma nor in the liver codon changes were detected at position 282 that cause resistance to nucleos(t)ide analogues. However, in 10 patients the V321I change conferring resistance to nucleos(t)ide NS5B polymerase inhibitors and in 16 patients the C316N/Y/H non-nucleoside inhibitors were found mainly in liver samples. In conclusion, 454-deep sequencing of liver and plasma compartments in treatment naïve patients provides insight into viral quasispecies and the pre-existence of some drug-resistant variants in the liver, which are not necessarily present in plasma.

## INTRODUCTION

Hepatitis C virus (HCV) is a positive-strand enveloped RNA virus, classified in the genus *Hepacivirus*, family *Flaviviridae*. This virus displays very high genetic variability, a primary problem for the development of an effective HCV vaccine and an explanation for the emergence of resistance during antiviral therapy. Accumulation of nucleotide substitutions in the virus has resulted in diversification into numerous subtypes and distinct genotypes [95,253]. The genetic diversity is due to an error-prone RNA-dependent RNA polymerase, which generates on average $1.7 \times 10^{-3}$ base substitutions per site per year, a high virion production rate (up to $1 \times 10^{12}$ particles produced per day), recombination, and deletion [218,254,255]. Consequentially, mutations including those associated with drug resistance are spontaneously generated many times daily in each patient and drug-resistant variants, therefore, may already pre-exist as a minor population within a pool of closely related virus variants called quasi-species [256,257].

Until recently, the only treatment for patients with chronic hepatitis C was a combination of pegylated interferon alpha (Peg-IFN) and ribavirin (RBV), shown to be relatively ineffective with a viral eradication rate of approximately only 50% [258]. Besides, this antiviral therapy is associated with numerous side effects, which excluded up to 50% of patients upfront from antiviral therapy [259]. There is a clear medical need for more efficacious therapies, and nowadays, several novel direct-acting antivirals (DAAs) that target NS3/NS4A protease, NS5B polymerase and NS5A protein either combined with Peg-IFN/ RBV or IFN-free combinations of DAAs have shown potent antiviral effects resulting in high cure rates in HCV- infected patients [260-262]. Antiviral therapy suppressing the wild-type virus but not the pre-existing resistant minority viruses, in due process, may function as a positive selective pressure leading to the rapid outgrowth of drug-resistant variants [263-267].

The most common method of detecting drug-resistant variants in HCV-infected patients is population-based Sanger sequencing. Using standard Sanger sequencing methods, the abundant HCV diversity in chronically HCV-infected patients cannot be fully mapped as relevant proportions of minority variants can be missed [262]. Deep sequencing (DPS), however, now allows for identification of rare minority drug-resistant human immunodeficiency virus variants which are not detectable by standard sequencing techniques [268,269], and recent studies also identified minor drug-resistant variants in plasma of HCV-infected patients [262,264-267]. In this study, we developed a DPS approach to obtain insight into HCV NS5B viral quasi-species and the presence of drug-resistance associated NS5B variants in the plasma and liver tissue of treatment naïve chronic HCV infected patients.

## RESULTS

### Validation of the deep sequencing assay

To allow analysis of HCV quasi-species in liver and plasma of HCV-infected treatment-naïve patients and in-depth analysis of the presence of drug-resistant HCV variants, a DPS approach was developed to analyze a 339 bp nucleotide genome fragment spanning amino acid positions 226 to 337 of the NS5B region. The frequency and nature of potential errors were analyzed by comparing DPS sequence reads of an HCV-1b NS5B plasmid to a consensus sequence obtained from the same construct by Sanger sequencing. To examine errors introduced by PCR and DPS, the plasmid was quantified and diluted to $10^6$ copies per ml, amplified by conventional PCR, and sequenced using the deep sequencing protocol. The contribution of reverse transcription to the error rate of the protocol was analysed using RNA synthesized from the plasmid, after which cDNA, PCR, and 454-sequencing were performed.

The raw sequence data generated in these experiments contained many errors, not uniformly distributed over the amplified region (Figure 1A). Especially GC-rich areas in the sequence (8 to 11 bp GC-rich stretches) led to an increased number of insertions and deletions. A position-specific error rate per nucleotide was determined before read-cleaning algorithms were applied for the control experiments (Figure 1A). The average error rate before read cleaning in all six experiments across the amplicon was similar and ranged from 0.3-0.34%, with on average 1.02-1.16 errors per read. The average error rate after read cleaning in all six experiments across the analyzed amplicon was more dissimilar although not significantly different, ranging from 0.014-0.018% for plasmids and 0.0023-0.0065% for transcripts (Figure 1A). On average, every read retained 0.04-0.06 (plasmids) and 0.008-0.02 (transcripts) errors. These data indicate that most sequencing errors were introduced by PCR and DPS and not by reverse transcription. Similarly, position-specific error rates per amino acid were determined for translated amino acid sequences from the cleaned reads in the six control experiments (Figure 1B). The average error rate after read cleaning was 0.042-0.056% errors per amino acid for plasmids (0.14-0.19 errors per translated read on average) and 0.010-0.021% for transcripts (0.03-0.1 errors per translated read on average) (Figure 1B). This suggests that error introduction through the deep sequencing protocol is less pronounced when starting from RNA transcripts compared to plasmid DNA, although differences did not reach statistical significance.

The read cleaning approach aims to purge errors potentially caused by DPS (but not by PCR or reverse transcription). To examine the effectiveness of the deep

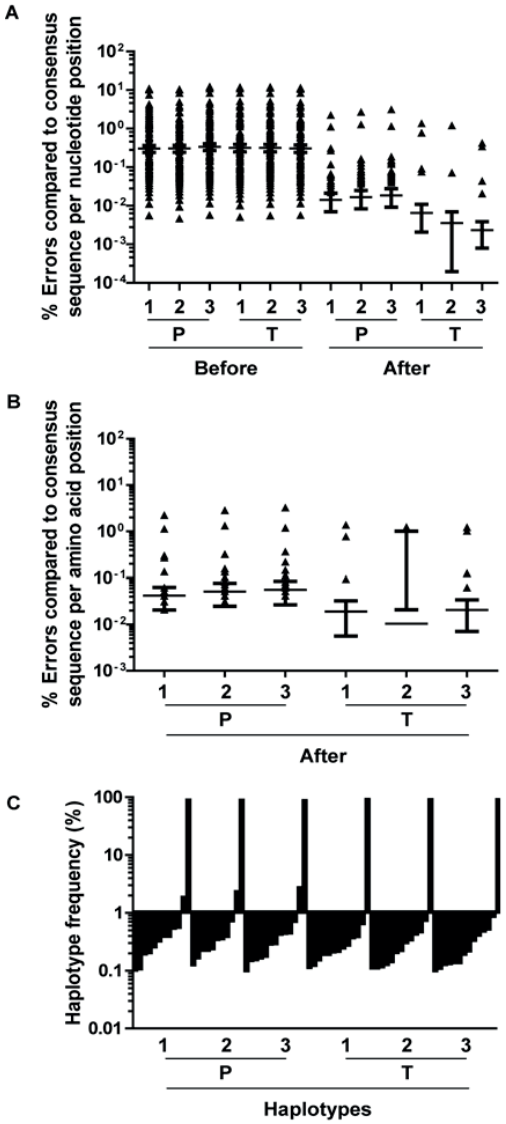



**Figure 1. Validation of the next-generation sequencing assay.** (**A**) A position-specific error rate per nucleotide was determined before or after read-cleaning algorithms were applied for the control experiments with plasmid (P) or RNA transcript (T) in 3 independent next-generation sequencing experiments (1-3) each. Mean ± sem are shown. (**B)** For confirmation, a position-specific error rate per amino acid position was determined after read-cleaning algorithms were applied for the control experiments with plasmid (P) or RNA transcript (T) in 3 independent next-generation sequencing experiments (1-3) each. Mean ± sem are shown. (**C**) Haplotypes and their frequencies in the dataset after read-cleaning were determined for the control experiments with plasmid (P) or RNA transcript (T) in 3 independent next-generation sequencing experiments (1-3) and plotted against each other with a cut-off value of 0.1%.

sequencing analysis, two sets of reads with different 454-specific error profiles were artificially simulated from the sequence of the plasmid. Before read cleaning, the average error rate of the simulated reads was 0.77% and 3.2%, respectively, thereby exceeding the number of errors effectively encountered in the six control experiments. After analysis, all errors were removed (average error rate of ~0% at every position, not shown), indicating the effectiveness of the read cleaning approach in the removal of errors that are typically associated with the deep sequencing technique.

For haplotype reconstruction, the set of reads after read cleaning was translated into amino acid sequences and analyzed for redundancy (Figure 1C). One dominant haplotype was encountered in these control experiments, with a frequency of ~90%. Multiple minor haplotypes that are generated by remaining sequencing errors, which together add up to ~10%, are still present. Most of these errors occur in less than 0.1% of the reads but at certain amino acid positions especially in the control experiments starting from the plasmid DNA, a haplotype generated by sequencing error can occur in percentages as high as ~2.5% of the total cleaned read population (Figure 1C).

As the error rates per nucleotide position are highly diverse, it is unreliable to define a cut-off value based on average error rate per nucleotide position. Instead, each nucleotide or amino acid position should be evaluated separately, which can be achieved for specific positions of special interest, for example, known drug resistance positions. If the variation across the entire genome is of interest, another type of analysis with reconstruction of haplotypes can be performed and the cut-off value can be determined based on data from control experiments with plasmid DNA or RNA transcripts, which in our case was placed at a haplotype threshold frequency of 1% (Figure 1C).

### Analysis of HCV quasispecies in liver and plasma of chronically HCV-infected patients

The HCV quasispecies in liver tissue and plasma from eighteen chronically infected, untreated individuals infected with HCV-1b were analyzed using DPS. All samples were successfully amplified by the NS5B nested-PCR method. The number of RNA molecules subjected to cDNA synthesis was not significantly different between plasma and liver samples (data not shown). Phylogenetic analysis of either nucleotide or amino acid consensus sequences determined by Sanger sequencing revealed that plasma and corresponding liver sequences clustered together (Figure 2). After DPS, we obtained a total of 862,618 reads from 18 paired liver tissue and plasma samples with a median number of reads per sample of ~24,000 (range 8,605 to 37,164, Supplementary Table S3). Subsequently, the

data were cleaned and corrected for DPS sequencing errors to obtain relatively error-free reads. During this process, reads were discarded from each sample and a median of ~23,000 (8,400 to 36,292) reads remained for each sample.

For haplotype reconstruction, the set of corrected reads after read cleaning was translated into amino acid sequences and analyzed for redundancy. The frequency of the haplotypes was determined by counting the number of amino acid sequences with an identical sequence. Most haplotypes in each sample were found at very low frequencies (less than 0.1%). Based on our control experiments with plasmid DNA and RNA transcripts, these haplotypes could be generated through sequencing errors, not cleaned from the reads. To achieve comparability between the different samples while maintaining a concise number of haplotypes,

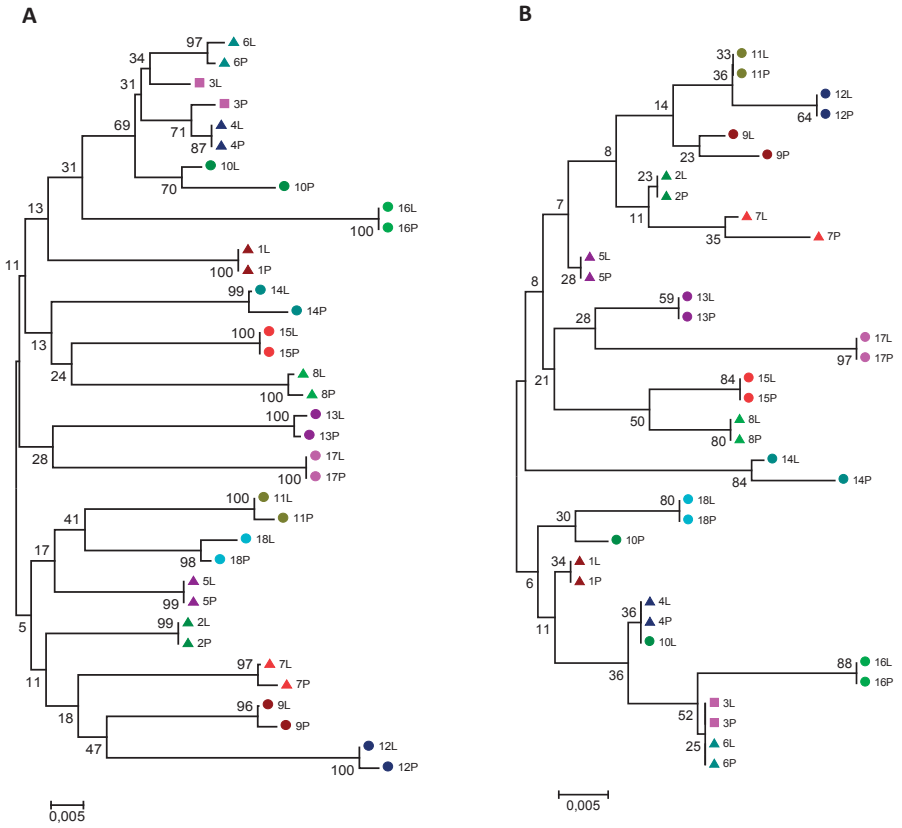



**Figure 2. Phylogenetic analysis of the partial NS5B consensus nucleotide (A) and deduced amino acid (B) sequences from plasma and corresponding liver biopsies from HCV patients.** Phylogenetic trees were generated using MEGA5, with the neighbour-joining method with p-distance model and 1,000 bootstrap replicates. Bootstrap values are shown. The different patients are indicated by colour, shape, and numbering with liver (L) and plasma (P) sequences indicated.

only haplotypes occurring at a frequency of 1% or more were taken into account for further phylogenetic and population analyses (Figure 1C). On average, 5.8 (range 2 to 14) and 4.6 protein haplotypes (range 2 to 7) were observed per liver or plasma sample, respectively. These haplotypes represented 79.3% of the whole population of sampled reads (Supplementary Table S3). Each sample consisted of one or two major protein haplotypes and several minor protein haplotypes (Figure 3). In almost all patients, the most prevalent haplotype in plasma was also present in the liver (Figure 3), except for patient 3, consistent with the results presented in Figure 2A. Compartment unique sequences were observed in all patients (Figure 3). The Simpson's diversity index of the haplotype population was estimated, considering the number of haplotypes, the total number of sequences, and the proportion of the total number of reads found for each haplotype. The diversity index in both compartments was found to be comparable for most patients analyzed (Supplementary Table S3).

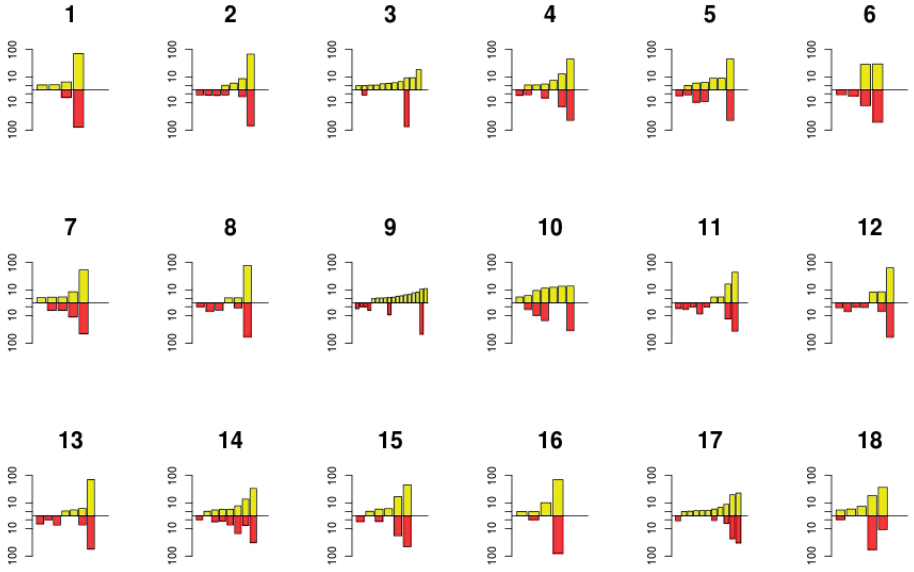



**Figure 3. The number and variation of HCV haplotypes in 18 paired liver and plasma samples.** The graphs indicate the number and frequency of each haplotype with a cut-off of a frequency of 1% or higher per haplotype, per liver (yellow) and plasma (red) sample of each patient (indicated by a number above the graphs).

### Variability at drug resistance positions

To determine whether NS5B gene drug-resistant variants pre-existed, specific codon positions implicated in resistance to NS5B inhibitors or associated with response to IFN/ribavirin treatment were analyzed. To note that within the region analyzed, codon positions 282, 316, and 321 have been implicated in resistance to NS5B inhibitors [261,270-272]; the S282T mutation, for example, confers resistance to 2'-modified nucleotide analogues including the recently approved sofosbuvir.

In our patient cohort, no codon changes from S282 (cut-off position-specific error rate 0) were detected either in plasma or in the liver by either Sanger or deep sequencing (Supplementary Table S4). Sanger sequencing did detect C316N, C316Y, and/or C316H mutations in 9 patient liver and/or plasma samples (50%), whereas deep sequencing detected such a mutation as minor variant in an additional 7 patients in liver and/or plasma samples (89%) with a cut-off of 0.2% based on the position-specific error rate determined in the control experiments (Figure 4 and Supplementary Table S4). The V321I substitution was detected by Sanger sequencing in 1 patient (5.6%) and deep sequencing detected this mutation as a minor variant in 10 patients (55.6%) with a cut-off of 0 based on the position-

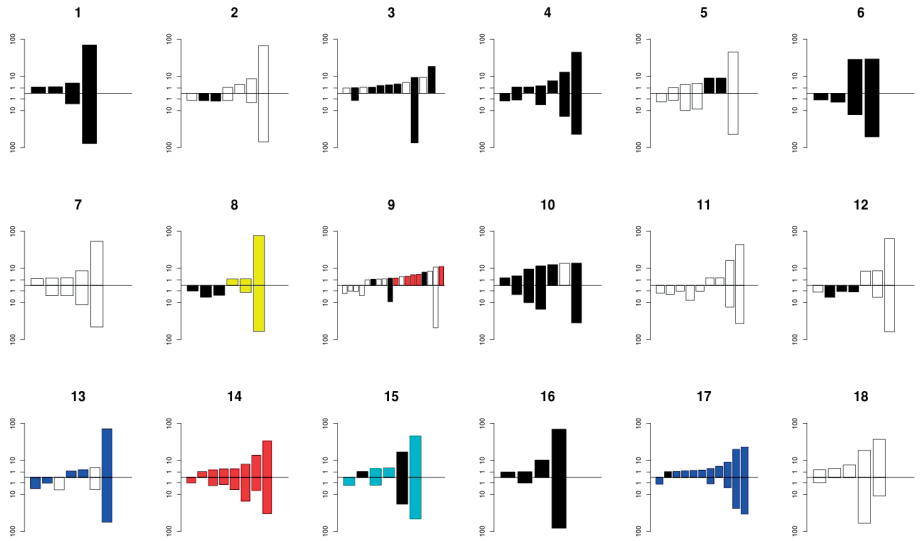



**Figure 4. The number and variation of HCV haplotypes in 18 paired liver and plasma samples.** The graphs indicate the number and frequency of each haplotype with a resistance mutation. The upper bars indicate the liver samples, whereas the lower bars represent plasma samples from the same patient (indicated by a number above the graphs). The colours represent different variants: Black, C316N; Red, C316H+V321I; Blue, Q309R; Cyan, C316N+Q309R; Yellow, C316Y+Q309R. Only haplotypes with a cut-off of a frequency of 1% or higher per haplotype are depicted.

specific error rate determined in the control experiments (Supplementary Table S4). Of interest, double mutations of C316H and V321I were observed in 2 patients in liver and/or plasma samples.

We analyzed six additional codons (310, 329, 244, 309, 326 and 333) that have been associated with a decreased response to IFN/ribavirin therapy (Supplementary Table S4), although none of these variants have been confirmed to be associated with virologic treatment outcome and nothing is known about potential functional associations with the mechanism of action of interferon and/or ribavirin [273-275]. All codon positions showed a position-specific error rate in the control experiment of 0, except for position 333 for which the specific error rate was 0.4%. The A333E mutation was not encountered in any of the patients. The D310N mutation was found in 9 liver and 8 plasma samples as a minor variant in a total of 11 patients. The D244N mutation was found in 5 liver and 9 plasma samples as a minor variant in a total of 10 patients, whereas the S326G mutation was found in 8 liver and 9 plasma samples as a minor variant in a total of 12 patients. The T329I mutation was observed in 11 liver and plasma samples from 11 patients and the Q309R mutation was found in 14 liver and plasma samples as a minor variant and in 4 patients as a major variant. The Q309R mutation was found in conjunction with C316 mutations in three patients. Overall, 22/198 variants were detected in the liver but not in the plasma whereas less unique variants were observed in the plasma (12/198).

## DISCUSSION

In this study, a deep sequencing approach was developed and validated to analyze liver and plasma HCV NS5B quasi-species and drug resistance-associated variants from eighteen treatment naïve patients. We optimised DPS protocols, data cleaning, and error correction strategies using a reference plasmid as input. The data indicated that the intra-assay precision was very high and that most sequencing errors were introduced by PCR and deep sequencing and not by reverse transcription. Although the average error rate after read cleaning was very low in the control experiments, ~0.01% and ~0.03% on the nucleotide and amino acid level, respectively, the range of the error rates per nucleotide or amino acid was 0%-3%. Similar observations in control experiments with a similar deep sequencing platform for HIV-1 quasi-species analysis were obtained previously [268,269]. Thus, each nucleotide- or amino acid-specific position should be evaluated separately, which can be achieved for specific positions of special interest, for example, known drug resistance positions. For the analysis of NS5B

quasi-species in liver and plasma, haplotypes were reconstructed and based on the control experiments, a conservative cut-off was placed at a haplotype frequency of 1%. Interpretation of the data generated in this study requires some caution as the liver needle biopsy specimen may not be representative for an entire liver.

