Technische Universität München

TUM School of Life Sciences

Evolution of hepatotropic viruses during chronic infection in humans

Evolution of HBV and HDV

Zainab Usman

Vollständiger Abdruck der von der TUM School of Life Sciences der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktor der Naturwissenschaften

(Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender:

Prof. Dr. Johan Philipp Benz

Prüfende der Dissertation:

1. Prof. Dr. Dmitrij Frischmann
2. Prof. Dr. Ulrike Protzer

Die Dissertation wurde am 28.09.2020 bei der Technischen Universität München eingereicht und durch die TUM School of Life Sciences am 13.01.2021 angenommen.

Dedication

To my boys,

My husband, Usman,

for your patience, love, and utmost support.

And

My newborn son, Musa,

for your giggles and smiles that fills my heart with pure joy.

Abstract

Viruses are infecting wide variety of organisms such as bacteria, fungi, plants, animals, and human beings. While most viruses are benign, some viruses can cause serious health conditions in their host. Several viruses are organ specific and e.g. attack exclusively hepatocytes. To date, there are five different types of hepatitis viruses, A, B, C, D and E, affecting millions of people around the globe. Hepatitis B virus (HBV) is the most common infectious disease of the liver and a major cause of chronic hepatitis and cirrhosis, with increased risk of developing hepatocellular carcinoma (HCC). Whereas, Hepatitis D causes a severe form of viral hepatitis that can only be acquired either as a co-infection with HBV or as super-infection in people chronically infected with HBV.

The advent of Next Generation Sequencing (NGS) has replaced the traditional Sanger sequencing over the past decade. NGS has enabled the researchers to perform large scale genome sequencing much more efficiently to understand the low-frequency genetic variations underlying infectious diseases and investigate the drug resistant mutations associated with disease progression and severity. The first publication of this thesis focuses on harnessing the NGS technology to analyse the evolutionary dynamics of HBV quasispecies response to therapy with an amphipathic DNA polymer (REP 2139) in chronic HBV infection among the responders and non-responders, in order to establish a relationship between the HBV quasispecies complexity in response to the therapy and disease outcome. In addition, mutations occurring within the “a-determinant” during the treatment, which could affect the HBsAg production, secretion or detection were also assessed.

Over the past decade, a large number of Hepatitis D virus (HDV) sequences have been identified and sequenced. This poses a much-needed re-classification of Hepatitis D viruses to identify new genotypes or subtypes within the existing genotypes. The second publication of the thesis covers the sequence and phylogenetic analysis of HDV sequences. The phylogenetic analysis of 116 L-HDAg isolates from our cohort and 621 L-HDAg sequences from GenBank, confirmed the existing genotypes (1-8) and revealed five new subtypes within HDV genotype -1, two subtypes for HDV-2 and HDV-4 respectively. In addition, we also identified specific amino acid signatures for the proposed subtypes within the genotypes HDV-1, -2 and -4.

The third publication covers the creation of Hepatitis Delta Virus database (HDVdb: http://hdvdb.bio.wzw.tum.de/), hosted on the server of Department of Bioinformatics of Technical University Munich. The platform provides a manually curated dataset containing HDV sequences that includes set of complete genomes, L-HDAg, S-HDAg and partial cds

sequences. The database also provides a comprehensive set of tools that allows users to view and retrieve sequences as well as perform various sequence analysis, such as identifying genotypes, primer design, phylogenetic tree construction, and multiple sequence alignments.

Zusammenfassung

Viren infizieren eine Vielzahl von Organismen wie, Bakterien, Pilze, Pflanzen, Tiere und Menschen infizieren. Während die meisten Viren gutartig sind, können einige Viren in ihrem Wirt ernsthafte Krankheiten auslösen. Viele Viren, sind organspeziefisch und können z.B. nur Leberzellen (Hepatozyten) infizieren. Diese werden als Hepatitis-Viren bezeichnet. Bis heute gibt es fünf verschiedene Typen von Hepatitis-Viren, A, B, C, D und E, die Millionen von Menschen auf der ganzen Welt befallen. Hepatitis B Virus (HBV) ist die häufigste Infektion der Leber und eine Hauptursache für chronische Hepatitis und Leberzirrhose mit einem erhöhten Risiko, ein hepatozelluläres Karzinom (HCC) zu entwickeln. Hepatitis D virus verursacht eine schwerste Form der Virushepatitis, die nur entweder als Koinfektion mit HBV oder als Superinfektion bei chronisch mit HBV infizierten Menschen erworben werden kann.