HCV quasi-species diversity has been implicated to play a role in HCV clearance and disease progression with a limited diversity being favourable for HCV clearance but not for disease progression [276-279]. In the majority of patients, the most prevalent haplotype(s) were identical in plasma and liver but compartment unique sequences were observed as reported previously [256,263,280-284]. Comparison of the Simpson's diversity index showed that the extent of diversity was relatively similar between the liver and plasma compartments per patient in 15 patients. A deterministic evolution selecting for the fittest (dominant) strain can be envisaged, based on the apparent presence of the same dominant haplotypes in both plasma and liver in most patients. Although the absence of a unique minority haplotype from plasma does not imply the absence of that haplotype in the liver and vice versa, quite a number of compartment unique haplotypes were obtained. Two hypotheses may explain these observations. One possible explanation could be that extrahepatic HCV replication occurs or a constant and dynamic flow of viral minority quasi-species between the two compartments, plasma and liver, may not occur. Alternatively, minority haplotypes are being generated in some liver areas, but their fitness constraints apparently do not allow them to occupy a large part of the sequence space and therefore are not detected in plasma.

A new era in HCV therapeutics has arrived with the development of direct-acting antivirals therapy and the management of antiviral resistance. We determined whether pre-existing direct-acting antiviral drug resistance mutations in NS5B in plasma and liver tissue of treatment naïve chronic HCV infected patients occurred. Already quite a number of drug resistance mutations have been described using *in vitro* selection protocols or *in vivo*, among which the S282T, C316N,H, Y, and V321Y mutations confer resistance to NS5B (non)-nucleoside inhibitors [270-272,285]. The NS5B S282T variant is associated with a decrease in replicative fitness *in vitro* and has hardly been yet encountered in clinical trials in patients [270,272,286-288]. For instance, in a recent comprehensive analysis of 1344 HCV isolates focussing on the NS5B gene, S282T was present in just one isolate for each genotype 1a, 1b, 3, and 4 at frequencies of 0.17%, 0.24%, 1.24%, and 1.63%, respectively [289]. In our cohort of treatment-naïve patients, the S282T mutation was not detected not even as a minority variant either in plasma or liver tissue. However, the other nucleos(t)ide inhibitors resistant variant V321A was detected in 10 out of 18 patients (~56%) mainly as minority haplotypes. As non-nucleoside inhibitors bind more distantly to the active site of NS5B, resistance-associated variants

often occur more frequently with these compounds [290]. Mutations that confer resistance to non-nucleoside inhibitors at position 316 in NS5B *in vivo* have been described in treatment-naïve patients at frequencies of 0.19% - 24% by Sanger sequencing analyses [274,291-293]. We noticed this mutation in 16 out of 18 patients (~89%) either as dominant (50%) or minority haplotypes (39%). These values are much higher than those obtained with Sanger sequencing and suggest that the presence of certain drug-resistant variants prior to treatment and minority variants are relatively high. In addition, variants containing two mutations in the same genomic strand involved in drug resistance against different compounds were encountered. Two patients showed mutations described in conferring resistance to non-nucleoside compound HCV796 and PSI-352938, a cyclic monophosphate prodrug of 2'-alfa-F-2'-beta-C-methylguanosine. Double mutants in position C316, involved in resistance against compound HCV796, and Q309R, which is associated with a decreased response to IFN/ribavirin therapy, were also detected in two patients. In three out of four patients, the double mutant haplotype was also the dominant haplotype. As drug-resistant mutations can confer a decrease in viral fitness compared to wildtype viruses, it is surprising that they were observed as dominant haplotype in all patient. Possibly compensatory mutations may have evolved in these viruses to increase viral fitness. Unlike deep-sequencing platforms with very short read lengths, such as Illumina, our 454-sequencing approach provides the opportunity to look at the linkage of mutations and identification of double-resistant virus variants, provided that they are located in the same amplicon.

Several other mutations associated with resistance to response to IFN/ribavirin therapy were observed at much higher frequencies in DPS than with Sanger sequencing approaches [273-275]. The D244N, S326G, T329I, and D310N mutations were encountered as a minority variant in 10-12 patients (~60%). The Q309R mutation was encountered most frequently (in all patients) and even as a major haplotype in 4 patients, which is similar to previous observations [274]. However, when we analysed the presence of drug resistance populations in both compartments we noticed that for mutations at most positions with discrepancies between plasma and liver, the prevalence of the resistant variants was higher in the liver as compared to plasma. Thus, the use of plasma is most likely sufficient to detect HCV quasispecies and drug-resistance associated variants. However, additional studies with large cohorts of paired samples, including analysis of other genome regions targeted by DAAs would be needed to reveal the clinical implications of the findings. Overall, our data thus provide insight into the HCV NS5B quasi-species population in liver and plasma in treatment-naïve patients obtained through state-of-the-art sensitive sequencing technologies.

## MATERIALS AND METHODS

### Patients and samples

A total of eighteen patients chronically infected with HCV genotype 1b (HCV-1b) and naïve to direct-acting antivirals (DAAs) treatment were included. Plasma samples and liver biopsies were obtained and stored at -80°C and RNA*later*, respectively, until testing. The baseline clinical characteristics of patients are summarized in Supplementary Table S1 online. The amount of HCV RNA was determined by RT-PCR using Cobas Amplicor HCV Monitor version 2.0 (Roche Diagnostics, Branchburg, NJ) and HCV genotype was determined using INNO-LiPA HCV II (Innogenetics N.V., Ghent). All experimental protocols in this study was approved by the institutional ethics committees of Erasmus Medical Center, Rotterdam, the Netherlands. Informed consent was obtained from all subjects. All methods were performed in accordance with the relevant guidelines and regulations. Paired liver biopsies and plasma were only available from patients infected with genotype 1b in accordance with the approved study protocol for these specific patients.

### Viral RNA isolation, cDNA synthesis and PCR amplification

Viral RNA was extracted from 140 to 280 µl of plasma using the QIAamp viral RNA mini kit (Qiagen) and from liver biopsies (approximately 10 mg tissue was used) using RNeasy mini kit (Qiagen) and the RNA was eluted with 40 µl buffer according to the manufacturer's instructions. To ensure a sufficient amount of viral copies for reliable detection of minor variants, the number of HCV RNA copies in the extracted RNA sample was determined by real-time quantitative polymerase chain reaction (qPCR). Ten microliter RNA was reverse transcribed with the Superscript III first-strand synthesis system (Invitrogen Corp) using random hexamers. The cDNA was used to amplify a 401 bp nucleotide genome fragment spanning amino acid positions 215 to 348 (nucleotides positions 8242 to 8642 according to GenBank accession no AJ238799) of the HCV NS5B polymerase gene with a nested PCR approach using the Hotstar Hifidelity Taq DNA polymerase (Qiagen). Both PCRs were carried out as follows: one initial denaturation step of 95°C for 5 min, followed by 35 cycles of 95°C for 30 s, 48°C for 30 s, 72°C for 40 s and a final extension of 72°C for 10 min. The primers used were, external sense primer Pri3, external antisense primer Pri4, inner sense primer Pri1, inner antisense primer Pri2 [200]. The inner sense and antisense primers were linked to DPS adapters A and B, respectively. To distinguish each sample in the multiplexed DPS, eight unique sequence tags were inserted between the adapter and the gene specific primer (Supplementary Table S2).

### 454 deep sequencing

PCR amplicons were purified from gel using the QIAquick gel extraction kit (Qiagen) and extracted DNA was again purified using Agencourt AMPure XP PCR purification system (Beckman Coulter). The quality and length of the amplicons were verified using Agilent 2100 bioanalyzer (Agilent Life Science, Santa Clara, California) and the concentration was quantified using Quant-iT PicoGreen dsDNA Reagent (Invitrogen) on a TECAN fluorometer (TECAN infinite F200). After quantification, amplicons were pooled in equimolar concentrations, followed by emulsion PCR (emPCR) and bead enrichment according to the 454 Titanium emPCR and enrichment bead recovery protocols (Roche), according to instructions of the manufacturer. The enriched beads were sequenced in both forward and reverse direction on the 454 Life Science platform (454 GS Junior, Roche Applied Science) according to the manufacturer's instructions.

### Sanger sequencing

To analyse the deep sequencing data and to verify the sample authenticity, the NS5B polymerase gene from all plasma and liver tissue samples were PCR amplified as described above and sequenced directly on both strands using the BigDye Terminator version 3.1 Cycle sequencing kit on an ABI PRISM 3100 genetic analyser (Applied Biosystems). To exclude contamination between samples, the Sanger sequences were used to reconstruct a neighbour-joining phylogenetic tree with MEGA 5.05 software using the p-distance model with gamma distributed rate across sites ($\alpha = 0.5$) [294]. Statistical support for internal branches in the tree was obtained by 1000 bootstrap replicates. The GenBank accession numbers of the sequences obtained in this study are KF730703-KF730738.

### Determination of errors introduced by reverse transcription, PCR and DPS

Errors caused by cDNA synthesis, PCR, or DPS are referred to as sequencing errors throughout the study. To quantify the frequency and nature of potential sequencing errors, we synthesized an HCV-1b NS5B polymerase plasmid (GenBank accession No: AJ238799) with T7 promoter (Eurogentech, Belgium). To discriminate between synthetic plasmid and natural HCV sequences, we introduced two mutations in the evolutionarily very well-conserved GDD (GHH) motif present in the active site for the HCV NS5B polymerase in the plasmid. To ensure this synthetic plasmid contained only one plasmid sequence, *E.coli* bacteria were transformed and a plasmid was isolated from a single bacterial clone. The sequence of the plasmid was confirmed by Sanger sequencing.

To assess the error rate of the polymerases in PCR in combination with DPS, the plasmid was quantified using Quant-iT PicoGreen dsDNA reagent, diluted to

$10^6$ copies per ml (an input copy number similar to that of HCV RNA in samples), and amplified by conventional PCR using the Hotstar Hifidelity Taq DNA polymerase (Qiagen). To measure the error rate of reverse transcriptase, RNA was synthesized from the plasmid using RiboMAX large scale RNA production system SP6 and T7 (Promega Corporation, Madison, WI, USA). Subsequently, template DNA was digested with DNase1 and absence of template plasmid in the RNA transcript was verified from serially diluted transcript using the nested PCR system as mentioned above. DNA-free RNA was reverse transcribed using the Superscript III reverse transcriptase and amplified by PCR using Hotstar Hifidelity Taq DNA polymerase (Qiagen). PCR amplified products were analyzed using the DPS protocol; the entire procedure from sample preparation to DPS was repeated three times in the two experiments using plasmid DNA or transcribed RNA.

**Deep sequencing data analysis**

The deep sequencing reads were sorted into their sample of origin according to their unique sequence tag in the primers (Supplementary Table S2). The primer, tag, and adapter sequences were trimmed from the reads. If an average Phred quality score lower than 12 was encountered in a window of four bases across a read, the low-quality bases were removed and the read split at the respective position. Remaining reads were kept if they were longer than 200 bases. The trimming procedure was performed with scripts written in the Python programming language (Python 2.7.3) using Biopython tools (version 1.5.9) [295].

Reads of each sample were aligned to the respective reference sequence with MOSAIK (version 1.3.88, https://code.google.com/p/mosaik-aligner/) as implemented in runMosaik.pl provided in ReadClean454 v1 (RC454) [296]. Aligned reads were corrected for homopolymer stretch polymorphisms (in homopolymers larger than 2N), for 'carry forward and incomplete extension' (CAFIE) errors [297] and for insertion and deletions (InDels) causing frameshifts if they occurred in less than 25% of the reads with RC454. The corrected reads were passed to V-Phaser v1.0 [298], which combines information regarding covariation ("phasing") in reads and an expectation maximization algorithm. Nucleotide and amino acid frequency tables were extracted with V-Profiler [296]. The amount of variation in the deep sequencing reads per nucleotide or translated amino acid position was derived from the frequency tables. Variations were considered as such if at least two high-quality reads exhibited the variant.

The average percentage of variation was determined for the entire amplicon of the control experiments. Furthermore, the variation per position encountered

before read cleaning, which included the errors that were subsequently cleaned, was compared to the variation per position after read cleaning.

**Validation of read cleaning by deep sequencing analysis**

To assess the validity of the read cleaning approach, two sets of reads with different 454 deep sequencing-specific error profiles were simulated from the plasmid sequence using Mason methodology [299]. First, 20000 reads were sampled with an error rate with the standard parameters of Mason-454, resulting in an average of 2.6 errors per reads. Second, 20000 reads with a higher error rate were simulated, with an average of 10.8 errors per read. The simulated reads were subjected to deep sequencing analysis as described above and the variation after read cleaning determined.

**Variant reconstruction**

For haplotype reconstruction, the set of corrected reads was analyzed for redundancy. The amplified product was sequenced as a single fragment; therefore, the frequency of the haplotypes was directly estimated by counting the number of reads with an identical sequence. Each read was translated in all three possible reading frames. Translated reads that contained either an undetermined amino acid or a STOP codon were excluded as these do not represent genuine viruses. To ensure comparability among the samples and to exclude haplotypes generated by sequencing errors, only haplotypes that occurred at a frequency of at least 1% were selected for comparison. Alignments of the amino acid haplotype with a frequency of greater than or equal to 1% from liver and plasma of each patient was performed with Muscle v3.7 and distribution of haplotypes visualized using R statistical software version 2.13.1.

The diversity of the haplotypes was estimated using Simpson's index of diversity 1-D [300], which considers the number of haplotypes, the total number of sequences, and the proportion of each haplotype of the total number of sequences. This value ranges from 0-1 with 0 indicating low diversity and 1 indicating high diversity. Samples were grouped based on the average Simpson's diversity index per patient (liver and plasma) and were considered to have low and high diversity when Simpson's diversity index $\leq 0.5$ or $> 0.5$, respectively.

**Acknowledgements**

This study was funded by Roche and the Virgo consortium, which is funded by the Netherlands Genomics Initiative and by the Dutch Government (FES0908).

## Author Contributions

V.S.R., S.L.S., I.N., C.A.B. and B.L.H. conceived and designed experiments. V.S.R. and S.L.S. implemented and performed the experiments. G.B.H., A.C.S., V.S.R., S.L.S., S.D.P. and B.L.H. analyzed the data. H.L.A.J., R.J.K. and A.D.M.E.O. provided materials and offered helpful discussion. V.S.R., G.B.H., S.L.S., S.L.P., I.N., C.A.B. and B.L.H. interpreted the results. V.S.R., G.B.H., S.L.S. and B.L.H. wrote the manuscript. All authors read and approved the final manuscript.

## Competing interests

Dr. A.D.M.E. Osterhaus and Dr. S.L. Smits are part time chief scientific officer and senior scientist respectively of Viroclinics Biosciences B.V. Dr. H.L.A. Janssen received grants from and is a consultant for: Bristol Myers Squibb, Gilead Sciences, Novartis, InnoGenetics, Roche and Merck. Dr C.A.B Boucher received grants from and is consultant for Merck, Roche and Abvie.

## SUPPLEMENTARY INFORMATION

*Supplementary Tables*

**Table S1: Patient characteristics at baseline**

| Patient | Age (Years) | Sex[a] | Serum viral load (IU/ml) | ALT | Fibrosis |
|---|---|---|---|---|---|
| 1 | 55 | M | $2.60 \times 10^6$ | 195 | 5 |
| 2 | 48 | F | $4.95 \times 10^6$ | 91 | 3 |
| 3 | 45 | M | $3.66 \times 10^6$ | 79 | 3 |
| 4 | 29 | M | $1.63 \times 10^6$ | 67 | 2 |
| 5 | 43 | F | $2.30 \times 10^6$ | 35 | 1 |
| 6 | 25 | F | $3.90 \times 10^6$ | 48 | 1 |
| 7 | 47 | M | $4.60 \times 10^6$ | 68 | 2 |
| 8 | 36 | M | $6.98 \times 10^6$ | 85 | NA |
| 9 | 54 | M | $1.71 \times 10^6$ | 107 | 4 |
| 10 | 54 | M | $4.20 \times 10^6$ | 59 | 3 |
| 11 | 37 | F | $7.81 \times 10^5$ | 31 | 3 |
| 12 | 45 | F | $1.65 \times 10^6$ | 43 | 0 |
| 13 | 53 | M | $1.84 \times 10^6$ | 19 | 1 |
| 14 | 39 | M | $5.13 \times 10^6$ | 86 | 5 |
| 15 | 27 | M | $1.65 \times 10^6$ | 61 | 2 |
| 16 | 21 | M | $4.81 \times 10^5$ | 65 | 1 |
| 17 | 31 | M | $2.23 \times 10^6$ | 41 | NA |
| 18 | 60 | M | $6.46 \times 10^6$ | 73 | 3 |

[a] M, male; F, female; NA, not available

**Table S2: DPS primer sequence**

| No | Primer name | Adopter | Taq | MID | Genespecific primer |
|---|---|---|---|---|---|
| 1 | Pr1FNS5bMID1 | 5'-CGTATCGCCTCCCTCGCGCCA | TCAG | ACGAGTGCGT | TGGGGATCCCGTATGATACCCGCTGCTTTGA-3' |
| 2 | Pr2RNS5bMID2 | 5'-CTATGCGCCTTGCCAGCCCGC | TCAG | ACGCTCGACA | GGCGGAATTCCTGGTCATAGCCTCCGTGAA-3' |
| 3 | Pr2RNS5bMID3 | 5'-CTATGCGCCTTGCCAGCCCGC | TCAG | AGACGCACTC | GGCGGAATTCCTGGTCATAGCCTCCGTGAA-3' |
| 4 | Pr2RNS5bMID4 | 5'-CTATGCGCCTTGCCAGCCCGC | TCAG | AGCACTGTAA | GGCGGAATTCCTGGTCATAGCCTCCGTGAA-3' |
| 5 | Pr2RNS5bMID5 | 5'-CTATGCGCCTTGCCAGCCCGC | TCAG | ATCAGACACG | GGCGGAATTCCTGGTCATAGCCTCCGTGAA-3' |
| 6 | Pr2RNS5bMID6 | 5'-CTATGCGCCTTGCCAGCCCGC | TCAG | ATATCGCGAG | GGCGGAATTCCTGGTCATAGCCTCCGTGAA-3' |
| 7 | Pr2RNS5bMID7 | 5'-CTATGCGCCTTGCCAGCCCGC | TCAG | CGTGTCTCTA | GGCGGAATTCCTGGTCATAGCCTCCGTGAA-3' |
| 8 | Pr2RNS5bMID8 | 5'-CTATGCGCCTTGCCAGCCCGC | TCAG | CTCGCGTGTC | GGCGGAATTCCTGGTCATAGCCTCCGTGAA-3' |

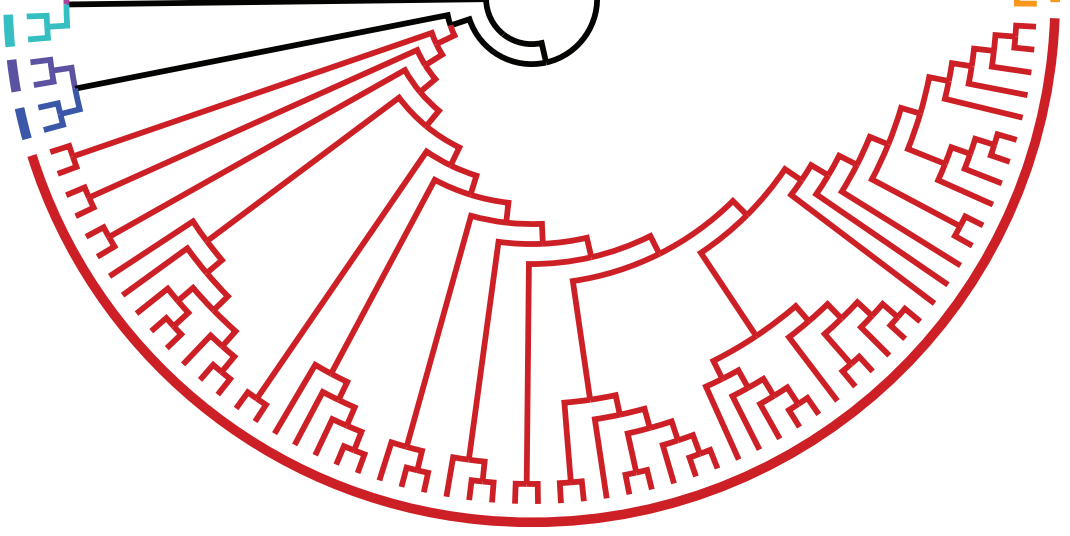
**Table S3: Amount of HCV reads, haplotypes, and Simpson's diversity index per liver (L) or plasma (P) sample of each patient.**

| Patient | No of reads | No of "cleaned" reads | No reads of haplotypes with frequencies of more than 1% | No haplotypes | Simpson's Diversity Index |
|---|---|---|---|---|---|
| 1L | 16049 | 15650 | 13588 | 4 | 0.14 |
| 1P | 17913 | 17540 | 15725 | 2 | 0.08 |
| 2L | 18885 | 18411 | 16381 | 4 | 0.23 |
| 2P | 18361 | 18010 | 16240 | 6 | 0.21 |
| 3L | 14094 | 13672 | 8185 | 11 | 0.77 |
| 3P | 27949 | 27102 | 23350 | 2 | 0.06 |
| 4L | 27468 | 26241 | 21966 | 6 | 0.48 |
| 4P | 29496 | 28706 | 23879 | 5 | 0.48 |
| 5L | 29960 | 28893 | 23641 | 6 | 0.47 |
| 5P | 31467 | 30685 | 24433 | 5 | 0.45 |
| 6L | 25874 | 24743 | 19848 | 2 | 0.5 |
| 6P | 26824 | 25742 | 21612 | 4 | 0.37 |
| 7L | 29932 | 27883 | 22646 | 5 | 0.29 |
| 7P | 27569 | 26828 | 21128 | 4 | 0.41 |
| 8L | 24848 | 24111 | 21144 | 3 | 0.06 |
| 8P | 21393 | 20198 | 16800 | 5 | 0.24 |
| 9L | 17643 | 16750 | 10546 | 14 | 0.89 |
| 9P | 21213 | 20267 | 15883 | 6 | 0.37 |
| 10L | 13760 | 12948 | 9933 | 7 | 0.82 |
| 10P | 26634 | 25865 | 21036 | 4 | 0.58 |
| 11L | 16836 | 16013 | 13191 | 4 | 0.45 |
| 11P | 24273 | 23343 | 19020 | 7 | 0.58 |
| 12L | 14365 | 13498 | 12056 | 3 | 0.28 |
| 12P | 22374 | 21430 | 18806 | 6 | 0.3 |
| 13L | 17245 | 16327 | 14248 | 4 | 0.15 |
| 13P | 23966 | 22655 | 19090 | 5 | 0.33 |
| 14L | 33251 | 31536 | 24757 | 7 | 0.61 |
| 14P | 36010 | 34662 | 28486 | 7 | 0.64 |
| 15L | 28058 | 27543 | 24066 | 5 | 0.49 |
| 15P | 37164 | 36292 | 31941 | 4 | 0.48 |
| 16L | 32415 | 30081 | 22151 | 4 | 0.24 |
| 16P | 22707 | 22075 | 19664 | 2 | 0.02 |
| 17L | 28402 | 27409 | 21067 | 10 | 0.72 |
| 17P | 25854 | 24885 | 21212 | 5 | 0.56 |
| 18L | 8605 | 8400 | 7089 | 5 | 0.56 |
| 18P | 23761 | 23209 | 19571 | 3 | 0.26 |

**Table S4: Detection of HCV-NS5B resistant variants by DPS (% of total cleaned reads).**

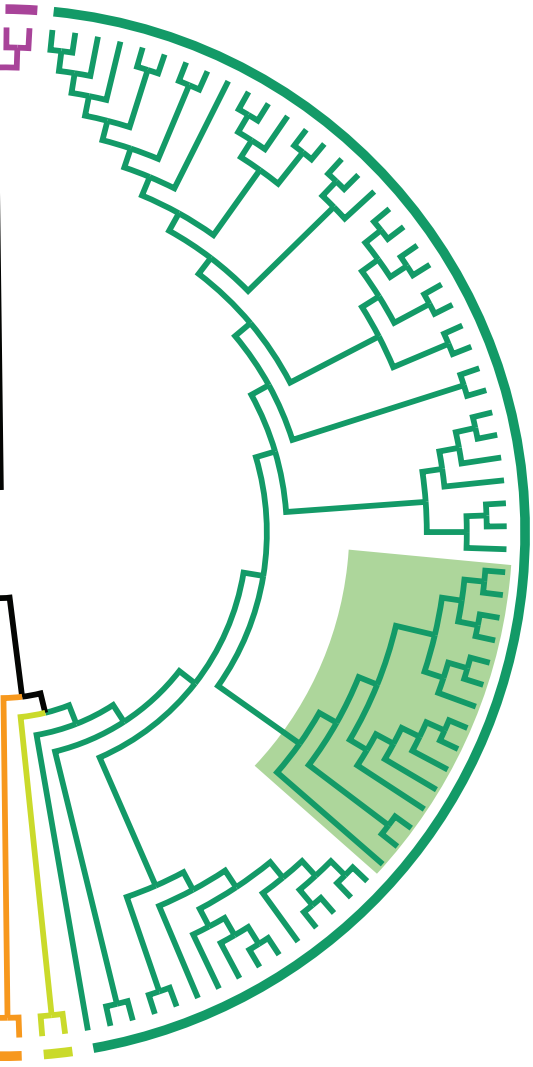
| Patient | S282T | C316N | C316H | C316Y | V321I | D244N | Q309R | D310N | S326G | T329I | A333E |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *cutoff* | 0 | 0.2 | 0.2 | 0.2 | 0 | 0 | 0 | 0 | 0 | 0 | 0.4 |
| 1L | 0 | 99.9 | 0 | 0 | 0 | 0 | 0.23 | 0.02 | 0 | 0 | 0 |
| 1P | 0 | 99.92 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2L | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2P | 0 | 4.38 | 0 | 0 | 0.03 | 0 | 0.54 | 0 | 0 | 0 | 0 |
| 3L | 0 | 65.66 | 3.12 | 5.63 | 0 | 0.02 | 5.82 | 0 | 0 | 0.02 | 0 |
| 3P | 0 | 99.81 | 0 | 0 | 0 | 0 | 0.18 | 0.03 | 0 | 0.07 | 0 |
| 4L | 0 | 99.93 | 0 | 0 | 0.02 | 0 | 0.11 | 0 | 0 | 0 | 0 |
| 4P | 0 | 99.88 | 0 | 0 | 0 | 0.02 | 0.19 | 0.04 | 0.03 | 0 | 0 |
| 5L | 0 | 20.77 | 0 | 0 | 0 | 0 | 0.25 | 0.22 | 0.03 | 0.04 | 0 |
| 5P | 0 | 0 | 0 | 0 | 0.04 | 0.05 | 0.21 | 0.08 | 0.06 | 0.07 | 0 |
| 6L | 0 | 99.92 | 0 | 0 | 0.02 | 0 | 0.35 | 0.88 | 0 | 0 | 0 |
| 6P | 0 | 99.83 | 0 | 0 | 0 | 0 | 0.21 | 0 | 0 | 0 | 0 |
| 7L | 0 | 0.28 | 0 | 0 | 0.02 | 0.11 | 0.51 | 0.02 | 0.02 | 0.08 | 0 |
| 7P | 0 | 0 | 0 | 0 | 0.01 | 0.04 | 0.62 | 0.03 | 0.06 | 0.09 | 0 |
| 8L | 0 | 0.34 | 0 | 99.47 | 0.03 | 0.07 | 99.53 | 0.07 | 0.02 | 0.14 | 0 |
| 8P | 0 | 13.66 | 0.8 | 85.26 | 0 | 0.04 | 85.28 | 0.03 | 0.08 | 0.07 | 0 |
| 9L | 0 | 11.55 | 43.91 | 1.39 | 0.02 | 0.01 | 2.42 | 0.2 | 0.17 | 0.2 | 0 |
| 9P | 0 | 13.27 | 0 | 0 | 0 | 0.13 | 0.48 | 0.02 | 0.03 | 0.16 | 0 |
| 10L | 0 | 73.55 | 0.64 | 0 | 0 | 0 | 0.9 | 0.1 | 0 | 0.12 | 0 |
| 10P | 0 | 99.56 | 0 | 0 | 0 | 0.08 | 0.4 | 0 | 0.05 | 0.07 | 0 |
| 11L | 0 | 0 | 0 | 0 | 0.03 | 0 | 0.68 | 0 | 0.07 | 0.03 | 0 |
| 11P | 0 | 0 | 0 | 0 | 0 | 0 | 0.42 | 0.03 | 0.06 | 0.11 | 0 |
| 12L | 0 | 0.3 | 0.65 | 0 | 0 | 0 | 0.16 | 0 | 0 | 0 | 0 |
| 12P | 0 | 7.31 | 0 | 0 | 0 | 0.04 | 0.28 | 0 | 0.03 | 0 | 0 |
| 13L | 0 | 0 | 1.7 | 0 | 0 | 0 | 94.17 | 0 | 0 | 0.01 | 0 |
| 13P | 0 | 0 | 0 | 0 | 0 | 0 | 83.41 | 0 | 0.11 | 0.17 | 0 |
| 14L | 0 | 0 | 97.41 | 0 | 0 | 0 | 2.53 | 0.09 | 0.28 | 0.05 | 0 |
| 14P | 0 | 0 | 99.73 | 0 | 0.02 | 0.01 | 0.88 | 0.02 | 0 | 0.04 | 0 |
| 15L | 0 | 99.93 | 0 | 0 | 0.03 | 0.09 | 72.61 | 0.02 | 0 | 0.85 | 0 |
| 15P | 0 | 99.91 | 0 | 0 | 0 | 0.06 | 71.42 | 0 | 0 | 0.8 | 0 |
| 16L | 0 | 99.84 | 0 | 0 | 0 | 0 | 0.25 | 0 | 0.05 | 0 | 0 |
| 16P | 0 | 97.78 | 0 | 0 | 0 | 0 | 0.18 | 0 | 0 | 0 | 0 |
| 17L | 0 | 2.9 | 0 | 0 | 0 | 0 | 94.69 | 0 | 0.01 | 0.03 | 0 |
| 17P | 0 | 0.55 | 0 | 0 | 0 | 0 | 98.82 | 0 | 0 | 0.08 | 0 |
| 18L | 0 | 0 | 0 | 0 | 0 | 0 | 0.23 | 0 | 0 | 0 | 0 |
| 18P | 0 | 0 | 0 | 0 | 0 | 0 | 0.39 | 0 | 0 | 0 | 0 |

L, liver; P, plasma. Cut-off, position specific error rate determined in control experiments.



# Chapter 8

## Summarizing discussion



## Overview

Hepatitis B and C viruses are a major cause of morbidity and mortality worldwide [6]. According to the World Health Organization (WHO) report, viral hepatitis caused 1.34 million deaths in 2015, a number comparable to annual deaths caused by tuberculosis but higher than those caused by either HIV or malaria [301]. However, in contrast to these other three infectious diseases, viral hepatitis has received relatively little attention. Worldwide, only 9% of HBV and 20% of HCV infected individuals have access to affordable hepatitis testing. Even more worrisome is the fact that antiviral treatment is available for only 8% of those diagnosed with HBV and 7.4% of those diagnosed with HCV [301]. This is even more problematic in countries like Ethiopia where more than 80% of the population are living in rural areas with a very limited access to healthcare. Besides that, viral hepatitis has never been considered a health priority in Ethiopia due to a lack of awareness and the fact that no national antiviral treatment and diagnostic guidelines or surveillance were developed. In addition, only in 2007, the universally available HBV vaccine was incorporated into the national neonatal immunization program in three doses of pentavalent vaccine (DTP-HepB-Hib). Consequently, these viruses continue to spread and the burden remains substantial. On-top-of-that, both HBV and HCV exhibit high genetic diversity that affects clinical and treatment outcomes. Most importantly, this genetic diversity revealed by different genotypes, subgenotypes, subtypes, quasispecies, and recombinants, differs in geographic distribution [54,102]. Therefore, for a proper management and prevention of these viruses in Ethiopia, understanding their prevalence and genetic diversity is extremely important. In **chapter** 2 we investigated the seroprevalence of HBV and HCV infection in Ethiopia. The molecular epidemiology and genetic diversity of HBV in different geographic regions of Ethiopia were described in **chapter 3**. The in-depth complete genome analyses of important and selected strains were determined in **chapter 4**. The genetic diversity of HCV in Ethiopia was explored in **chapter 5.** In **chapter 6** the relevance of HBV quasispecies variant diversity and their association with patients' clinical characteristics such as HBeAg status and viral load was assessed. Finally, in **chapter 7** potential pre-existence of direct-acting antivirals resistance-associated variants were investigated in treatment naïve chronic HCV patients.

## Hepatitis B and C viruses in Ethiopia

In Ethiopia, due to the lack of national surveillance program and availability of routine diagnostic tests for patients, the burden of HBV and HCV infection largely remains unknown. Previous studies showed seroprevalence variations in HBV (2.1 to 25.0%) and HCV (0.7 to 13.3%) infections [113-115,119]. However, most

of these reports are either from a limited number of subjects (small sample size) or among specific groups (HIV-infected subjects, liver patients, commercial or replacement blood donors, pregnant women etc.) and may not represent the true national burden of viral hepatitis in the general population. In the present study, we tried to fill in this gap by recruiting a large number of voluntary healthy blood donors (56 885) drawn from five distinct geographic regions that are approximately 100 to 1 500 kilometers distant from each other and differ in their socioeconomic, culture and ethnicity (**Chapter 2**). We found an overall nationwide seroprevalence of 3.9% (ranging from 3.7 to 5.91%) for HBV and 0.52% (ranging from 0.37 to 0.65%) for HCV. These prevalence rates classify the country as an intermediate (2 –7%) endemic area for HBV infection and low (< 1.5%) for HCV rather than a default classification of the country as a highly endemic for HBV (> 8%) because of its geographic location in sub-Saharan Africa. Countrywide, the prevalence of HBV was higher in the eastern region (Adama) compared to the other regions. This could be due to occupational and lifestyle differences as the majority of the population in the eastern region are pastoralists who commonly share unsterile and re-used equipment. In addition, this region is the main route of trade in and out of the country, associated also with changes in human behavior such as increased risk for sexual transmitted diseases.

The most recent global assessment of the country-level population prevalence of chronic HBV infection reported a seroprevalence of 6.03% in Ethiopia [8], which is in coherence with our findings (intermediate prevalence). HCV seroprevalence of 0.9%, 1.0%, and 0.3% was found in the general population of the neighboring countries Somalia, Sudan, and Djibouti, respectively [302]. Our nationwide seroprevalence data undoubtedly indicate the overall picture of HBV and HCV epidemiology in the country but might be underestimated due to exclusion of blood donors with clinical signs and symptoms of hepatitis. In summary, the burden due to viral hepatitis in Ethiopia is substantial and there is a need for robust national-level surveillance and public campaigns to raise awareness of viral hepatitis.



**HBV molecular epidemiology and genetic diversity in Ethiopia**

HBV exhibits a high genetic diversity with 10 main genotypes, more than forty subgenotypes and numerous recombinants recognized worldwide. In Africa, mainly three major genotypes circulate with different epidemiological trends. Genotype A (subgenotype A1) is predominant in Eastern and Southern Africa, genotype D (D1 and D7) in Northern Africa, and genotype E in Western and Central Africa [53]. In Ethiopia, we found two main genotypes, A (78%) and D (22%) based on S/pol gene sequence analysis of 371 samples obtained from

five regions: Addis Ababa (central), Adama (central-east), Gondar (northwest), Jimma (southwest), and Mekelle (north) (Chapter 3, Figure 2). Both genotypes are prevailing throughout the country but with a different epidemiological pattern. Genotype D was more prevalent in the northern region of the country compared to the central and southwestern regions. For instance, about 40% of the HBV isolates in Mekelle belongs to genotype D whereas in Jimma and Adama the prevalence of genotype D is below 10%. In contrast, genotype A is predominantly distributed in the central and southwestern regions (Chapter 3, Figure 2). When we analyze the epidemiology of HBV in the neighboring countries, genotype A is the predominant genotype in Somalia, Kenya, Rwanda, and Uganda [133,303], while genotype D is the most prevalent genotype in Sudan and Egypt [151,304]. It is noteworthy to say that Ethiopia is at the border of the genotype A and D geographic distribution in Africa based on these observations.

Of note, all genotype A isolates from Ethiopia belong to subgenotype A1 similar as in other eastern and southern Africa countries [30,133,150]. However, they are genetically divergent from the majority of strains from these regions, because these strains from Ethiopia cluster in both the African and the Asian–American clades (Chapter 3 Figure S2). Fascinatingly, most of the Ethiopian strains clustered more closely to the Asian-American clades (strains from Brazil, Haiti, Japan, India, Philippines) rather than to the African clade (Rwanda, Malawi, Uganda, Tanzania or South Africa). It has been suggested that genotype A originates in Africa and did spread from Eastern Africa to Southern Asia during the last 1000-2000 years by coastline travel and trade [305]. Similar studies in Central and South America also found that A1 originates in East Africa and were introduced to these regions during the 19$^{th}$ century slave trade and human migration [179,180]. Taking into consideration the early split of A1 strains from Ethiopia in the phylogenetic tree and Ethiopia's location at a strategic geographic position between Africa and Eurasia, combined with the knowledge on the main route of human migration out of Africa as well as the origin of modern humans [172-174], we hypothesize that HBV strains from Ethiopia could be the oldest lineage of the African and the Asian-American A1 clade.

Genotype D, the second most prevalent genotype in Ethiopia is genetically highly diverse and differentiated into five subgenotypes: D1, D2, D4, D7 and one unassigned subgenotype (now named D10, see chapter 4). Interestingly, the majority of genotype D strains (85%) belonged to two subgenotypes: D2 (55%) and the novel D10 (30%). It is important to recall that subgenotype D2 is prevalent in Eastern Europe including Russia and the Baltic countries [157], but less common in Africa where D1 is the dominant subgenotype. Nevertheless, the Ethiopian D2 strains clustered separately and more distantly to D2 of the

European and Asian origin with strong bootstrap support (100%) (Chapter 3, Figure 3B). The D2 strains from Ethiopia are found throughout the country and are quite divergent. Furthermore, Ethiopia does not have a historical relationship with those European countries where D2 predominates. Interestingly, D2 was also reported among Ethiopian immigrants living in Australia [306]. Thus, we suggest that D2 from Ethiopia evolved separately in a distinct ethno-geographic to that of the European D2 and has a long natural history in the Ethiopian population.

**Novel HBV subgenotype D10**

As discussed in chapter 3, two main genotypes (A and D) and around 6 subgenotypes are prevailing in Ethiopia, with subgenotypes A1, D2 and the unassigned subgenotype hereafter called D10 being dominant. However, as that study was limited to the small S gene sequence analysis, several strains remained unclassified and the ambiguity of some sequences undetermined. So, to better descripe HBV evolution in Ethiopia 29 HBV strains — 10 genotype A and 19 genotype D (14 unassigned strains, 3 subgenotype D2, 1 D4 and 1 D/A recombinant) — were selected for full-genome sequence analysis. Except for strain AA44Eth (Figure 4A, chapter 4) which was a putative A1/D recombinant, all genotype A strains belonged to A1, and the majority of them clustered in the Asian-American clade together with Somalia strains than to the African clade (Figure 1, chapter 4), confirming our earlier findings. Similarly, the three D2 strains were once again classified again into subgenotype D2 but with a distinct branch (Figure 2, chapter 4). Notably, strain MK141ETH that was suspected to be a D/A recombinant in the previous study, was found to be a triple recombinant (D/A/E) when analyzing the complete genome sequence (Figure 4B, chapter 4). Although genotype E has been largely reported in the neighboring country Sudan [151], it has not been reported in Ethiopia until most recently [307]. However, the two genotype E strains reported among human immunodeficiency virus-positive patients were based on core region fragment sequence analysis [307] and could be D/E recombinant as D strains showed a higher signal for genotype E than genotype D in the X to core region[39]. In addition, D/E recombinant strains were also found among Ethiopian children living in Australia (unpublished data). In general, these putative genotype E signal indicates that genotype E could also circulate in Ethiopia and thus emphasizes a need for more HBV sequences from this region of the world.

Interestingly, the unassigned strains display a distinct monophyletic tree and possess unique distinguishing features, which is ample evidence to classify them into a novel subgenotype D10. In the first place, these strains were obtained from unrelated blood donors from different regions of Ethiopia. Secondly, they formed

a distinct monophyletic cluster strongly supported by high bootstrap value both in the complete genome (Figure 2, chapter 4) as well as in the four separate open reading frames (ORFs) phylogenetic trees (Figure 3, chapter 4). Thirdly, the mean inter-subgenotype nucleotide divergence compared with reference sequences of all existing subgenotypes D was >4% and <7.5%, which is the minimum requirement for the definition of a new subgenotype [53]. Lastly, they had nine unique amino acid substitution positions in the surface, polymerase, and X genes; five of those were exclusively unique to D10 (Table 2, Chapter 4). Moreover, unlike its preceding subgenotypes D9 [41] and D8 [38] which are recombinant of C/D and D/E respectively, the novel D10 is free of any recombination. This can be noticed in the four ORFs phylogenetic trees which were further confirmed by different HBV recombinant analysis methods (RDP4, SimPlot and jpHMM). Furthermore, the D10 strains exhibited an unusual mutation in different regions of the HBV genome. For instance, 67% of the D10 strains harbor the PC premature stop codon mutation (G1896A) and 27% the preS2 deletion (6 to 51 nucleotide deletion). The clinical importance of these mutations was shown in several studies [65,186,187,226], however, the biological/molecular reason why D10 strains have high mutations and whether these mutations contributed to the evolution of D10 needs further study.

In summary, the identification of novel D10 and a unique distribution of subgenotype D2 in Ethiopia, indicate the high genetic variability and evolution of HBV in Africa, particularly in Ethiopia.

**HCV genetic diversity in Ethiopia**

HCV, the etiologic agent of 185 million chronic hepatitis infections and the leading cause of liver cirrhosis and HCC, is characterized by a great genetic variability. The virus is classified into 7 genotypes and more than 100 subtypes [94]. Each of these genotypes is unique with respect to its nucleotide sequence, geographic distribution, and therapeutic response. Genotypes 1, 2 and 3 have a worldwide distribution, while genotypes 4, 5, 6, and 7 have restricted geographic distribution [102]. In Africa, the dominant genotypes are 1, 2, 4 and 5, with different epidemiological patterns. Genotype 1 and 2 are most prevalent in Central and West Africa, 4 in Central and North Africa, and 5 in South Africa [96]. Nevertheless, little is known about the prevailing genotypes and subtypes in Eastern Africa, including Ethiopia. In the present nationwide study, we found that 4 major genotypes (1, 2, 4, and 5) and seven subtypes (1b, 1c, 2c, 4d, 4l, 4r, and 4v) are circulating in Ethiopia (chapter 5). Genotype 4 is by far the most dominant accounting for 78% of the HCV strains characterized, followed by genotype 2c. The most frequent subtypes are 4d and 4r, each accounting for 35% of the HCV

strains circulate. This genotype prevails throughout the country as seen from the data analyzed from the five geographic regions. However, unlike HBV, regional differences in the distribution of HCV genotypes could not be established due to a limited number of samples studied. In addition, a potential natural recombinant virus was identified inconsistent with the presence of the two main genotypes (4 and 2) through analysis of two distinct regions of the HCV genome (Chapter 5 Figure 1 and 2). Overall, the findings from the present study underline the need for further molecular characterization of HCV from this region.

It is well known that knowledge of HCV genotype information is useful for making treatment decisions, such as treatment duration, type of medicine to use, and the likelihood of treatment response. Genotype 4 is one of the most difficult to treat genotypes with the available standard treatment, peginterferon alpha plus ribavirin [210]. Nowadays, pan-genotypic direct-acting antivirals are the choice of treatment in the western world [308]. However, use of these agents in resource-limited countries endemic for HCV genotype 4 is currently hindered by the very high costs. The ever-increasing epidemiology of genotype 4 in western countries due to immigration from endemic regions [191,309] could make these drugs available in resource-limited countries subsequently. Hence, having ample information on the prevalent genotype/subtype and presence of mutations associated with resistance to direct-acting antivirals is highly essential for the future development of treatment guidelines and patient management in regions like Ethiopia where no national hepatitis treatment policy in place at the moment.