Das Aufkommen der Next Generation Sequencing (NGS) hat in den letzten zehn Jahren die traditionelle Sanger-Sequenzierung ersetzt. Die NGS ermöglicht den Forschern, die Genomsequenzierung in großem Maßstab viel effizienter durchzuführen, um die niederfrequenten genetischen Variationen, die bei Infektionskrankheiten auftreten, zu erkennen und arzneimittelresistente Mutationen, die mit dem Fortschreiten und der Schwere der Krankheit verbundenen sind zu charakterisieren. Die erste Veröffentlichung der Dissertation konzentriert sich auf die Nutzung der NGS-Technologie zur Analyse der evolutionären Dynamik von HBV-Quasis-Spezies während der antiviralen Therapie mit einem amphipathischen DNA Polymer (REP 2139) bei chronischer Hepatitis B in Respondern und Non-Respondern, um eine Beziehung zwischen der Komplexität der HBV-Quasis-Spezies als Reaktion auf die Therapie und den Krankheitsausgang herzustellen. Darüber hinaus wurden auch die während der Behandlung innerhalb der "a-Determinante" auftretenden Mutationen bewertet, die die HBsAg-Produktion, -Sekretion oder -Detektion beeinflussen könnten.

In den letzten zehn Jahren wurde eine große Anzahl von Hepatitis-D-Virussequenzen identifiziert und sequenziert. Dies erfordert dringend eine Neuklassifizierung der Hepatitis-D-Viren dar, um neue Genotypen oder Subtypen innerhalb der bestehenden Genotypen zu identifizieren. Die zweite Veröffentlichung der Arbeit befasst sich mit der Sequenz und der phylogenetischen Analyse der Sequenzen des Hepatitis-D-Delta-Virus. Die phylogenetische Analyse von 116 L-HDAg-Isolaten aus unserer Kohorte und 621 L-HDAg-Sequenzen aus der GenBank bestätigte die bestehenden Genotypen (1-8) und ergab fünf neue Subtypen innerhalb des HDV-Genotyps -1, zwei Subtypen für HDV-2 bzw. HDV-4. Darüber hinaus identifizierten wir auch spezifische Aminosäuresignaturen für die vorgeschlagenen Subtypen innerhalb der Genotypen HDV-1, -2 und -4.

Die dritte Publikation befasst sich mit der Erstellung einer Datenbank für das Hepatitis-Delta-Virus (HDVdb: (http://hdvdb.bio.wzw.tum.de/), die auf dem Server der Abteilung für Bioinformatik der Technischen Universität München gehostet wird. Die Plattform bietet einen manuell kuratierten Datensatz mit HDV-Sequenzen, der einen Satz vollständiger Genom-, L-HDAg-, S-HDAg- und partieller cds-Sequenzen enthält. Die Datenbank stellt auch einen umfassenden Satz von Werkzeugen zur Verfügung, mit denen die Benutzer Sequenzen anzeigen und abrufen sowie verschiedene Sequenzanalysen durchführen können, wie z.B. die Identifizierung von Genotypen, Primerdesign, phylogenetische Baumkonstruktion und multiple Sequenzalignments.

Acknowledgements

Pursuing a Ph.D is a difficult and sometimes a lonely path. In pursuing this path, first of all, I would like to express my respectful gratitude to Prof Dr Dmitrij Frishman for giving me the opportunity to join his lab and introducing me to academic research. His extensive knowledge, technical guidance, and practical suggestions supported me throughout my doctoral program.

I would like to express my sincere gratitude to my mentor/advisor Prof Dr Michael Roggendorf for his unwavering guidance, encouragement and tremendous support during my research. He spent endless hours proofreading my manuscripts and gave me valuable suggestions which always resulted in improved versions of manuscripts. I am very grateful to Prof Dr Johan Phillip Benz who have kindly agreed to be chair of examination committee.

I would also like extend my appreciation to people I collaborated with during the course of my dissertation. Dr Hadi Karemzadeh, Dr Hrvoje Mijocevic, Dr Dieter Hoffman, Dr Suliman Afridi, Dr Hassan Moeini, and Stoyan Velkov helped me understand challenges in virology with many fruitful discussions and critical input including experimental data from the patients with HBV infection and NoroVirus.

I am grateful to all my former and present colleagues at our department for scientific discussions, entertaining lunch breaks and Christmas parties. I would like to thank Leonie Corry, Martina Rüttger, Roswitha Weinbrunn and Erna Look for helping with all the bureaucratic work and always having some time to spare to have a word and ask about ones well-being. Thanks to Drazen Jalsovec for managing all of our soft and hardware, installing various software and resolving dependencies issues. My gratitude also extends to Michael Kiening, Peter Hönigschmid, Evans Kataka, Bo Zeng, Fei Qu, Jinlong Ru, Jianfeng Sun, Hengyuan Liu, Hongen Xu, Alec Steep for being great colleagues whom I had the pleasure to work with during my doctoral studies.

I am thankful for the research scholarship I was granted by the DAAD Deutscher Akademischer Austauschdienst to finance my dissertation work.