**HBV quasispecies variants association with patients HBeAg status and viral load**

In the last decade, ample information has been obtained on the HBV genetic diversity. Although deep-sequencing of selected genome regions allowed the detection of viral variants of low frequencies [221], variant analysis at the full genome level has not been performed in detail. We studied the association of HBV quasispecies variants using full genome sequencing with HBeAg status and viral load in chronically infected HBV patients. We identified a total of 4615 different quasispecies variants across the full-genome of five HBV genotypes (A to D) in 186 patients at a threshold of 1% frequency. This quasispecies variant diversity was significantly higher in viruses from HBeAg-negative patients compared to HBeAg-positive patients (Figure 2, chapter 6). HBeAg is clinically important because its presence in serum is a marker of viral replication and its seroconversion is a potential endpoint of therapy. However, there are HBV variants that can reduce or eliminate the HBeAg production, allowing the virus to evade host immune attack and continued replication [28,65]. As a result, the loss

of HBeAg alone can't be indicative of viral clearance or disease remission, it is of utmost importance to identify viral variants that are associated with HBeAg status. We identified multiple variants that are significantly associated with HBeAg status (Table 4 and Figure 3, chapter 6). The majority of these variants were located in core gene followed by surface gene, which might be resulted from the highly selective immune pressure exerted in these regions [28]. In addition, we noted that about one-third of these variants found to be associated with low viral load, indicating the correlation between HBeAg and HBV DNA levels. Indeed, the prevalence of HBeAg negative chronic hepatitis B has been increasing over the past few decades and has become the commonest type of HBV infection in many countries of the world [310]. Overall, these results show that in addition to the well-known PC and BC variants that reduce or eliminate the HBeAg production, other variants could contribute to HBeAg reduction and therefore helpful in monitoring patients. To conclude, a high-quality HBV genome deep-sequencing approach can provide novel insight into HBV quasispecies variant diversity and its relationship to patient clinical characteristics.

**Direct-acting antivirals resistance associated HCV variants**

In the early 2000s, the combination of pegylated interferon plus ribavirin was the standard-of-care treatment for HCV infection. However, these drugs are associated with frequent and sometimes serious adverse effects with limited efficacy [308], leaving researchers to look for alternative options. Since 2011, the HCV treatment has dramatically changed due to the advent of DAAs. These drugs target the nonstructural coding genes (NS3/4A, NS5A, NS5B), which plays an important role in the cleavage of the viral polyprotein as well as viral replication, resulting in high sustained virological response rates (> 90%) [311]. However, drug resistance-associated variants (RAVs) have emerged as a potential challenge for the treatment of HCV infection due to the high genetic variability in the HCV quasispecies population [93]. These RAVs can be present in treatment-naive patients as naturally occurring variants usually at low levels, although more frequently they are selected upon treatment failure and become the predominant variants (Figure 1). In chapter 7, we investigate the pre-existence of RAVs in the NS5B region (spanning amino acid positions 215 to 348 known to confer resistance to NS5B polymerase inhibitors) in the liver and plasma compartments of treatment naïve patients infected with HCV genotype 1b by 454-deep sequencing.

In our cohort of 18 paired treatment-naïve patients, the nucleoside inhibitor (NI) resistance-associated variant S282T was not detected not even as a minority variant either in plasma or liver tissue. S282T variant has rarely been detected *in vivo* due to the low replicative fitness of this variant [93]. For instance, in a recent

comprehensive NS5B gene analysis of 1,344 HCV isolates, S282T was present at frequencies of 0.17%, 0.24%, 1.24%, and 1.63%, in just one isolate for genotypes 1a, 1b, 3, and 4, respectively [289]. Nevertheless, the other NI resistant variant V321A was detected in 10 out of 18 patients (~56%) mainly as minor variant (Table S4). On the other hand, we noticed mutations that confer resistance to non-nucleoside inhibitors at position 316 (C316N/Y/H) in 16 out of 18 patients (~89%) either as dominant (50%) or minor variant (39%). Another study also found a high frequency of C316N as naturally occurring polymorphisms in never treated patients, especially in persons infected with HCV genotype 1b [312]. Overall, the quasispecies distribution in both plasma and liver samples showed a remarkable overlap in each patient (Chapter 7, Figure 2), underlining the use of plasma is most likely sufficient enough to detect HCV quasispecies and RAVs. In conclusion, although some of the variants in the present study confer only a low level of resistance, the presence of mutant HCV variants at baseline should be taken into consideration in the context of DAA therapy.

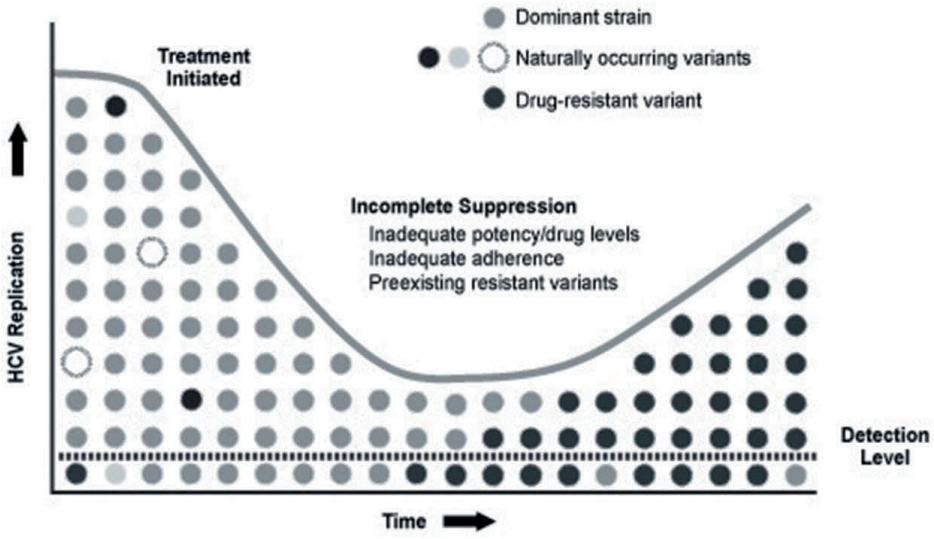



**Figure 1. Emergence and selection of drug resistant variants. Adopted from [93].**

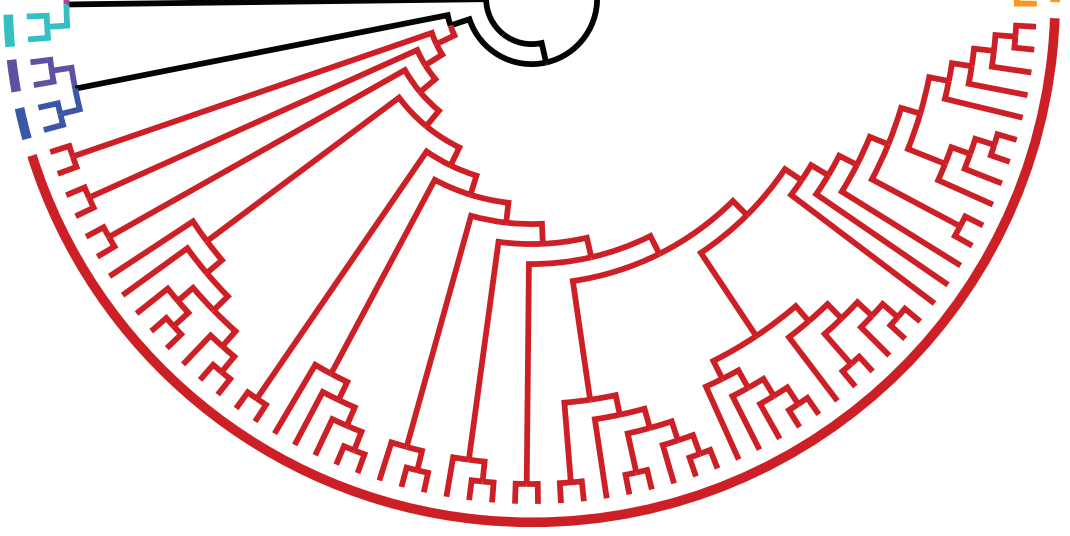
## Conclusion and future remarks

The work in this thesis describes insights into the molecular and genetic diversity of HBV and HCV in Ethiopia. This is the first such study from this region that can serve as a benchmark for future studies. Due to their largely asymptomatic nature, HBV and HCV infections remain a silent killer for decades. Most people are unaware of their infection and do not seek treatment until the disease has

progressed to serious liver damage. This problem is enormous in countries like Ethiopia where there is little awareness and knowledge of viral hepatitis both in the general population and even among key health policymakers. In Ethiopia, there is a widespread belief that hepatitis "yewof beshita", a liver disease manifests as yellow skin/eyes and dark urine, is caused by bat urine (bat flying overhead). A person with such symptoms is traditionally advised to go to traditional healers. As such, about 80% of the Ethiopian population still relies on traditional medicines for primary health cares. Hence, much more work is needed in this country to create awareness and knowledge at the national and regional levels regarding the impact of viral hepatitis.

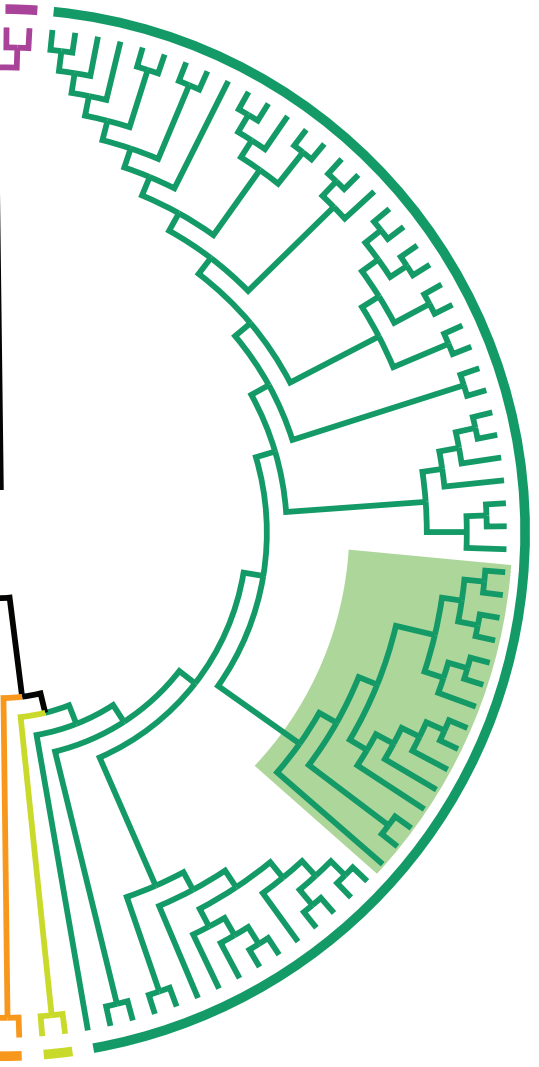
When we started this study in 2012, viral hepatitis has been largely ignored as a health priority in low-income countries, including Ethiopia. Nowadays, owing to the increasing disease burden, viral hepatitis is gaining better attention at the global level. The WHO declared (Glasgow Declaration on Hepatitis – September 2015) viral hepatitis as a serious public health threat and set a new global strategy to reduce new infections by 90% & mortality by 65%, by 2030. To achieve this goal the strategic framework defines a set of priority actions to be undertaken: increasing information system to understand viral hepatitis epidemics, expanding service coverage of testing and treatment, delivering hepatitis services through a public health package, financing to sustainability for universal health coverage, and accelerating innovations [301]. In response to WHO's call to action, Ethiopia also needs to prepare national policy and strategies in areas including surveillance, identification of risk group, diagnosis, treatment, resource mobilization, public awareness and advocacy. Such efforts need a strong political commitment and a high degree of attention equal to what has been relentlessly given to HIV, TB, and malaria. In addition, since the standard serological method cannot detect occult hepatitis B infection — HBsAg negative but viral DNA positive — Ethiopia should start incorporating nucleic acid testing in blood donor screening at national and regional blood banks.

Genotypes and genetic variants determination in chronic HBV and HCV infection are important in estimating disease progression and planning optimal antiviral treatments especially with the current direct-acting antivirals and pan-genotypic drugs. In Ethiopia, as shown in this study both HBV and HCV exhibit a wide range of genotypic diversity. Therefore, guidelines for the prevention, care, and treatment of HBV and HCV infection should take into account the genotypes that circulate in the country. In addition, the clinical impact of HBV and HCV genotypes in Ethiopia should be studied, which haven't been addressed in the present study.



# Chapter 9

## References



## REFERENCES


1 Blumberg, B. S. & Alter, H. J. A "new" antigen in leukemia sera. *JAMA* **191**, 541-546 (1965).

2 Prince, A. M. An antigen detected in the blood during the incubation period of serum hepatitis. *Proc Natl Acad Sci U S A* **60**, 814-821 (1968).

3 Farci, P. Choo QL, Kuo G, Weiner AJ, Overby LR, Bradley DW, Houghton M. Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome [Science 1989;244:359-362]. *J Hepatol* **36**, 582-585 (2002).

4 Khuroo, M. S. Discovery of hepatitis E: the epidemic non-A, non-B hepatitis 30 years down the memory lane. *Virus Res* **161**, 3-14 (2011).

5 Te, H. S. & Jensen, D. M. Epidemiology of hepatitis B and C viruses: a global overview. *Clin Liver Dis* **14**, 1-21, vii (2010).

6 Stanaway, J. D. *et al.* The global burden of viral hepatitis from 1990 to 2013: findings from the Global Burden of Disease Study 2013. *Lancet* **388**, 1081-1088 (2016).

7 Trepo, C., Chan, H. L. & Lok, A. Hepatitis B virus infection. *Lancet* **384**, 2053-2063 (2014).

8 Schweitzer, A., Horn, J., Mikolajczyk, R. T., Krause, G. & Ott, J. J. Estimations of worldwide prevalence of chronic hepatitis B virus infection: a systematic review of data published between 1965 and 2013. *Lancet* **386**, 1546-1555 (2015).

9 Datta, S., Chatterjee, S., Veer, V. & Chakravarty, R. Molecular Biology of the Hepatitis B Virus for Clinicians. *J Clin Exp Hepatol* **2**, 353-365, doi:http://dx.doi.org/10.1016/j.jceh.2012.10.003 (2012).

10 Revill, P., Testoni, B., Locarnini, S. & Zoulim, F. Global strategies are required to cure and eliminate HBV infection. *Nat Rev Gastroenterol Hepatol* **13**, 239-248 (2016).

11 Datta, S. An overview of molecular epidemiology of hepatitis B virus (HBV) in India. *Virol J* **5**, 156, doi:10.1186/1743-422x-5-156 (2008).

12 Block, T. M., Guo, H. & Guo, J.-T. Molecular Virology of Hepatitis B Virus for Clinicians. *Clinics in Liver Disease* **11**, 685-706, doi:http://dx.doi.org/10.1016/j.cld.2007.08.002 (2007).

13 Lamontagne, R. J., Bagga, S. & Bouchard, M. J. Hepatitis B virus molecular biology and pathogenesis. *Hepatoma Research* **2**, 163-186 (2016).

14 Perz, J. F., Armstrong, G. L., Farrington, L. A., Hutin, Y. J. & Bell, B. P. The contributions of hepatitis B virus and hepatitis C virus infections to cirrhosis and primary liver cancer worldwide. *J Hepatol* **45**, 529-538 (2006).

15 Te, H. S. & Jensen, D. M. Epidemiology of Hepatitis B and C Viruses: A Global Overview. *Clinics in Liver Disease* **14**, 1-21, doi:http://dx.doi.org/10.1016/j.cld.2009.11.009 (2010).

16 Jayaraman, S., Chalabi, Z., Perel, P., Guerriero, C. & Roberts, I. The risk of transfusion-transmitted infections in sub-Saharan Africa. *Transfusion* **50**, 433-442 (2010).

17 Bloch, E. M., Vermeulen, M. & Murphy, E. Blood transfusion safety in Africa: a literature review of infectious disease and organizational challenges. *Transfus Med Rev* **26**, 164-180 (2012).

18 Ott, J. J., Stevens, G. A., Groeger, J. & Wiersma, S. T. Global epidemiology of hepatitis B virus infection: New estimates of age-specific HBsAg seroprevalence and endemicity. *Vaccine* **30**, 2212-2219, doi:http://dx.doi.org/10.1016/j.vaccine.2011.12.116 (2012).

19 Drexler, J. F. *et al.* Bats carry pathogenic hepadnaviruses antigenically related to hepatitis B virus and capable of infecting human hepatocytes. *Proc Natl Acad Sci U S A* **110**, 16151-16156 (2013).

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

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

22 Liang, T. J. Hepatitis B: The virus and disease. *Hepatology* **49**, S13-S21, doi:10.1002/hep.22881 (2009).

23 Clark, D. N. & Hu, J. Unveiling the roles of HBV polymerase for new antiviral strategies. *Future Virol* **10**, 283-295 (2015).

24 Yan, H. *et al.* Sodium taurocholate cotransporting polypeptide is a functional receptor for human hepatitis B and D virus. *Elife* **3** (2012).

25 Balmasova, I. P. *et al.* Immunopathogenesis of chronic hepatitis B. *World J Gastroenterol* **20**, 14156-14171 (2014).

26 Locarnini, S. & Zoulim, F. Molecular genetics of HBV infection. *Antivir Ther* **15 Suppl 3**, 3-14 (2010).

27 Kramvis, A. Genotypes and genetic variability of hepatitis B virus. *Intervirology* **57**, 141-150 (2014).

28 Zhang, Z. H. *et al.* Genetic variation of hepatitis B virus and its significance for pathogenesis. *World J Gastroenterol* **22**, 126-144 (2016).

29 Norder, H., Courouce, A. M. & Magnius, L. O. Molecular basis of hepatitis B virus serotype variations within the four major subtypes. *J Gen Virol* **73 ( Pt 12)**, 3141-3145 (1992).

30 Norder, H. *et al.* Genetic diversity of hepatitis B virus strains derived worldwide: genotypes, subgenotypes, and HBsAg subtypes. *Intervirology* **47**, 289-309 (2004).

31 Okamoto, H. *et al.* Typing Hepatitis B Virus by Homology in Nucleotide Sequence: Comparison of Surface Antigen Subtypes. *J Gen Virol* **69**, 2575-2583, doi:doi:10.1099/0022-1317-69-10-2575 (1988).

32 Stuyver, L. *et al.* A new genotype of hepatitis B virus: complete genome and phylogenetic relatedness. *J Gen Virol* **81**, 67-74 (2000).

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

34 Tran, T. T., Trinh, T. N. & Abe, K. New complex recombinant genotype of hepatitis B virus identified in Vietnam. *J Virol* **82**, 5657-5663 (2008).

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

36 Tatematsu, K. *et al.* A genetic variant of hepatitis B virus divergent from known human and ape genotypes isolated from a Japanese patient and provisionally assigned to new genotype J. *J Virol* **83**, 10538-10547 (2009).

37 Kramvis, A. *et al.* Relationship of serological subtype, basic core promoter and precore mutations to genotypes/subgenotypes of hepatitis B virus. *J Med Virol* **80**, 27-46 (2008).

38 Abdou Chekaraou, M. *et al.* A novel hepatitis B virus (HBV) subgenotype D (D8) strain, resulting from recombination between genotypes D and E, is circulating in Niger along with HBV/E strains. *J Gen Virol* **91**, 1609-1620 (2010).

39 Meldal, B. H., Moula, N. M., Barnes, I. H., Boukef, K. & Allain, J. P. A novel hepatitis B virus subgenotype, D7, in Tunisian blood donors. *J Gen Virol* **90**, 1622-1628 (2009).

40 Olinger, C. M. *et al.* 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**, 1163-1173 (2006).

41 Ghosh, S. *et al.* New HBV subgenotype D9, a novel D/C recombinant, identified in patients with chronic HBeAg-negative infection in Eastern India. *J Viral Hepat* **20**, 209-218 (2013).

42 Paraskevis, D. *et al.* Dating the origin and dispersal of hepatitis B virus infection in humans and primates. *Hepatology* **57**, 908-916 (2013).

43 Araujo, N. M. Hepatitis B virus intergenotypic recombinants worldwide: An overview. *Infect Genet Evol* **36**, 500-510 (2015).

44 Sugiyama, M. *et al.* Influence of hepatitis B virus genotypes on the intra- and extracellular expression of viral DNA and antigens. *Hepatology* **44**, 915-924 (2006).

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

46 Shi, W. *et al.* Identification of novel inter-genotypic recombinants of human hepatitis B viruses by large-scale phylogenetic analysis. *Virology* **427**, 51-59 (2012).

47 Chen, X. *et al.* A novel B/C inter-genotype recombinant of hepatitis B virus identified in north-west China. *J Gen Virol* **95**, 153-155 (2014).

48 Kojima, Y. *et al.* Identification of Novel Recombinant Forms of Hepatitis B Virus Generated from Genotypes Ae and G in HIV-1-Positive Japanese Men Who Have Sex with Men. *AIDS Res Hum Retroviruses* **31**, 760-767 (2015).

49 Kurbanov, F. *et al.* A new subtype (subgenotype) Ac (A3) of hepatitis B virus and recombination between genotypes A and E in Cameroon. *J Gen Virol* **86**, 2047-2056 (2005).

50 Mahgoub, S., Candotti, D., El Ekiaby, M. & Allain, J. P. Hepatitis B virus (HBV) infection and recombination between HBV genotypes D and E in asymptomatic blood donors from Khartoum, Sudan. *J Clin Microbiol* **49**, 298-306 (2011).

51 Cui, C. *et al.* The dominant hepatitis B virus genotype identified in Tibet is a C/D hybrid. *J Gen Virol* **83**, 2773-2777 (2002).

52 Kurbanov, F., Tanaka, Y. & Mizokami, M. Geographical and genetic diversity of the human hepatitis B virus. *Hepatol Res* **40**, 14-30 (2010).

53 Pourkarim, M. R., Amini-Bavil-Olyaee, S., Kurbanov, F., Van Ranst, M. & Tacke, F. Molecular identification of hepatitis B virus genotypes/subgenotypes: revised classification hurdles and updated resolutions. *World J Gastroenterol* **20**, 7152-7168 (2014).

54 Cao, G. W. Clinical relevance and public health significance of hepatitis B virus genomic variations. *World J Gastroenterol* **15**, 5761-5769 (2009).

55 Thedja, M. D. *et al.* Ethnogeographical structure of hepatitis B virus genotype distribution in Indonesia and discovery of a new subgenotype, B9. *Arch Virol* **156**, 855-868 (2011).

56 Yousif, M. & Kramvis, A. Genotype D of hepatitis B virus and its subgenotypes: An update. *Hepatol Res* **43**, 355-364 (2013).

57 Shi, Y. H. Correlation between hepatitis B virus genotypes and clinical outcomes. *Jpn J Infect Dis* **65**, 476-482 (2012).

58 McMahon, B. J. The influence of hepatitis B virus genotype and subgenotype on the natural history of chronic hepatitis B. *Hepatol Int* **3**, 334-342 (2009).

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

60 Kim, B. K., Revill, P. A. & Ahn, S. H. HBV genotypes: relevance to natural history, pathogenesis and treatment of chronic hepatitis B. *Antivir Ther* **16**, 1169-1186 (2011).

61 Janssen, H. L. *et al.* Pegylated interferon alfa-2b alone or in combination with lamivudine for HBeAg-positive chronic hepatitis B: a randomised trial. *Lancet* **365**, 123-129 (2005).

62 Kong, L. N., Qin, B., Ma, Q., Li, L. & Yao, Y. Relationship between hepatitis B virus genotype B and C and response to interferon therapy in HBeAg positive chronic hepatitis B patients: A meta-analysis. *J Gastroenterol Hepatol* **29**, 1387-1395 (2014).

63 Pollicino, T., Cacciola, I., Saffioti, F. & Raimondo, G. Hepatitis B virus PreS/S gene variants: pathobiology and clinical implications. *J Hepatol* **61**, 408-417 (2014).

64 Mina, T. *et al.* Genomic Diversity of Hepatitis B Virus Infection Associated With Fulminant Hepatitis B Development. *Hepat Mon* **15**, e29477 (2015).

65 Caligiuri, P., Cerruti, R., Icardi, G. & Bruzzone, B. Overview of hepatitis B virus mutations and their implications in the management of infection. *World J Gastroenterol* **22**, 145-154 (2016).

66 Hollinger, F. B. Hepatitis B virus genetic diversity and its impact on diagnostic assays. *J Viral Hepat* **14 Suppl 1**, 11-15 (2007).

67 El Chaar, M., Candotti, D., Crowther, R. A. & Allain, J. P. Impact of hepatitis B virus surface protein mutations on the diagnosis of occult hepatitis B virus infection. *Hepatology* **52**, 1600-1610 (2010).

68 Tong, S., Kim, K. H., Chante, C., Wands, J. & Li, J. Hepatitis B Virus e Antigen Variants. *Int J Med Sci* **2**, 2-7 (2005).

69 Sonneveld, M. J. *et al.* Presence of precore and core promoter mutants limits the probability of response to peginterferon in hepatitis B e antigen-positive chronic hepatitis B. *Hepatology* **56**, 67-75 (2012).

70 Buti, M., Rodriguez-Frias, F., Jardi, R. & Esteban, R. Hepatitis B virus genome variability and disease progression: the impact of pre-core mutants and HBV genotypes. *J Clin Virol* **34 Suppl 1**, S79-82 (2005).

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

72 Gomes-Gouvea, M. S. *et al.* HBV carrying drug-resistance mutations in chronically infected treatment-naive patients. *Antivir Ther* **20**, 387-395 (2015).

73 Zoulim, F. & Locarnini, S. Hepatitis B virus resistance to nucleos(t)ide analogues. *Gastroenterology* **137**, 1593-1608 e1591-1592 (2009).

74 Jardi, R. *et al.* Hepatitis B virus polymerase variants associated with entecavir drug resistance in treatment-naive patients. *J Viral Hepat* **14**, 835-840 (2007).

75 Mohd Hanafiah, K., Groeger, J., Flaxman, A. D. & Wiersma, S. T. Global epidemiology of hepatitis C virus infection: new estimates of age-specific antibody to HCV seroprevalence. *Hepatology* **57**, 1333-1342 (2013).

76 Bukh, J. The history of hepatitis C virus (HCV): Basic research reveals unique features in phylogeny, evolution and the viral life cycle with new perspectives for epidemic control. *J Hepatol* **65**, S2-S21 (2016).

77 Lozano, R. *et al.* Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* **380**, 2095-2128 (2012).

78 Lavanchy, D. Evolving epidemiology of hepatitis C virus. *Clin Microbiol Infect* **17**, 107-115 (2011).

79 Castello, G., Scala, S., Palmieri, G., Curley, S. A. & Izzo, F. HCV-related hepatocellular carcinoma: From chronic inflammation to cancer. *Clin Immunol* **134**, 237-250 (2010).

80 Antiretroviral Therapy Cohort Collaboration. Causes of death in HIV-1-infected patients treated with antiretroviral therapy, 1996-2006: collaborative analysis of 13 HIV cohort studies. *Clin Infect Dis* **50**, 1387-1396 (2010).

81 Bostan, N. & Mahmood, T. An overview about hepatitis C: a devastating virus. *Crit Rev Microbiol* **36**, 91-133 (2010).

82 Alter, M. J. Epidemiology of hepatitis C virus infection. *World J Gastroenterol* **13**, 2436-2441 (2007).

83 Pfaender, S., Brown, R. J., Pietschmann, T. & Steinmann, E. Natural reservoirs for homologs of hepatitis C virus. *Emerg Microbes Infect* **3**, e21 (2014).

84 Houghton, M. Discovery of the hepatitis C virus. *Liver Int* **29 Suppl 1**, 82-88 (2009).

85 Pybus, O. G. & Theze, J. Hepacivirus cross-species transmission and the origins of the hepatitis C virus. *Curr Opin Virol* **16**, 1-7 (2016).

86 Smith, D. B. *et al.* Proposed update to the taxonomy of the genera Hepacivirus and Pegivirus within the Flaviviridae family. *J Gen Virol* **97**, 2894-2907 (2016).

87 Catanese, M. T. *et al.* Ultrastructural analysis of hepatitis C virus particles. *Proc Natl Acad Sci U S A* **110**, 9505-9510 (2013).

88 Sharma, S. D. Hepatitis C virus: molecular biology & current therapeutic options. *Indian J Med Res* **131**, 17-34 (2010).

89 Moradpour, D. & Penin, F. Hepatitis C virus proteins: from structure to function. *Curr Top Microbiol Immunol* **369**, 113-142 (2013).

90 Myrmel, H., Ulvestad, E. & Asjo, B. The hepatitis C virus enigma. *Apmis* **117**, 427-439 (2009).

91 Phan, T., Beran, R. K., Peters, C., Lorenz, I. C. & Lindenbach, B. D. Hepatitis C virus NS2 protein contributes to virus particle assembly via opposing epistatic interactions with the E1-E2 glycoprotein and NS3-NS4A enzyme complexes. *J Virol* **83**, 8379-8395 (2009).

92 Horner, S. M. & Gale, M., Jr. Regulation of hepatic innate immunity by hepatitis C virus. *Nat Med* **19**, 879-888 (2013).

93 Esposito, I., Trinks, J. & Soriano, V. Hepatitis C virus resistance to the new direct-acting antivirals. *Expert Opin Drug Metab Toxicol* **12**, 1197-1209 (2016).

94 Smith, D. B. *et al.* Expanded classification of hepatitis C virus into 7 genotypes and 67 subtypes: updated criteria and genotype assignment web resource. *Hepatology* **59**, 318-327 (2014).

95 Simmonds, P. *et al.* Consensus proposals for a unified system of nomenclature of hepatitis C virus genotypes. *Hepatology* **42**, 962-973 (2005).

96 Messina, J. P. *et al.* Global distribution and prevalence of hepatitis C virus genotypes. *Hepatology* **61**, 77-87 (2015).

97 Domingo, E., Sheldon, J. & Perales, C. Viral quasispecies evolution. *Microbiol Mol Biol Rev* **76**, 159-216 (2012).

98 Hnatyszyn, H. J. Chronic hepatitis C and genotyping: the clinical significance of determining HCV genotypes. *Antivir Ther* **10**, 1-11 (2005).

99 Galli, A. & Bukh, J. Comparative analysis of the molecular mechanisms of recombination in hepatitis C virus. *Trends Microbiol* **22**, 354-364 (2014).

100 Iles, J. C. *et al.* Characterization of Hepatitis C Virus Recombination in Cameroon by Use of Nonspecific Next-Generation Sequencing. *J Clin Microbiol* **53**, 3155-3164 (2015).

101 Gonzalez-Candelas, F., Lopez-Labrador, F. X. & Bracho, M. A. Recombination in hepatitis C virus. *Viruses* **3**, 2006-2024 (2011).

102 Gower, E., Estes, C., Blach, S., Razavi-Shearer, K. & Razavi, H. Global epidemiology and genotype distribution of the hepatitis C virus infection. *J Hepatol* **61**, S45-57 (2014).

103 Murphy, D. G. *et al.* Hepatitis C virus genotype 7, a new genotype originating from central Africa. *J Clin Microbiol* **53**, 967-972 (2015).

104 Markov, P. V. *et al.* Colonial history and contemporary transmission shape the genetic diversity of hepatitis C virus genotype 2 in Amsterdam. *J Virol* **86**, 7677-7687 (2012).

105 Iles, J. C. *et al.* Phylogeography and epidemic history of hepatitis C virus genotype 4 in Africa. *Virology* **464-465**, 233-243 (2014).

106 Markov, P. V. *et al.* Phylogeography and molecular epidemiology of hepatitis C virus genotype 2 in Africa. *J Gen Virol* **90**, 2086-2096 (2009).

107 Pybus, O. G. *et al.* Genetic history of hepatitis C virus in East Asia. *J Virol* **83**, 1071-1082 (2009).

108 Magiorkinis, G. *et al.* The global spread of hepatitis C virus 1a and 1b: a phylodynamic and phylogeographic analysis. *PLoS Med* **6**, e1000198 (2009).

109 Li, C., Lu, L., Murphy, D. G., Negro, F. & Okamoto, H. Origin of hepatitis C virus genotype 3 in Africa as estimated through an evolutionary analysis of the full-length genomes of nine subtypes, including the newly sequenced 3d and 3e. *J Gen Virol* **95**, 1677-1688 (2014).

110 Verbeeck, J. *et al.* Investigating the origin and spread of hepatitis C virus genotype 5a. *J Virol* **80**, 4220-4226 (2006).

111 Tsega, E. Current views on liver diseases in Ethiopia. *Ethiop Med J* **15**, 75-82 (1977).

112 Tsega, E., Nordenfelt, E., Hansson, B. G., Mengesha, B. & Lindberg, J. Chronic liver disease in Ethiopia: a clinical study with emphasis on identifying common causes. *Ethiop Med J* **30**, 1-33 (1992).

113 Belyhun, Y., Maier, M., Mulu, A., Diro, E. & Liebert, U. G. Hepatitis viruses in Ethiopia: a systematic review and meta-analysis. *BMC Infect Dis* **16**, 761 (2016).

114 Shimelis, T. *et al.* Hepatitis B virus infection among people attending the voluntary counselling and testing centre and anti-retroviral therapy clinic of St Paul's General Specialised Hospital, Addis Ababa, Ethiopia. *Sex Transm Infect* **84**, 37-41 (2008).

115 Abebe, A. *et al.* Seroepidemiology of hepatitis B virus in Addis Ababa, Ethiopia: transmission patterns and vaccine control. *Epidemiol Infect* **131**, 757-770 (2003).

116 Bisetegen, F. S., Bekele, F. B., Ageru, T. A. & Wada, F. W. Transfusion-Transmissible Infections among Voluntary Blood Donors at Wolaita Sodo University Teaching Referral Hospital, South Ethiopia. *Can J Infect Dis Med Microbiol* **2016**, 8254343 (2016).

117 Gebreegziabher, D., Asfeha, G. G. & Gebreyesus, H. A. Seroprevalence of hepatitis B virus surface antigen (HBsAg) among clients visiting 'Tefera Hailu' memorial hospital, Sekota, Northern Ethiopia. *BMC Infect Dis* **16**, 383 (2016).

118 Kefene, H. *et al.* Ethiopian national hepatitis B study. *Journal of Medical Virology* **24**, 75-84, doi:10.1002/jmv.1890240110 (1988).

119 Tessema, B. *et al.* Seroprevalence of HIV, HBV, HCV and syphilis infections among blood donors at Gondar University Teaching Hospital, Northwest Ethiopia: declining trends over a period of five years. *BMC Infect Dis* **10**, 111 (2010).

120 Mohammed, Y. & Bekele, A. Seroprevalence of transfusion transmitted infection among blood donors at Jijiga blood bank, Eastern Ethiopia: retrospective 4 years study. *BMC Res Notes* **9**, 129 (2016).

121 Abdo, A. E., Mohammed, D. A. & Satti, M. Prevalence of hepatitis B virus among blood donors and assessment of blood donor's knowledge about HBV in Sudan. . *HIV/AIDS Res Treat Open J.* **2**, 76-80 (2015).

122 Fessehaye, N., Naik, D. & Fessehaye, T. Transfusion transmitted infections - a retrospective analysis from the National Blood Transfusion Service in Eritrea. *Pan Afr Med J* **9**, 40 (2011).

123 Babanejad, M., Izadi, N., Najafi, F. & Alavian, S. M. The HBsAg Prevalence Among Blood Donors From Eastern Mediterranean and Middle Eastern Countries: A Systematic Review and Meta-Analysis. *Hepat Mon* **16**, e35664 (2016).

124 Baha, W. *et al.* Prevalence and risk factors of hepatitis B and C virus infections among the general population and blood donors in Morocco. *BMC Public Health* **13**, 50 (2013).

125 Yong-Lin, Y. *et al.* Hepatitis B surface antigen variants in voluntary blood donors in Nanjing, China. *Virol J* **9**, 82 (2012).

126 Behal, R. *et al.* Seroprevalence and risk factors for hepatitis B virus infection among general population in Northern India. *Arq Gastroenterol* **45**, 137-140 (2008).

127 Zou, S. *et al.* Current incidence and residual risk of hepatitis B infection among blood donors in the United States. *Transfusion* **49**, 1609-1620 (2009).

128 Rao, V. B. *et al.* Hepatitis C seroprevalence and HIV co-infection in sub-Saharan Africa: a systematic review and meta-analysis. *Lancet Infect Dis* **15**, 819-824 (2015).

129 Konstantinou, D. & Deutsch, M. The spectrum of HBV/HCV coinfection: epidemiology, clinical characteristics, viralinteractions and management. *Ann Gastroenterol* **28**, 221-228 (2015).

130 Blumberg, B. S. Sex differences in response to hepatitis b virus. *Arthritis & Rheumatism* **22**, 1261-1266, doi:10.1002/art.1780221114 (1979).

131 Yuen, M. F., Hou, J. L., Chutaputti, A. & Asia Pacific Working Party on Prevention of Hepatocellular, C. Hepatocellular carcinoma in the Asia pacific region. *J Gastroenterol Hepatol* **24**, 346-353 (2009).

132 WHO. *Hepatitis B. Fact sheet N°204*, 2015).

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

134 WHO. *Department of Communicable Diseases Surveillance and Response. Hepatitis B.*, <http://www.who.int/csr/disease/hepatitis/HepatitisB_whocdscsrlyo2002_2.pdf> (2002).

135 Rahlenbeck, S. I., Yohannes, G., Molla, K., Reifen, R. & Assefa, A. Infection with HIV, syphilis and hepatitis B in Ethiopia: a survey in blood donors. *Int J STD AIDS* **8**, 261-264 (1997).

136 Yu, H. *et al.* Molecular and phylogenetic analyses suggest an additional hepatitis B virus genotype "I". *PLoS One* **5**, e9297 (2010).

137 Tian, Y. *et al.* The amino Acid residues at positions 120 to 123 are crucial for the antigenicity of hepatitis B surface antigen. *J Clin Microbiol* **45**, 2971-2978 (2007).

138 Wu, C. *et al.* Biological significance of amino acid substitutions in hepatitis B surface antigen (HBsAg) for glycosylation, secretion, antigenicity and immunogenicity of HBsAg and hepatitis B virus replication. *J Gen Virol* **91**, 483-492 (2010).

139 Sayiner, A. A., Ozcan, A. & Sengonul, A. Naturally occurring MHR variants in Turkish patients infected with hepatitis B virus. *J Med Virol* **80**, 405-410 (2008).

140 Pourkarim, M. R. *et al.* Evolutionary analysis of HBV "S" antigen genetic diversity in Iranian blood donors: a nationwide study. *J Med Virol* **86**, 144-155 (2014).

141 Pas, S. D., Fries, E., De Man, R. A., Osterhaus, A. D. & Niesters, H. G. Development of a quantitative real-time detection assay for hepatitis B virus DNA and comparison with two commercial assays. *J Clin Microbiol* **38**, 2897-2901 (2000).

142 Pas, S. D. & Niesters, H. G. Detection of HBV DNA using real time analysis. *J Clin Virol* **25**, 93-94 (2002).

143 Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* **30**, 2725-2729 (2013).

144 Schultz, A. K. *et al.* jpHMM: recombination analysis in viruses with circular genomes such as the hepatitis B virus. *Nucleic Acids Res* **40**, W193-198 (2012).

145 Purdy, M. A., Talekar, G., Swenson, P., Araujo, A. & Fields, H. A new algorithm for deduction of hepatitis B surface antigen subtype determinants from the amino acid sequence. *Intervirology* **50**, 45-51 (2007).

146 Warner, N. & Locarnini, S. The antiviral drug selected hepatitis B virus rtA181T/sW172* mutant has a dominant negative secretion defect and alters the typical profile of viral rebound. *Hepatology* **48**, 88-98 (2008).

147 Kao, J. H. Hepatitis B viral genotypes: clinical relevance and molecular characteristics. *J Gastroenterol Hepatol* **17**, 643-650 (2002).

148 Chu, C. J. & Lok, A. S. Clinical significance of hepatitis B virus genotypes. *Hepatology* **35**, 1274-1276 (2002).

149 Mwangi, J. *et al.* Molecular genetic diversity of hepatitis B virus in Kenya. *Intervirology* **51**, 417-421 (2008).

150 Kwange, S. O. *et al.* Hepatitis B virus subgenotype A1, occurrence of subgenotype D4, and S gene mutations among voluntary blood donors in Kenya. *Virus Genes* **47**, 448-455 (2013).

151 Yousif, M., Mudawi, H., Bakhiet, S., Glebe, D. & Kramvis, A. Molecular characterization of hepatitis B virus in liver disease patients and asymptomatic carriers of the virus in Sudan. *BMC Infect Dis* **13**, 328 (2013).

152 Pourkarim, M. R., Amini-Bavil-Olyaee, S., Lemey, P., Maes, P. & Van Ranst, M. Are hepatitis B virus "subgenotypes" defined accurately? *J Clin Virol* **47**, 356-360 (2010).

153 Pourkarim, M. R., Amini-Bavil-Olyaee, S., Lemey, P., Maes, P. & Van Ranst, M. HBV subgenotype misclassification expands quasi-subgenotype A3. *Clin Microbiol Infect* **17**, 947-949 (2011).

154 Kramvis, A. & Paraskevis, D. Subgenotype A1 of HBV--tracing human migrations in and out of Africa. *Antivir Ther* **18**, 513-521 (2013).

155 Hubschen, J. M. *et al.* Exceptional genetic variability of hepatitis B virus indicates that Rwanda is east of an emerging African genotype E/A1 divide. *J Med Virol* **81**, 435-440 (2009).

156 Gulube, Z. *et al.* Molecular characterization of hepatitis B virus isolates from Zimbabwean blood donors. *J Med Virol* **83**, 235-244 (2011).

157 Tallo, T. *et al.* D2: major subgenotype of hepatitis B virus in Russia and the Baltic region. *J Gen Virol* **89**, 1829-1839 (2008).

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

159 Echevarria, J. M. & Avellon, A. Hepatitis B virus genetic diversity. *J Med Virol* **78 Suppl 1**, S36-42 (2006).

160 Stramer, S. L. *et al.* Nucleic acid testing to detect HBV infection in blood donors. *N Engl J Med* **364**, 236-247 (2011).

161 Cooreman, M. P., Leroux-Roels, G. & Paulij, W. P. Vaccine- and hepatitis B immune globulin-induced escape mutations of hepatitis B virus surface antigen. *J Biomed Sci* **8**, 237-247 (2001).

162 Coleman, P. F., Chen, Y. C. & Mushahwar, I. K. Immunoassay detection of hepatitis B surface antigen mutants. *J Med Virol* **59**, 19-24 (1999).

163 Ho, M. *et al.* Patterns of circulating hepatitis B surface antigen variants among vaccinated children born to hepatitis B surface antigen carrier and non-carrier mothers. A population-based comparative study. *J Biomed Sci* **5**, 355-362 (1998).

164 Avellon, A. & Echevarria, J. M. Frequency of hepatitis B virus 'a' determinant variants in unselected Spanish chronic carriers. *J Med Virol* **78**, 24-36 (2006).

165 Carman, W. F. The clinical significance of surface antigen variants of hepatitis B virus. *J Viral Hepat* **4 Suppl 1**, 11-20 (1997).

166 Hou, J. *et al.* Prevalence of naturally occurring surface gene variants of hepatitis B virus in nonimmunized surface antigen-negative Chinese carriers. *Hepatology* **34**, 1027-1034 (2001).

167 Locarnini, S. Primary resistance, multidrug resistance, and cross-resistance pathways in HBV as a consequence of treatment failure. *Hepatol Int* **2**, 147-151 (2008).

168 Zheng, J. *et al.* Prevalence and significance of Hepatitis B reverse transcriptase mutants in different disease stages of untreated patients. *Liver Int* **32**, 1535-1542 (2012).

169 Hundie, G. B. *et al.* Molecular epidemiology and genetic diversity of hepatitis B virus in Ethiopia. *J Med Virol* (2015).

170 Hubschen, J. M. *et al.* Detection of a new subgenotype of hepatitis B virus genotype A in Cameroon but not in neighbouring Nigeria. *Clin Microbiol Infect* **17**, 88-94 (2011).

171 Haile-Selassie, Y. Late Miocene hominids from the Middle Awash, Ethiopia. *Nature* **412**, 178-181 (2001).

172 White, T. D. *et al.* Ardipithecus ramidus and the paleobiology of early hominids. *Science* **326**, 75-86 (2009).

173 Johanson, D. C. & White, T. D. A systematic assessment of early African hominids. *Science* **203**, 321-330 (1979).

174 McDougall, I., Brown, F. H. & Fleagle, J. G. Stratigraphic placement and age of modern humans from Kibish, Ethiopia. *Nature* **433**, 733-736 (2005).

175 White, T. D. *et al.* Pleistocene Homo sapiens from Middle Awash, Ethiopia. *Nature* **423**, 742-747 (2003).

176 Beggel, B. *et al.* Full genome ultra-deep pyrosequencing associates G-to-A hypermutation of the hepatitis B virus genome with the natural progression of hepatitis B. *J Viral Hepat* **20**, 882-889 (2013).

177 Martin, D. & Rybicki, E. RDP: detection of recombination amongst aligned sequences. *Bioinformatics* **16**, 562-563 (2000).

178 Lole, K. S. *et al.* Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J Virol* **73**, 152-160 (1999).

179 Andernach, I. E., Nolte, C., Pape, J. W. & Muller, C. P. Slave trade and hepatitis B virus genotypes and subgenotypes in Haiti and Africa. *Emerg Infect Dis* **15**, 1222-1228 (2009).

180 Lago, B. V., Mello, F. C., Kramvis, A., Niel, C. & Gomes, S. A. Hepatitis B virus subgenotype A1: evolutionary relationships between Brazilian, African and Asian isolates. *PLoS One* **9**, e105317 (2014).

181 Chu, C. J. *et al.* Prevalence of HBV precore/core promoter variants in the United States. *Hepatology* **38**, 619-628 (2003).

182 Chandra, P. K. *et al.* 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**, 749-756 (2009).

183 Tacke, F. *et al.* Basal core promoter and precore mutations in the hepatitis B virus genome enhance replication efficacy of Lamivudine-resistant mutants. *J Virol* **78**, 8524-8535 (2004).

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

185 Hu, F., Bi, S., Yan, H., Shi, Y. & Sheng, J. Associations between hepatitis B virus basal core promoter/pre-core region mutations and the risk of acute-on-chronic liver failure: a meta-analysis. *Virol J* **12**, 87 (2015).

186 Fang, Z. L. *et al.* Hepatitis B virus pre-S deletion mutations are a risk factor for hepatocellular carcinoma: a matched nested case-control study. *J Gen Virol* **89**, 2882-2890 (2008).

187 Qu, L. S. *et al.* Association of hepatitis B virus pre-S deletions with the development of hepatocellular carcinoma in Qidong, China. *PLoS One* **9**, e98257 (2014).

188 de Oliveria Andrade, L. J. *et al.* Association between hepatitis C and hepatocellular carcinoma. *J Glob Infect Dis* **1**, 33-37 (2009).

189 Deuffic-Burban, S., Poynard, T., Sulkowski, M. S. & Wong, J. B. Estimating the future health burden of chronic hepatitis C and human immunodeficiency virus infections in the United States. *J Viral Hepat* **14**, 107-115 (2007).

190 Jackowiak, P. *et al.* Phylogeny and molecular evolution of the hepatitis C virus. *Infect Genet Evol* **21**, 67-82 (2014).

191 de Bruijne, J. *et al.* Emergence of hepatitis C virus genotype 4: phylogenetic analysis reveals three distinct epidemiological profiles. *J Clin Microbiol* **47**, 3832-3838 (2009).

192 Njouom, R. *et al.* Phylogeography, risk factors and genetic history of hepatitis C virus in Gabon, central Africa. *PLoS One* **7**, e42002 (2012).

193 Nopporpanth, S. *et al.* Complete genome analysis of hepatitis C virus subtypes 6t and 6u. *J Gen Virol* **89**, 1276-1281 (2008).

194 Frommel, D. *et al.* A survey of antibodies to hepatitis C virus in Ethiopia. *Am J Trop Med Hyg* **49**, 435-439 (1993).

195 Hundie, G. B., Raj, V. S., Gebre Michael, D. & Haagmans, B. L. Seroepidemiology of hepatitis B and C virus infections among blood donors in Ethiopia. *J Med Virol* (2017).

196 Candotti, D., Temple, J., Sarkodie, F. & Allain, J. P. Frequent recovery and broad genotype 2 diversity characterize hepatitis C virus infection in Ghana, West Africa. *J Virol* **77**, 7914-7923 (2003).

197 Iles, J. C. *et al.* Phylogeography and epidemic history of hepatitis C virus genotype 4 in Africa. *Virology* **464–465**, 233-243, doi:http://dx.doi.org/10.1016/j.virol.2014.07.006 (2014).

198 Abreha, T. *et al.* Genotypes and viral load of hepatitis C virus among persons attending a voluntary counseling and testing center in Ethiopia. *J Med Virol* **83**, 776-782 (2011).

199 Mellor, J., Holmes, E. C., Jarvis, L. M., Yap, P. L. & Simmonds, P. Investigation of the pattern of hepatitis C virus sequence diversity in different geographical regions: implications for virus classification. The International HCV Collaborative Study Group. *J Gen Virol* **76 ( Pt 10)**, 2493-2507 (1995).

200 Sandres-Saune, K. *et al.* Determining hepatitis C genotype by analyzing the sequence of the NS5b region. *J Virol Methods* **109**, 187-193 (2003).

201 Ronquist, F. *et al.* MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* **61**, 539-542 (2012).

202 Milne, I. *et al.* TOPALi v2: a rich graphical interface for evolutionary analyses of multiple alignments on HPC clusters and multi-core desktops. *Bioinformatics* **25**, 126-127 (2009).

203 Rose, R., Markov, P. V., Lam, T. T. & Pybus, O. G. Viral evolution explains the associations among hepatitis C virus genotype, clinical outcomes, and human genetic variation. *Infect Genet Evol* **20**, 418-421 (2013).

204 Kamal, S. M. & Nasser, I. A. Hepatitis C genotype 4: What we know and what we don't yet know. *Hepatology* **47**, 1371-1383 (2008).

205 Ndong-Atome, G. R. *et al.* High prevalence of hepatitis C virus infection and predominance of genotype 4 in rural Gabon. *J Med Virol* **80**, 1581-1587 (2008).

206 Ciccozzi, M. *et al.* Hepatitis C virus genotype 4d in Southern Italy: reconstruction of its origin and spread by a phylodynamic analysis. *J Med Virol* **84**, 1613-1619 (2012).

207 Kamal, S. M. Hepatitis C virus genotype 4 therapy: progress and challenges. *Liver Int* **31 Suppl 1**, 45-52 (2011).

208 Khattab, M. A. *et al.* Management of hepatitis C virus genotype 4: recommendations of an international expert panel. *J Hepatol* **54**, 1250-1262 (2011).

209 Shire, A. M. *et al.* Viral hepatitis among Somali immigrants in Minnesota: association of hepatitis C with hepatocellular carcinoma. *Mayo Clin Proc* **87**, 17-24 (2012).

210 Wantuck, J. M., Ahmed, A. & Nguyen, M. H. Review article: the epidemiology and therapy of chronic hepatitis C genotypes 4, 5 and 6. *Aliment Pharmacol Ther* **39**, 137-147 (2014).

211 Abdel-Ghaffar, T. Y., Sira, M. M. & El Naghi, S. Hepatitis C genotype 4: The past, present, and future. *World J Hepatol* **7**, 2792-2810 (2015).

212 Al Ashgar, H. I. *et al.* Hepatitis C genotype 4: genotypic diversity, epidemiological profile, and clinical relevance of subtypes in Saudi Arabia. *Saudi J Gastroenterol* **19**, 28-33 (2013).

213 Pasquier, C. *et al.* Distribution and heterogeneity of hepatitis C genotypes in hepatitis patients in Cameroon. *J Med Virol* **77**, 390-398 (2005).

214 Njouom, R. *et al.* Predominance of hepatitis C virus genotype 4 infection and rapid transmission between 1935 and 1965 in the Central African Republic. *J Gen Virol* **90**, 2452-2456 (2009).

215 Cantaloube, J. F. *et al.* Analysis of hepatitis C virus strains circulating in Republic of the Congo. *J Med Virol* **82**, 562-567 (2010).

216 Purdy, M. A. *et al.* A re-evaluation of the origin of hepatitis C virus genotype 2 in West Africa. *Journal of General Virology* **96**, 2157-2164, doi:doi:10.1099/vir.0.000153 (2015).

217 Hundie, G. B. *et al.* A novel hepatitis B virus subgenotype D10 circulating in Ethiopia. *J Viral Hepat* **24**, 163-173 (2017).

218 Noppornpanth, S. *et al.* Identification of a naturally occurring recombinant genotype 2/6 hepatitis C virus. *J Virol* **80**, 7569-7577 (2006).

219 Bull, R. A. *et al.* A method for near full-length amplification and sequencing for six hepatitis C virus genotypes. *BMC Genomics* **17**, 247 (2016).

220 Gao, S., Duan, Z. P. & Coffin, C. S. Clinical relevance of hepatitis B virus variants. *World J Hepatol* **7**, 1086-1096 (2015).

221 Rybicka, M., Stalke, P. & Bielawski, K. P. Current molecular methods for the detection of hepatitis B virus quasispecies. *Rev Med Virol* **26**, 369-381 (2016).

222 Chotiyaputta, W. & Lok, A. S. Hepatitis B virus variants. *Nat Rev Gastroenterol Hepatol* **6**, 453-462 (2009).

223 Cao, L. *et al.* Coexistence of hepatitis B virus quasispecies enhances viral replication and the ability to induce host antibody and cellular immune responses. *J Virol* **88**, 8656-8666 (2014).

224 Li, F. *et al.* Whole genome characterization of hepatitis B virus quasispecies with massively parallel pyrosequencing. *Clin Microbiol Infect* **21**, 280-287 (2015).

225 Coffin, C. S. *et al.* Hepatitis B virus quasispecies in hepatic and extrahepatic viral reservoirs in liver transplant recipients on prophylactic therapy. *Liver Transpl* **17**, 955-962 (2011).

226 Wei, F., Zheng, Q., Li, M. & Wu, M. The association between hepatitis B mutants and hepatocellular carcinoma: A meta-analysis. *Medicine (Baltimore)* **96**, e6835 (2017).

227 Malik, A., Singhal, D. K., Albanyan, A., Husain, S. A. & Kar, P. Hepatitis B virus gene mutations in liver diseases: a report from New Delhi. *PLoS One* **7**, e39028 (2012).

228 Shen, T. & Yan, X. M. Hepatitis B virus genetic mutations and evolution in liver diseases. *World J Gastroenterol* **20**, 5435-5441 (2014).

229 Kim, D. W., Lee, S. A., Hwang, E. S., Kook, Y. H. & Kim, B. J. Naturally occurring precore/core region mutations of hepatitis B virus genotype C related to hepatocellular carcinoma. *PLoS One* **7**, e47372 (2012).

230 Keshvari, M., Alavian, S. M. & Sharafi, H. Comparison of Serum Hepatitis B Virus DNA and HBsAg Levels Between HBeAg-Negative and HBeAg-Positive Chronic Hepatitis B Patients. *Jundishapur J Microbiol* **8**, e21444 (2015).

231 Chen, C. J., Yang, H. I., Iloeje, U. H. & Group, R.-H. S. Hepatitis B virus DNA levels and outcomes in chronic hepatitis B. *Hepatology* **49**, S72-84 (2009).

232 Liu, X. *et al.* Correlation between hepatitis B virus DNA levels and diagnostic tests for HBsAg, HBeAg, and PreS1-Ag in chronic hepatitis B. *Genet Mol Res* **15** (2016).

233 Thompson, A. J. *et al.* Serum hepatitis B surface antigen and hepatitis B e antigen titers: disease phase influences correlation with viral load and intrahepatic hepatitis B virus markers. *Hepatology* **51**, 1933-1944 (2010).

234 Kim, H., Lee, S. A., Kim, D. W., Lee, S. H. & Kim, B. J. Naturally occurring mutations in large surface genes related to occult infection of hepatitis B virus genotype C. *PLoS One* **8**, e54486 (2013).

235 Lee, S. A. *et al.* Hepatitis B virus preS1 deletion is related to viral replication increase and disease progression. *World J Gastroenterol* **21**, 5039-5048 (2015).

236 Margeridon-Thermet, S. *et al.* Ultra-deep pyrosequencing of hepatitis B virus quasispecies from nucleoside and nucleotide reverse-transcriptase inhibitor (NRTI)-treated patients and NRTI-naive patients. *J Infect Dis* **199**, 1275-1285 (2009).

237 Bayliss, J. *et al.* Deep sequencing shows that HBV basal core promoter and precore variants reduce the likelihood of HBsAg loss following tenofovir disoproxil fumarate therapy in HBeAg-positive chronic hepatitis B. *Gut* (2016).

238 Xuan, J., Yu, Y., Qing, T., Guo, L. & Shi, L. Next-generation sequencing in the clinic: promises and challenges. *Cancer Lett* **340**, 284-295 (2013).

239 Han, Y. *et al.* Prediction of virological response by pretreatment hepatitis B virus reverse transcriptase quasispecies heterogeneity: the advantage of using next-generation sequencing. *Clin Microbiol Infect* **21**, 797 e791-798 (2015).

240 Yin, F. *et al.* Resistant mutations and quasispecies complexity of hepatitis B virus during telbivudine treatment. *J Gen Virol* **96**, 3302-3312 (2015).

241 Zhang, A. Y. *et al.* Deep sequencing analysis of quasispecies in the HBV pre-S region and its association with hepatocellular carcinoma. *J Gastroenterol* (2017).

242 Gong, L. *et al.* Comparison of next-generation sequencing and clone-based sequencing in analysis of hepatitis B virus reverse transcriptase quasispecies heterogeneity. *J Clin Microbiol* **51**, 4087-4094 (2013).

243 Yan, L. *et al.* Deep sequencing of hepatitis B virus basal core promoter and precore mutants in HBeAg-positive chronic hepatitis B patients. *Sci Rep* **5**, 17950 (2015).

244 Gregori, J. *et al.* Viral quasispecies complexity measures. *Virology* **493**, 227-237 (2016).

245 Lok, A. S. & McMahon, B. J. Chronic hepatitis B. *Hepatology* **45**, 507-539 (2007).

246 Lin, S. M. *et al.* Interferon therapy in HBeAg positive chronic hepatitis reduces progression to cirrhosis and hepatocellular carcinoma. *J Hepatol* **46**, 45-52 (2007).

247 van Zonneveld, M. *et al.* Long-term follow-up of alpha-interferon treatment of patients with chronic hepatitis B. *Hepatology* **39**, 804-810 (2004).

248 Chen, B. F. *et al.* High prevalence and mapping of pre-S deletion in hepatitis B virus carriers with progressive liver diseases. *Gastroenterology* **130**, 1153-1168 (2006).

249 Lapalus, M. *et al.* Precore/Core promoter variants to predict significant fibrosis in both HBeAg positive and negative chronic hepatitis B. *Liver Int* **35**, 2082-2089 (2015).

250 Ghabeshi, S., Sharifi, Z., Hosseini, S. M. & Mahmoodian Shooshtari, M. Correlation between viral load of HBV in chronic hepatitis B patients and precore and Basal core promoter mutations. *Hepat Mon* **13**, e7415 (2013).

251 Fang, Z. L. *et al.* The association of HBV core promoter double mutations (A1762T and G1764A) with viral load differs between HBeAg positive and anti-HBe positive individuals: a longitudinal analysis. *J Hepatol* **50**, 273-280 (2009).

252 Ren, X. *et al.* Hepatitis B virus genotype and basal core promoter/precore mutations are associated with hepatitis B-related acute-on-chronic liver failure without pre-existing liver cirrhosis. *J Viral Hepat* **17**, 887-895 (2010).

253 Simmonds, P. Genetic diversity and evolution of hepatitis C virus--15 years on. *J Gen Virol* **85**, 3173-3188 (2004).

254 Neumann, A. U. *et al.* Hepatitis C viral dynamics in vivo and the antiviral efficacy of interferon-alpha therapy. *Science* **282**, 103-107 (1998).

255 Ogata, N., Alter, H. J., Miller, R. H. & Purcell, R. H. Nucleotide sequence and mutation rate of the H strain of hepatitis C virus. *Proc Natl Acad Sci U S A* **88**, 3392-3396 (1991).

256 Martell, M. *et al.* Hepatitis C virus (HCV) circulates as a population of different but closely related genomes: quasispecies nature of HCV genome distribution. *J Virol* **66**, 3225-3229 (1992).

257 Pawlotsky, J. M. Hepatitis C virus population dynamics during infection. *Curr Top Microbiol Immunol* **299**, 261-284 (2006).

258 Tsubota, A., Fujise, K., Namiki, Y. & Tada, N. Peginterferon and ribavirin treatment for hepatitis C virus infection. *World J Gastroenterol* **17**, 419-432 (2011).

259 Sarrazin, C. The importance of resistance to direct antiviral drugs in HCV infection in clinical practice. *J Hepatol* **64**, 486-504 (2016).

260 Wyles, D. L. Antiviral resistance and the future landscape of hepatitis C virus infection therapy. *J Infect Dis* **207 Suppl 1**, S33-39 (2013).

261 Delang, L. *et al.* Hepatitis C virus-specific directly acting antiviral drugs. *Curr Top Microbiol Immunol* **369**, 289-320 (2013).

262 Svarovskaia, E. S., Martin, R., McHutchison, J. G., Miller, M. D. & Mo, H. Abundant drug-resistant NS3 mutants detected by deep sequencing in hepatitis C virus-infected patients undergoing NS3 protease inhibitor monotherapy. *J Clin Microbiol* **50**, 3267-3274 (2012).

263 Fonseca-Coronado, S. *et al.* Specific detection of naturally occurring hepatitis C virus mutants with resistance to telaprevir and boceprevir (protease inhibitors) among treatment-naive infected individuals. *J Clin Microbiol* **50**, 281-287 (2012).

264 Lenz, O. *et al.* Efficacy of re-treatment with TMC435 as combination therapy in hepatitis C virus-infected patients following TMC435 monotherapy. *Gastroenterology* **143**, 1176-1178 (2012).

265 Ninomiya, M. *et al.* Use of illumina deep sequencing technology to differentiate hepatitis C virus variants. *J Clin Microbiol* **50**, 857-866 (2012).

266 Sarrazin, C. *et al.* Dynamic hepatitis C virus genotypic and phenotypic changes in patients treated with the protease inhibitor telaprevir. *Gastroenterology* **132**, 1767-1777 (2007).

267 Susser, S. *et al.* Characterization of resistance to the protease inhibitor boceprevir in hepatitis C virus-infected patients. *Hepatology* **50**, 1709-1718 (2009).

268 Hedskog, C. *et al.* Dynamics of HIV-1 quasispecies during antiviral treatment dissected using ultra-deep pyrosequencing. *PLoS One* **5**, e11345 (2010).

269 Simen, B. B. *et al.* Low-abundance drug-resistant viral variants in chronically HIV-infected, antiretroviral treatment-naive patients significantly impact treatment outcomes. *J Infect Dis* **199**, 693-701 (2009).

270 Dutartre, H., Bussetta, C., Boretto, J. & Canard, B. General catalytic deficiency of hepatitis C virus RNA polymerase with an S282T mutation and mutually exclusive resistance towards 2'-modified nucleotide analogues. *Antimicrob Agents Chemother* **50**, 4161-4169 (2006).

271 McCown, M. F., Rajyaguru, S., Kular, S., Cammack, N. & Najera, I. GT-1a or GT-1b subtype-specific resistance profiles for hepatitis C virus inhibitors telaprevir and HCV-796. *Antimicrob Agents Chemother* **53**, 2129-2132 (2009).

272 Shi, S. T. *et al.* In vitro resistance study of AG-021541, a novel nonnucleoside inhibitor of the hepatitis C virus RNA-dependent RNA polymerase. *Antimicrob Agents Chemother* **52**, 675-683 (2008).

273 Asahina, Y. *et al.* Mutagenic effects of ribavirin and response to interferon/ribavirin combination therapy in chronic hepatitis C. *J Hepatol* **43**, 623-629 (2005).

274 Castilho, M. C. *et al.* Association of hepatitis C virus NS5B variants with resistance to new antiviral drugs among untreated patients. *Mem Inst Oswaldo Cruz* **106**, 968-975 (2011).

275 Hamano, K. *et al.* Mutations in the NS5B region of the hepatitis C virus genome correlate with clinical outcomes of interferon-alpha plus ribavirin combination therapy. *J Gastroenterol Hepatol* **20**, 1401-1409 (2005).

276 Bull, R. A. *et al.* Sequential bottlenecks drive viral evolution in early acute hepatitis C virus infection. *PLoS Pathog* **7**, e1002243 (2011).

277 Farci, P. *et al.* The outcome of acute hepatitis C predicted by the evolution of the viral quasispecies. *Science* **288**, 339-344 (2000).

278 Farci, P. *et al.* Early changes in hepatitis C viral quasispecies during interferon therapy predict the therapeutic outcome. *Proc Natl Acad Sci U S A* **99**, 3081-3086 (2002).

279 Laskus, T. *et al.* Analysis of hepatitis C virus quasispecies transmission and evolution in patients infected through blood transfusion. *Gastroenterology* **127**, 764-776 (2004).

280 Cabot, B. *et al.* Structure of replicating hepatitis C virus (HCV) quasispecies in the liver may not be reflected by analysis of circulating HCV virions. *J Virol* **71**, 1732-1734 (1997).

281 Farci, P. New insights into the HCV quasispecies and compartmentalization. *Semin Liver Dis* **31**, 356-374 (2011).

282 Maggi, F. *et al.* Differences in hepatitis C virus quasispecies composition between liver, peripheral blood mononuclear cells and plasma. *J Gen Virol* **78 ( Pt 7)**, 1521-1525 (1997).

283 Navas, S., Martin, J., Quiroga, J. A., Castillo, I. & Carreno, V. Genetic diversity and tissue compartmentalization of the hepatitis C virus genome in blood mononuclear cells, liver, and serum from chronic hepatitis C patients. *J Virol* **72**, 1640-1646 (1998).

284 Okuda, M. *et al.* Differences in hypervariable region 1 quasispecies of hepatitis C virus in human serum, peripheral blood mononuclear cells, and liver. *Hepatology* **29**, 217-222 (1999).

285 Gaudieri, S. *et al.* Hepatitis C virus drug resistance and immune-driven adaptations: relevance to new antiviral therapy. *Hepatology* **49**, 1069-1082 (2009).

286 Gane, E. J. Diabetes mellitus following liver transplantation in patients with hepatitis C virus: risks and consequences. *Am J Transplant* **12**, 531-538 (2012).

287 Ludmerer, S. W. *et al.* Replication fitness and NS5B drug sensitivity of diverse hepatitis C virus isolates characterized by using a transient replication assay. *Antimicrob Agents Chemother* **49**, 2059-2069 (2005).

288 Franco, S. *et al.* No detection of the NS5B S282T mutation in treatment-naive genotype 1 HCV/HIV-1 coinfected patients using deep sequencing. *J Clin Virol* **58**, 726-729 (2013).

289 Di Maio, V. C. *et al.* Hepatitis C virus genetic variability and the presence of NS5B resistance-associated mutations as natural polymorphisms in selected genotypes could affect the response to NS5B inhibitors. *Antimicrob Agents Chemother* **58**, 2781-2797 (2014).

290 Sarrazin, C. & Zeuzem, S. Resistance to direct antiviral agents in patients with hepatitis C virus infection. *Gastroenterology* **138**, 447-462 (2010).

291 Bartels, D. J. *et al.* Hepatitis C virus variants with decreased sensitivity to direct-acting antivirals (DAAs) were rarely observed in DAA-naive patients prior to treatment. *J Virol* **87**, 1544-1553 (2013).

292 Dryer, P. D. *et al.* Screening for hepatitis C virus non-nucleotide resistance mutations in treatment-naive women. *J Antimicrob Chemother* **64**, 945-948 (2009).

293 Le Pogam, S. *et al.* Existence of hepatitis C virus NS5B variants naturally resistant to non-nucleoside, but not to nucleoside, polymerase inhibitors among untreated patients. *J Antimicrob Chemother* **61**, 1205-1216 (2008).

294 Tamura, K. *et al.* MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* **28**, 2731-2739 (2011).

295 Cock, P. J. *et al.* Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**, 1422-1423 (2009).

296 Henn, M. R. *et al.* Whole genome deep sequencing of HIV-1 reveals the impact of early minor variants upon immune recognition during acute infection. *PLoS Pathog* **8**, e1002529 (2012).

297 Brockman, W. *et al.* Quality scores and SNP detection in sequencing-by-synthesis systems. *Genome Res* **18**, 763-770 (2008).

298 Macalalad, A. R. *et al.* Highly sensitive and specific detection of rare variants in mixed viral populations from massively parallel sequence data. *PLoS Comput Biol* **8**, e1002417 (2012).

299 Holtgrewe, M. Mason – a read simulator for second generation sequencingdata. http://publications.mi.fu-berlin.de/962/ (FU Berlin, 2010).

300 Simpson, E. H. Measurement of diversity. *Nature* **163**, 688 (1949).

301 WHO. *Global Hepatitis Report 2017*, <http://www.who.int/hepatitis/publications/global-hepatitis-report2017/en/> (2017).

302 Chaabna, K., Kouyoumjian, S. P. & Abu-Raddad, L. J. Hepatitis C Virus Epidemiology in Djibouti, Somalia, Sudan, and Yemen: Systematic Review and Meta-Analysis. *PLoS One* **11**, e0149966 (2016).

303 Ochwoto, M. *et al.* Hepatitis B infection is highly prevalent among patients presenting with jaundice in Kenya. *BMC Infect Dis* **16**, 101 (2016).

304 Habil, F. E., Mahdi, W. K., Abdelwahab, S. F. & Abdel-Hamid, M. Hepatitis B virus genotype D predominates HBsAg-positive egyptian blood donors and is mainly associated with a negative HBeAg serostatus. *Intervirology* **56**, 278-283 (2013).

305 Hannoun, C., Soderstrom, A., Norkrans, G. & Lindh, M. 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**, 2163-2167 (2005).

306 Thurnheer, M. C. *et al.* Genotypic profiles of hepatitis B in African immigrants and their clinical relevance. *J Med Virol* **89**, 1000-1007 (2017).

307 Deressa, T. *et al.* The burden of hepatitis B virus (HBV) infection, genotypes and drug resistance mutations in human immunodeficiency virus-positive patients in Northwest Ethiopia. *PLoS One* **12**, e0190149 (2017).

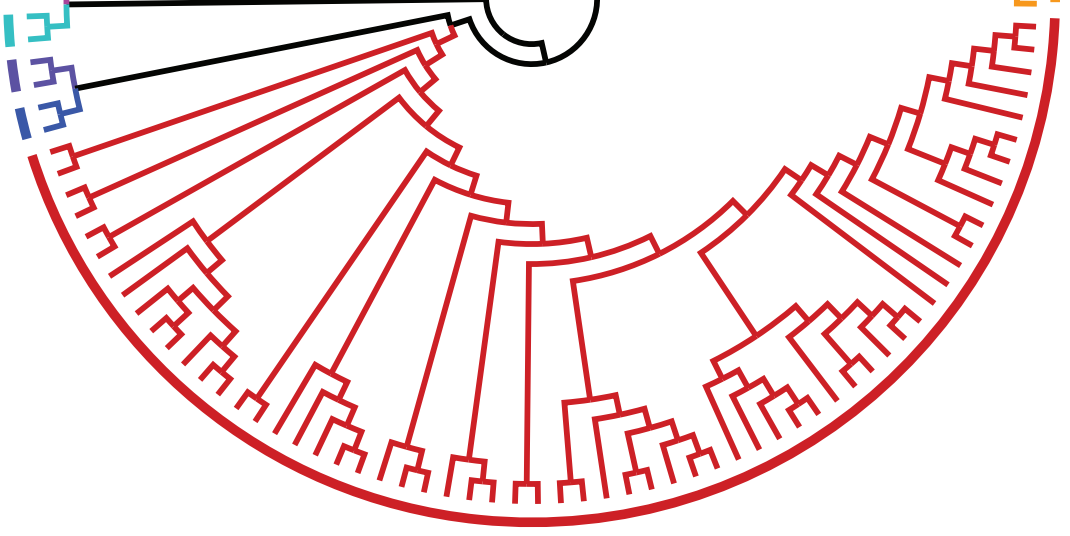
308 Scavone, C. *et al.* New era in treatment options of chronic hepatitis C: focus on safety of new direct-acting antivirals (DAAs). *Expert Opin Drug Saf* **15**, 85-100 (2016).

309 Kartashev, V. *et al.* New findings in HCV genotype distribution in selected West European, Russian and Israeli regions. *J Clin Virol* **81**, 82-89 (2016).

310 Alexopoulou, A. & Karayiannis, P. HBeAg negative variants and their role in the natural history of chronic hepatitis B virus infection. *World J Gastroenterol* **20**, 7644-7652 (2014).

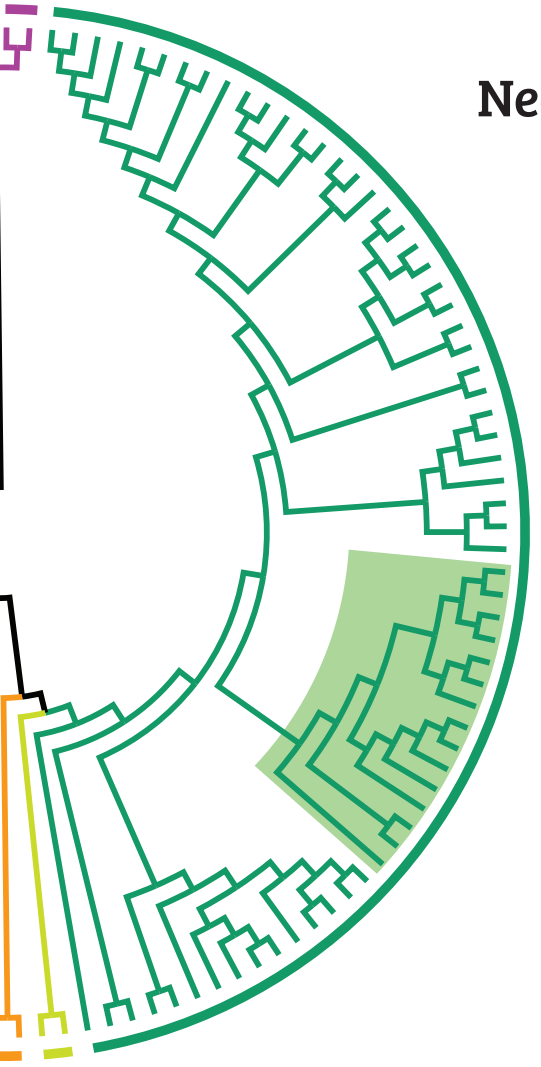
311 Jimenez-Perez, M. *et al.* Treatment of chronic hepatitis C with direct-acting antivirals: The role of resistance. *World J Gastroenterol* **22**, 6573-6581 (2016).

312 Ito, J. *et al.* Prevalence and characteristics of naturally occurring sofosbuvir resistance-associated variants in patients with hepatitis C virus genotype 1b infection. *Hepatol Res* **46**, 1294-1303 (2016).



# Chapter 10

## Nederlandse Samenvatting



Hepatitis B- en C-virussen zijn wereldwijd een belangrijke oorzaak van morbiditeit en mortaliteit. Volgens het rapport van de Wereldgezondheidsorganisatie (WHO) veroorzaakte virale hepatitis 1,34 miljoen doden in 2015, een aantal vergelijkbaar met jaarlijkse sterfgevallen veroorzaakt door tuberculose maar hoger dan die veroorzaakt door HIV of malaria. In tegenstelling tot deze andere drie infectieziekten heeft virale hepatitis echter relatief weinig aandacht gekregen. Wereldwijd heeft slechts 9% van hepatitis B virus (HBV) en 20% van de met hepatitis C virus (HCV) besmette personen toegang tot betaalbare hepatitis-tests. Nog zorgwekkender is het feit dat antivirale behandeling beschikbaar is voor slechts 8% van degenen met HBV en 7,4% van degenen met HCV. Dit is nog problematischer in landen als Ethiopië, waar meer dan 80% van de bevolking op het platteland woont met een zeer beperkte toegang tot gezondheidszorg. Daarnaast is virale hepatitis nog niet beschouwd als een prioriteit in Ethiopië vanwege een gebrek aan bewustzijn en het feit dat er geen nationale antivirale behandeling en diagnostische richtlijnen of surveillance zijn ontwikkeld. Bovendien werd pas in 2007 het universeel verkrijgbare HBV-vaccin opgenomen in het nationale neonatale immunisatieprogramma. Bijgevolg blijven deze virussen zich verspreiden. Daarom hebben we in dit proefschrift de seroepidemiologie en genetische diversiteit van HBV en HCV in Ethiopië bestudeerd en daarnaast analyseren we niuewe methodes om de differentiatie van HBV en HCV quasispecies met behulp van next generation sequencing-technologie mogelijk te maken.

**Hepatitis B- en C-virussen in Ethiopië**

Vanwege het gebrek aan nationaal surveillanceprogramma en de beschikbaarheid van routinematige diagnostische tests voor patiënten blijft de omvang van HBV- en HCV-infecties in Ethiopië grotendeels onbekend. Eerdere studies toonden aan dat de seroprevalentie van HBV (2,1 tot 25,0%) en HCV (0,7 tot 13,3%) infectie nogal varieerden. De meeste van deze meldingen zijn echter gedaan op basis van een beperkt aantal personen (kleine steekproefomvang) of van specifieke groepen (HIV-geïnfecteerde patiënten, leverpatiënten, commerciële of vervangende bloeddonoren, zwangere vrouwen, enz.)Deze vertegenwoordigen daarom mogelijk niet de echte nationale omvang van virale hepatitis in de algemene bevolking. In de huidige studie probeerden we deze lacune op te vullen door gebruik te maken van een groot aantal vrijwillige gezonde bloeddonoren (56 885) uit vijf verschillende geografische regio's die zich op ongeveer 100 tot 1 500 kilometer van elkaar bevinden en verschillen in hun sociaaleconomische, cultuur en etniciteit (hoofdstuk 2). We vonden een algemene landelijke seroprevalentie van 3,9% (variërend van 3,7 tot 5,91%) voor HBV en 0,52% (variërend van 0,37

tot 0,65%) voor HCV. Deze prevalentiepercentages classificeren het land als een intermediair (2 -7%) endemisch gebied voor HBV-infectie en laag (<1,5%) voor HCV. Dit in tegenstelling tot de standaardindeling van het land als zeer endemisch voor HBV (> 8%) vanwege de geografische locatie in Afrika ten zuiden van de Sahara. Ons resultaat is in overeenstemming met de meest recente algemene beoordeling van de prevalentie op landelijk niveau van chronische HBV-infectie die een seroprevalentie van 6,03% in Ethiopië rapporteerde. HCV-seroprevalentie van 0,9%, 1,0% en 0,3% werden gevonden in de algemene bevolking van de buurlanden Somalië, Soedan en Djibouti, respectievelijk. Onze landelijke seroprevalentie gegevens wijzen ongetwijfeld op het algemene beeld van de HBV- en HCV-epidemiologie in het land, maar kunnen mogelijk worden onderschat vanwege de uitsluiting van bloeddonoren met klinische tekenen en symptomen van hepatitis. Samengevat is de belasting van virale hepatitis in Ethiopië aanzienlijk en is er behoefte aan surveillance op nationaal niveau en publieke campagnes om het bewustzijn te vergroten.

**Moleculaire epidemiologie en genetische diversiteit van HBV in Ethiopië**

HBV vertoont een hoge genetische diversiteit met 10 genotypes, meer dan veertig subgenotypes en talloze recombinanten. In Afrika hebben hoofdzakelijk drie belangrijke genotypes de overhand met verschillende epidemiologische trends. Genotype A (subgenotype A1) is dominant in Oost- en Zuidelijk Afrika, genotype D (D1 en D7) in Noord-Afrika en genotype E in West- en Centraal-Afrika. In Ethiopië vonden we twee belangrijke genotypen, A (78%) en D (22%) op basis van sequentieanalyse van 371 monsters verkregen uit vijf regio's: Addis Ababa, Adama, Gondar, Jimma en Mekelle. Beide genotypes zijn overheersend in het hele land maar laten een ander epidemiologisch patroon zien. Genotype D kwam vaker voor in de noordelijke regio van het land in vergelijking met de centrale en zuidwestelijke regio's. Ggenotype A wordt voornamelijk gezien in de centrale en zuidwestelijke regio's (hoofdstuk 3). In de buurlanden is genotype A het overheersende genotype in Somalië, Kenia, Rwanda en Oeganda, terwijl genotype D het meest voorkomende genotype is in Soedan en Egypte. Het is opmerkelijk om te vermelden dat Ethiopië op basis van deze waarnemingen de grens van de geografische verspreiding van genotype A en D vormt.

Verder valt op te merken dat alle genotype A-isolaten uit Ethiopië behoren tot subgenotype A1, vergelijkbaar met die in andere Oost- en Zuid-Afrikaanse landen. Ze verschillen echter genetisch van de meeste stammen uit deze regio's, omdat deze stammen uit Ethiopië zich in zowel de Afrikaanse als de Aziatisch-Amerikaanse clades nestelen (hoofdstuk 3). Fascinerend genoeg zijn de meeste Ethiopische stammen nauwer geclusterd met de Aziatisch-Amerikaanse clades

(soorten uit Brazilië, Haïti, Japan, India, de Filippijnen) dan met de Afrikaanse clade (Rwanda, Malawi, Uganda, Tanzania of Zuid-Afrika). Er is gesuggereerd dat genotype A afkomstig is uit Afrika en zich in de laatste 1000-2000 jaar heeft uitgebreid van Oost-Afrika naar Zuid-Azië door reizen en handel langs de kust en naar Midden- en Zuid-Amerika tijdens de negentiende-eeuwse slavenhandel en door menselijke migratie. Rekening houdend met de vroege splitsing van A1-stammen uit Ethiopië in de fylogenetische boom en de locatie van Ethiopië op een strategische geografische positie tussen Afrika en Eurazië, gecombineerd met de kennis over de hoofdroute van menselijke migratie vanuit Afrika evenals de oorsprong van moderne mens, veronderstellen we dat HBV-stammen uit Ethiopië de oudste afstamming van subgenotype A1 zouden kunnen zijn.

Genotype D, het op een na meest voorkomende genotype in Ethiopië, is genetisch zeer divers en onderscheidt zich in vijf subgenotypes: D1, D2, D4, D7 en één niet-toegewezen subgenotype (nu D10 genoemd). Interessant genoeg behoorde het grootste deel van genotype D-stammen (85%) tot twee subgenotypen: D2 (55%) en de nieuwe D10 (30%). Het is belangrijk om te onthouden dat subgenotype D2 voorkomt in Oost-Europa, inclusief Rusland en de Baltische staten, maar minder gebruikelijk is in Afrika, waar D1 het dominante subgenotype is. Desondanks groepeerden de Ethiopische D2-stammen zich afzonderlijk van Europese en Aziatische virussen. De D2-stammen uit Ethiopië zijn overal in het land te vinden en variëren genetisch nogal behoorlijk. Bovendien heeft Ethiopië geen historische relatie met die Europese landen waar D2 de boventoon voert. We stellen daarom voor dat D2 uit Ethiopië afzonderlijk is geëvolueerd in een specifieke etno-geografische omgeving en een lange natuurlijke geschiedenis heeft in de Ethiopische bevolking.

Zoals besproken in hoofdstuk 3, identificeerden we één niet eerder toegewezen subgenotype, hierna D10 genoemd, gebaseerd op analyse van kleine genensequenties. Om de HBV-evolutie in Ethiopië beter te beschrijven, onderzochten we vervolgens 14 niet-toegewezen stammen door een sequentieanalyse van het volledige genoom (hoofdstuk 4). Interessant is dat de niet-toegewezen stammen een duidelijke monofyletische boom vertonen met unieke onderscheidende kenmerken, wat voldoende bewijs is om ze in te delen in een nieuw subgenotype D10. In de eerste plaats werden deze stammen verkregen van niet-verwante bloeddonoren uit verschillende regio's van Ethiopië. Ten tweede vormden ze een afzonderlijke monofyletische cluster welkee sterk werd ondersteund door een hoge bootstrapwaarde, zowel in het complete genoom als in de vier afzonderlijke fylogenetische bomen met open lees ramen (hoofdstuk 4). Ten derde was de gemiddelde inter-subgenotype nucleotide divergentie in vergelijking met de referentiesequenties van alle bestaande subgenotypes D

> 4% en <7,5%, wat de minimale vereiste is voor de definitie van een nieuw subgenotype. Ten slotte hadden ze negen unieke aminozuur substitutie posities in de oppervlakte-, polymerase- en X-genen; vijf ervan waren exclusief uniek voor D10. Bovendien is, in tegenstelling tot de voorgaande subgenotypes D9 en D8 die respectievelijk recombinant zijn van C / D en D / E, de nieuwe D10 vrij van enige recombinatie. Dit kan worden gezien in de vier fylogenetische bomen van de ORF's die verder werden bevestigd door verschillende HBV-recombinante analysetools. Samenvattend wijzen de identificatie van nieuwe D10 en een unieke verdeling van subgenotype D2 in Ethiopië, op de hoge genetische variabiliteit en evolutie van HBV in Afrika, met name in Ethiopië.

**Genetische diversiteit van HCV in Ethiopië**

HCV, de etiologische agens van 185 miljoen chronische hepatitisinfecties en de belangrijkste oorzaak van levercirrose en HCC, wordt gekenmerkt door een grote genetische variabiliteit. Het virus is ingedeeld in 7 genotypen en meer dan 100 subtypes. Elk van deze genotypen is uniek met betrekking tot de geografische distributie en therapeutische respons. Genotypen 1, 2 en 3 hebben een wereldwijde verdeling, terwijl genotypes 4, 5, 6 en 7 een beperkte geografische spreiding hebben. In Afrika zijn de dominante genotypen 1, 2, 4 en 5, met verschillende epidemiologische patronen. Toch is er weinig bekend over de heersende genotypen en subtypen in Oost-Afrika, waaronder Ethiopië. In de huidige landelijke studie vonden we dat 4 belangrijke genotypen (1, 2, 4 en 5) en zeven subtypes (1b, 1c, 2c, 4d, 4l, 4r en 4v) overheersen in Ethiopië (hoofdstuk 5). Genotype 4 is verreweg het meest dominant in de analyse van 78% van de HCV-stammen die werden gekarakteriseerd, gevolgd door genotype 2c. De meest voorkomende subtypen zijn 4d en 4r, waarbij elk verantwoordelijk is voor 35% van de HCV-stammen circuleren. Dit genotype prevaleert in het hele land, zoals blijkt uit de geanalyseerde gegevens van de vijf geografische regio's. Bovendien werd door analyse van twee verschillende gebieden van het HCV-genoom een potentieel natuurlijk recombinant virus geïdentificeerd consistent met de aanwezigheid van de twee belangrijkste genotypen (4 en 2) . Over het algemeen onderstrepen de bevindingen van de huidige studie de noodzaak van verdere moleculaire karakterisering van HCV uit dit gebied.

Het is bekend dat kennis van HCV-genotype-informatie nuttig is voor het nemen van behandelbeslissingen, zoals de duur van de behandeling, het type medicijn dat moet worden gebruikt en de waarschijnlijkheid van een behandelingsreactie. Genotype 4 is een van de moeilijkst te behandelen genotypes met de beschikbare standaardbehandeling, peginterferon alpha plus ribavirine. Tegenwoordig zijn pan-genotypische direct werkende antivirale middelen de keuze voor behandeling

in de westerse wereld. Daarom is het hebben van voldoende informatie over het gangbare HCV-genotype / -subtype van groot belang voor de toekomstige ontwikkeling van behandelingsrichtlijnen en patiëntbeheer in regio's zoals Ethiopië, waar op dit moment geen nationaal hepatitis-behandelingsbeleid bestaat.

**HBV quasispecies varianten**

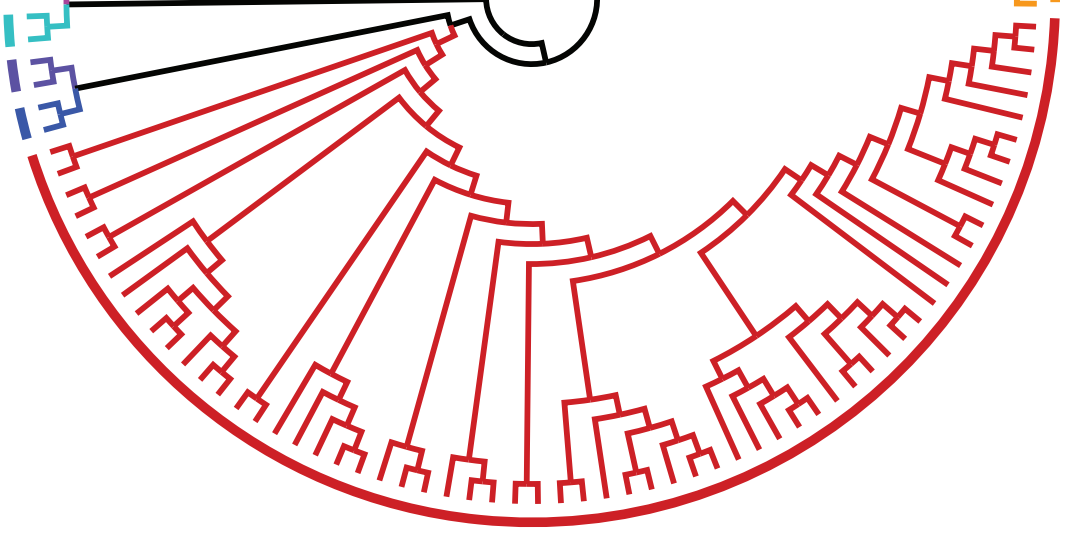
In het laatste decennium is ruime informatie verkregen over de genetische diversiteit van HBV. Hoewel deep-sequencing van geselecteerde genoom gebieden de detectie van virale varianten van lage frequenties mogelijk maakte, is variant analyse op het volledige genoom niveau niet in detail uitgevoerd. We bestudeerden de associatie van HBV quasispecies-varianten met behulp van volledige genoom sequencing met HBeAg-status en virale belasting bij chronisch geïnfecteerde HBV-patiënten. We identificeerden een totaal van 4615 verschillende quasispecies-varianten over het volledige genoom van vijf HBV-genotypen (A tot D) bij 186 patiënten. Deze quasispecies variante diversiteit was significant hoger in virussen van HBeAg-negatieve patiënten in vergelijking met HBeAg-positieve patiënten (hoofdstuk 6). HBeAg is klinisch belangrijk omdat de aanwezigheid ervan in serum een marker is van virale replicatie en de seroconversie ervan een potentieel eindpunt van de behandeling is. Er zijn echter HBV-varianten die de HBeAg-productie kunnen verminderen of elimineren, waardoor het virus de immuun aanval van de gastheer en de voortdurende replicatie kan omzeilen. Als gevolg hiervan kan het verlies van HBeAg alleen niet indicatief zijn voor virale klaring of remissie van de ziekte, het is van het grootste belang om virale varianten te identificeren die zijn geassocieerd met de HBeAg-status. We hebben meerdere varianten geïdentificeerd die significant geassocieerd zijn met de HBeAg-status (hoofdstuk 6). De meerderheid van deze varianten bevonden zich in het core gen gevolgd door het oppervlakte-gen, dat mogelijk het gevolg was van de zeer selectieve immunologische druk die in deze regio's werd uitgeoefend. Bovendien merkten we dat ongeveer een derde van deze varianten geassocieerd bleek te zijn met lage virale hoeveelheden in het bloed, wat de correlatie aangeeft tussen HBeAg- en HBV-DNA-niveaus. Over het algemeen laten deze resultaten zien dat naast de welbekende PC- en BC-varianten die de HBeAg-productie verminderen of elimineren, andere varianten ook kunnen bijdragen aan HBeAg-reductie en daarom nuttig zijn bij het controleren van patiënten. Concluderend kan een hoogwaardige HBV-genoom-deep-sequencing-benadering nieuw inzicht verschaffen in de diversiteit van HBV quasispecies en de relatie ervan tot klinische kenmerken van de patiënt.

**Direct werkende antivirale resistentie-geassocieerde HCV-varianten**

Tot voor kort was de combinatie van gepegyleerd interferon plus ribavirine de standaardbehandeling voor HCV-infectie. Deze medicijnen worden echter geassocieerd met frequente en soms ernstige bijwerkingen met beperkte werkzaamheid, waardoor onderzoekers op zoek gaan naar alternatieve opties. Sinds 2011 is de HCV-behandeling dramatisch veranderd als gevolg van de komst van DAA's. Deze geneesmiddelen richten zich op de niet-structurele coderende genen van het HCV-genoom, wat resulteert in hoge aanhoudende virologische responspercentages (> 90%). Echter, geneesmiddelresistentie-geassocieerde varianten (RAVs) zijn naar voren gekomen als een potentiële uitdaging voor de behandeling van HCV-infectie vanwege de hoge genetische variabiliteit in de HCV-quasispecies populatie. Deze RAV's kunnen aanwezig zijn in therapie naïeve patiënten als van nature voorkomende varianten gewoonlijk op lage niveaus, hoewel ze vaker worden geselecteerd na falen van de behandeling en de overheersende varianten worden. In hoofdstuk 7 onderzoeken we de pre-existentie van RAVs als we naar de NS5B-regio kijken (spreiding van aminozuurposities 215 tot 348 waarvan bekend is dat ze resistentie verlenen tegen NS5B-polymerase remmers) in de lever- en plasma compartimenten van naïeve patiënten die 454 met HCV genotype 1b zijn geïnfecteerd. In ons cohort van 18 gepaarde behandelingsnaïeve patiënten werd de nucleosideremmer (NI) resistentiegeassocieerde variant S282T niet als een minderheid variant gedetecteerd, noch in plasma noch in leverweefsel. Over het algemeen is de S282T-variant in vivo zelden gedetecteerd vanwege de lage replicatieve fitheid van deze variant. Echter werd de andere NI-resistente variant V321A gedetecteerd bij 10 van de 18 patiënten (~ 56%), voornamelijk als minder belangrijke variant. We zagen ook mutaties die resistentie verlenen aan niet-nucleosideremmers op positie 316 (C316N / Y / H) bij 16 van de 18 patiënten (~ 89%). Over het algemeen vertoonde de quasispecies distributie in zowel plasma- als levermonsters een opmerkelijke overlap in elke patiënt (Hoofdstuk 7), wat onderstreept dat het gebruik van plasma hoogstwaarschijnlijk voldoende is om HCV-quasispecies en RAV's te detecteren. Hoewel sommige van de varianten in de huidige studie slechts een laag niveau van resistentie verlenen, moet de aanwezigheid van HCV-varianten op baseline in aanmerking worden genomen in de context van DAA-therapie.

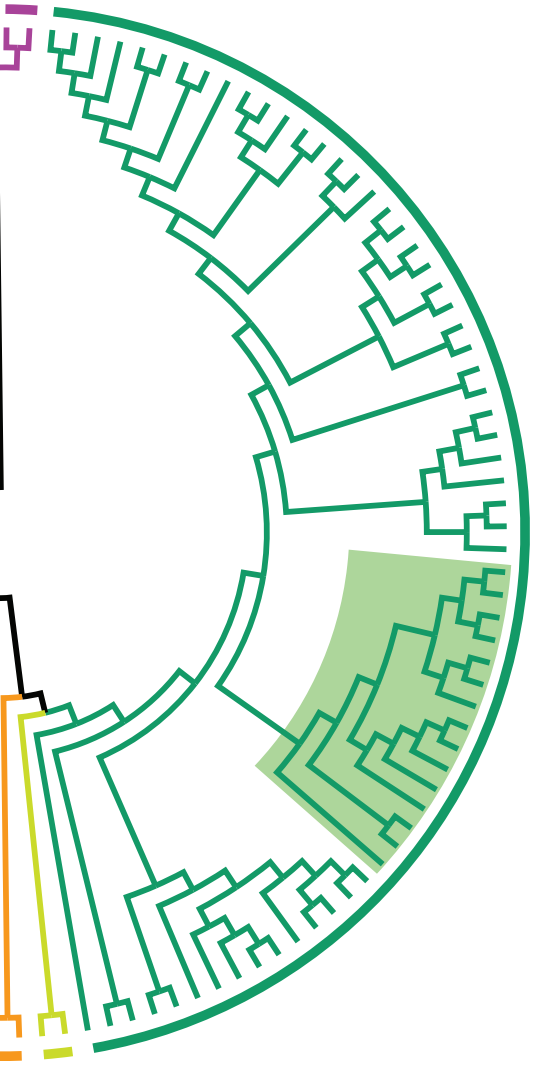
Het werk in dit proefschrift beschrijft nieuwe inzichten in de moleculaire en genetische diversiteit van HBV en HCV in Ethiopië. Dit is de eerste dergelijke studie uit deze regio die als referentie kan dienen voor toekomstige studies. Toen we in 2012 met dit onderzoek begonnen, werd virale hepatitis grotendeels genegeerd als gezondheidsprioriteit in landen met lage inkomens, waaronder Ethiopië. Tegenwoordig wint virale hepatitis, als gevolg van de toenemende

ziektelast, wereldwijd meer aandacht. De WHO verklaarde (Glasgow Declaration on Hepatitis - september 2015) virale hepatitis als een ernstige bedreiging voor de volksgezondheid en stelde in 2030 een nieuwe globale strategie om nieuwe infecties met 90% en mortaliteit te verminderen met 65%. Om dit doel te bereiken, definieert het strategisch kader een reeks prioritaire acties die moeten worden ondernomen: verbetering van het informatiesysteem voor het begrijpen van virale hepatitis-epidemieën, uitbreiding van de dekking van testen en behandeling door de gezondheidszorg, levering van hepatitis diensten via een pakket voor de volksgezondheid, financiering van duurzaamheid voor de dekking van de universele gezondheidszorg en versnelde innovaties. In reactie op de oproep tot actie van de WHO moet Ethiopië ook het nationale beleid en de nationale strategieën voorbereiden op gebieden als surveillance, identificatie van risicogroepen, diagnose, behandeling, mobilisatie van middelen, bewustmaking van het publiek en pleitbezorging. Dergelijke inspanningen vereisen een sterke politieke betrokkenheid en een hoge mate van aandacht die gelijk is aan hiv, tuberculose en malaria. Omdat de standaard serologische methode geen occulte hepatitis-B-infectie kan detecteren - HBsAg-negatief maar viraal DNA-positief - moet Ethiopië beginnen met het opnemen van nucleïnezuuronderzoek in bloeddonorscreening bij nationale en regionale bloedbanken. Bovendien zijn genotypes en bepaling van genetische varianten bij chronische HBV- en HCV-infectie belangrijk bij het schatten van ziekte progressie en het plannen van optimale antivirale behandelingen, vooral met de huidige direct werkende antivirale middelen en pan-genotypische geneesmiddelen. In Ethiopië, zoals aangetoond in deze studie, vertonen zowel HBV als HCV een breed scala aan genotypische diversiteit. Daarom moeten richtlijnen voor de preventie, zorg en behandeling van HBV- en HCV-infectie rekening houden met de genotypes die in het land circuleren. Bovendien moet de klinische impact van HBV- en HCV-genotypen in Ethiopië verder worden onderzocht.



# Chapter 11

## Epilogue



**ABOUT THE AUTHOR**

*Curriculum vitae*

*PhD portfolio*

*Publications*

**ACKNOWLEDGMENTS**

## CURRICULUM VITAE

The author of this thesis was born on 08 July 1975 in Addis Alem, Ethiopia. After finishing high school in his hometown, he started his BSc study in Biology at Addis Ababa University and continued his MSc study in Microbiology at the same university. During his MSc study, he developed and evaluated a rapid diagnostic method, the colorimetric MTT assay, for detection of isoniazid and rifampicin resistance in *Mycobacterium tuberculosis*. After his graduation (2007), he worked as a lecturer at the University of Gondar, Ethiopia. In 2012, he started as a PhD student at the Department of Viroscience in the Erasmus MC The Netherlands under the supervision of Dr. Bart L. Haagmans and Prof. Marion P. Koopmans. The focus of his PhD project was studying epidemiology and genetic diversity of HBV and HCV strains in Ethiopia and in addition characterizing the importance of HBV/HCV quasispecies by ultra-deep sequence, which has been resulted in this thesis. As part of his PhD-training, he studied research master in Infection and Immunity (2012-2014) at Erasmus MC Molecular Medicine Postgraduate School.

## PHD PORTFOLIO

| | |
|---|---|
| **Name**: | Gadissa Bedada Hundie |
| **Research Department**: | Viroscience, Erasmus MC |
| **Research school**: | Molecular Medicine (MolMed) Postgraduate School |
| **PhD period**: | 2012-2018 |
| **Promotor**: | Prof. Dr. MPG Koopmans |
| **Copromotor**: | Dr. Bart L. Haagmans |

**Education**

| | | |
|---|---|---|
| 2012-2018 | PhD Program | Erasmus MC |
| 2012-2014 | Master of Science (Infection & Immunity) | Erasmus MC |
| 2005-2007 | Master of Science (Microbiology) | Addis Ababa University |
| 2001-2004 | BSc (Biology) | Addis Ababa University |

**In-Depth courses / Certificates**

| | |
|---|---|
| Bayesians statistics and JASP | 2018 |
| Programming with Python | 2017 |
| Course in Adobe InDesign CS6 | 2017 |
| Course in a broad spectrum of NGS applications in Molecular Medicine | 2015 |
| Course in GraphPad Prism | 2015 |
| Basic Course on R | 2014 |
| Workshop on NCBI and other software | 2014 |
| Virology course and symposium | 2014 |
| Course in Biomedical English Writing and Communication | 2014 |
| Course in Molecular Medicine | 2013 |
| Translational Imaging workshop by AMIE: from mouse to man | 2013 |
| Course in Adobe Photoshop and Illustrator CS5 | 2012 |
| Course in Adobe InDesign CS5 | 2012 |
| Course in Statistics and Survival analysis | 2012 |
| Course in Genetic for Dummies | 2012 |
| Course in SPSS | 2012 |
| Course in Biomedical Research Techniques | 2012 |

**Attended conferences and presentations**

| | | |
|---|---|---|
| ASM Microbe 2017 meeting, New Orleans, USA (June 1-5) | (Poster) | 2017 |
| 21st Molecular Medicine Day, Rotterdam, The Netherlands (March 21) | (Poster) | 2017 |
| Training workshop PREDEMICS, Institute Pasteur, Paris, France (September 19-22) | | 2016 |

| | |
|---|---|
| 20th Molecular Medicine Day, Rotterdam, The Netherlands (March 3) | 2016 |
| Molecular Biology of Hepatitis B Viruses 2015 International meeting 2015, Bad Nauheim, Germany (October 4-8) | 2015 |
| 19th Molecular Medicine Day, Rotterdam, The Netherlands (March 19) | 2015 |
| 18th Molecular Medicine Day, Rotterdam, The Netherlands (February 20) | 2014 |
| 17th Molecular Medicine Day, Rotterdam, The Netherlands (February 13) | 2013 |

**Grant**
Nuffic grant (No. CF8605)

## LIST OF PUBLICATIONS

**Hundie GB**, van de Vijver DAM, Ad J.P. van der Meer, Margo J.H. van Campenhout, Wildum S, Garcia-Alcalde F, Mueller-Breckenridge AJ, Chopy D, Boucher CA, Haagmans BL .Hepatitis B virus quasispecies variants associated with HBeAg status and viral load as determined by full-length genome deep sequencing (manuscript in preparation).

Raj VS, **Hundie GB**, Schürch AC, Smits SL, Pas SD, Pogam SL, Janssen HLA, de Knegt RJ, Osterhaus ADME, Najera I, Boucher CA, Haagmans BL. Identification of HCV Resistant Variants against Direct Acting Antivirals in Plasma and Liver of Treatment Naïve Patients. Scientific Reports 2017; 7: 4688 | DOI:10.1038/s41598-017-04931-y.

**Hundie GB**, Raj VS, GebreMichael D, Pas SD, Haagmans BL. Genetic diversity of hepatitis C virus in Ethiopia**.** PLOS ONE. 2017 Jun 1;12(6):e0179064.

**Hundie GB**, Raj VS, GebreMichael D, Haagmans BL. Seroepidemiology of hepatitis B and C virus infections among blood donors in Ethiopia. J Med Virol. 2017; 89(7):1300–1303.

**Hundie GB**, Raj VS, Gebre Michael D, Pas SD, Koopmans MP, Osterhaus ADM E, Smits SL, Haagmans BL. A novel hepatitis B virus subgenotype D10 circulating in Ethiopia. J Viral Hepat. 2017; 24 (2): 163–173.

**Hundie GB**, Woldemeskel D, Gessesse A. Evaluation of Direct Colorimetric MTT Assay for Rapid Detection of Rifampicin and Isoniazid Resistance in *Mycobacterium tuberculosis*. PLoS ONE. 2016; 11(12): e0169188.

**Hundie GB**, Raj VS, GebreMichael D, Pas SD, Osterhaus ADM E, Koopmans MP, Smits SL, Haagmans BL. Molecular Epidemiology and Genetic Diversity of Hepatitis B Virus in Ethiopia. J Med Virol. 2016; 88(6):1035–1043.

Lamers MM, Smits SL, **Hundie GB**, Provacia LB, Koopmans M, Osterhaus AD, Haagmans BL, Raj VS. Naturally occurring recombination in ferret coronaviruses revealed by complete genome characterization. J Gen Virol. 2016; 97(9):2180-2186.

## ACKNOWLEDGMENTS

Finally, it would be remarkable to close this long journey with thankful words of appreciation to so many great people who have helped me in accomplishing this thesis. Forgive me if you are not on the list, it wasn't intentional. At the beginning of this study, I have very little information and knowledge about viral hepatitis in general except what I have read from literature. But I have a strong ambition and enthusiasm to work on these viruses as they have important health impact in Ethiopia; though they have been neglected by health sectors. Yet, the most challenging for me at that moment was where to work, with whom, and where and how to get funding. So, it was my daily duty to google, google, and…google in searching for my future mentor, that was how I came up with Ab Osterhaus.

Ab, thank you very much for everything you did for me including reading my research interest and responding to my first email that opened my chapter. Your positive words put me on the moon, the joyful day in my life. Not only that you sent my email to your workgroup leaders requesting them to contact me should they be interested in meeting me. Oh, that was fruitful! Ab, I am very grateful again for bringing me to Bart and opened the beginning of my new journey from Ethiopia to the Netherlands and to the viral hepatitis world.

Bart, I am so grateful to have you as my supervisor, you have been a tremendous mentor for me. It is a pleasure to work with you. Where to begin it is almost six years, unbelievable! Oh , from the first email you sent me to send you my proposal to have a look at it and see whether you can be of any help. Yes, you did it. You supported me in all the application process for Nuffic grant, which was successful!! You picked me from the airport in my new adventure to Europe. Since then, you have always been supportive at a time of need. Bart, thank you very much for your continuous support, enthusiastic encouragement, motivation, and guidance throughout my study and research. Your professional guidance, advice and constructive recommendations in all the time of my research, scientific writing as well as on my career have been priceless.

I would like to express my special thanks and appreciation to my promotor Marion. Dear Marion, your warmly brainstorming discussion, constructive ideas, suggestion and comments were the corner pillars of my thesis.

I am deeply grateful to the following national and regional blood bank services staffs (Ethiopia) for their assistance with the collection of my samples and

data: Daniel, Mengistu, Ejigayehu, and Azalu (Addis Ababa); Shibre and Tigist (Adama); Samuel and Teklewine (Mekelle), Woinshet and Habtamu (Gondar); Abebaw (Jimma), without their precious support this work hadn't been possible. My grateful thanks are also extended to Solomon G/Selassie for his valuable and constructive suggestions during the planning and development of this research project.

I would especially like to thank Stalin, Suzan, Saskia and Rogier for their unreserved support. Stalin, it was magnificent to have the opportunity to work with you. You are such an amazing person. Your hard work, determination and enthusiasm were my inspiration and driving forces. Thanks for all your valuable technical support and encouragement! Suzan, you were my resource powerhouse. You are always there and supportive when I need, thank you. I would also like to extend my thanks to Claudia Mulders and Claudia Schapendonk for their help in offering me the resources and technical assistance.

I am deeply grateful to Do, Nisreen and Bri for the stimulating discussions, debates, helpful ideas and suggestion, and for all the fun we have had in the last four years. Do, you are always supportive every time I stacked with adobe illustrator, tatatata tatatata! Also, you never forget to replenish our coffee stock; we need to boost our energy to work hard and do stupid things faster. Bri, thank you for the nice Indonesian coffee gift. Nisreen, thank you for the delicious Egyptian cookies. For our loudest room Ee1759 crew both onboard and off board: Alta, Amare, Cynthia, Mart, Marwah, Miruna, Rebecca, Risa, Shirley, Simone, Jeroen…….. thank you for the wonderful time we have spent together.

I am indebted to so many people in the Department of Viroscience for their encouragement and kindness. Special mention goes to, Loubina, Maria, Simone, Wim, Debby, Gijs, Werner, Sarah, Stefan, Oanh, Petra, Theo, Fasa, Noreen, Pascal, Sander, Rory, Miranda, Monique, Rik, Thijs, Ron, David, Guus, Charles, Georges, Nella, Felicity, Rachel.

My indebted thanks are also extended to Frank van Vliet for his valuable support during my I & I study and whenever I am financially in trouble.

I am thankful to my Ethiopian friends Tamirat, Girma, GebreMeskel, Adane, Daniel, Wondewosen, Bekele and Dejene for their encouragement, calling, skyping, and thus helping me not to feel homesickness and loneliness. My best friends, all your encouragement and memories are highly appreciated.

Lastly, I would like to thank my family for all their love and support. I have no words to express how grateful I am to my mom, brothers, sisters and uncle, who have provided me with moral and spiritual support and encouragement throughout my study. I heartily dedicate this thesis to my father who was always in my thoughts on this journey – you are missed. Above all, I thank Almighty God for giving me the strength, health and endurance.