Last not the least, my deepest gratitude goes to my husband and my best friend Usman Saeed, for believing in me and supporting me both emotionally and academically, teaching me programming skills and always coming to rescue in resolving system errors. None of this Ph.D work, however, would have been possible without your love, care and guidance.

Publications

Journal publications:

The three entries written in bold are part of this thesis.

- **Usman Z, Velkov S, Protzer U, Roggendorf M, Frishman D, Karimzadeh H. HDVdb: a comprehensive hepatitis D virus database. *Viruses*. 2020; 12: 538**
- **Usman Z, Mijočević H, Karimzadeh H, Däumer M, Al-Mathab M, Bazinet M, Frishman D, Vaillant A, Roggendorf M. Kinetics of hepatitis B surface antigen quasispecies during REP 2139-Ca therapy in HBeAg-positive chronic HBV infection. *J Viral Hepat*. 2019; 26 (12):1454-1464.**
- **Karimzadeh H, Usman Z, Frishman D, Roggendorf, M. Genetic diversity of hepatitis D virus genotype-1 in Europe allows classification into subtypes. *J Viral Hepat*. 2019; 26 (7): 900– 910.**
- Mijocevic H, Karimzadeh H, Seebach J, Usman Z, Al-Mahtab M, Bazinet M, Vaillant A, Roggendorf M. Variants of hepatitis B virus surface antigen observed during therapy with nucleic acid polymer REP 2139-Ca have no influence on treatment outcome and its detection by diagnostic assays. *J Viral Hepat*. 2019; 26(4): 485– 495.

Book publications:

- Karimzadeh H, Usman Z, Frishman D, Roggendorf M. “Expanded classification of Hepatitis D Viruses into subtypes.” In: Rizzetto M, Smedile A. Hepatitis D. Virology, Management and Methodology. Il Pensiero Scientifico. 2019. P 125

Contents

List of Figures

List of Tables

Acronyms

BCP	Basic Core promoter
Bp	Base pairs
BLAST	Basic local Alignment search tool
BWA	Burrows-Wheeler Aligner
CD-HIT	Cluster Database at High Identity with Tolerance
dN/dS	Ratio of nonsynonymous substitutions per nonsynonymous site dN to synonymous substitutions per synonymous site dS
DNA	Deoxyribonucleic Acid
HBV	Hepatitis B Virus
HBsAg	Hepatitis B surface antigen
HDV	Hepatitis D Virus
HCC	Hepatocellular carcinoma
HCV	Hepatitis C Virus
MCMC	Markov chain Monte Carlo
MEGA	Molecular Evolutionary Genetics Analysis
MHR	Major Histocompatibility Region
ML	Maximum likelihood
MP	Maximum Parsimony
NAP	Nucleic Acid Polymer
NGS	Next Generation Sequencing
NJ	Neighbor-Joining
NUC	Nucleos(t)ide analogues
ORF	Open Reading Frame
PCR	Polymerase Chain Reaction
PegIFN	Pegylated Interferon
PGRS	Proline Glycine Rich Segment
QS	Quasispecies
RefSeq	Reference Sequence database
RNA	Ribonucleic Acid
SD	Standard deviation
SDP	Specificity determining positions
SRA	Short Read Archive

1 Introduction

The following introduction is designed to provide information about the biological and methodological background as well as its implications for the approaches used in the published articles.

1.1 Viruses

Viruses are the most abundant and diverse biological entities on our planet ‘‘Earth‘‘, ubiquitous in both terrestrial and aquatic ecosystems, able to survive and thrive in extreme acidic, thermal, and saline environmental conditions. Viruses are small infectious agents that infects a living organism (i.e plants and animals) and hijacks the host cell’s machinery to reproduce. To date, researchers estimate that over 1 million virus species exist in mammals alone, with about half a million posing a significant threat to human health (Carlson et al. 2019). While most of the viruses (i.e plant viruses) are harmless, some viruses are very dangerous and can spread across different species with potentially fatal consequences eg Ebola, SARS, and Influenza etc.

1.1.1 Virus classification system

According to Baltimore classification system devised by David Baltimore (in 1971), viruses are categorized into seven groups depending on their genetic contents (i.e., DNA or RNA), strandedness (single-stranded or double-stranded), sense (+ or -), and replication strategies (Baltimore 1971). Viruses can be placed in one of the seven following groups:

I: Double-stranded DNA viruses

II: Single-stranded (+) sense DNA viruses

III: Double-stranded RNA viruses

IV: Single-stranded (+) sense RNA viruses

V: Single-stranded (-) sense RNA viruses

VI: Single-stranded (+) sense RNA viruses with DNA intermediate in life-cycle

VII: Double-stranded DNA viruses with RNA intermediate in life-cycle

1.2 Hepatitis B virus

Hepatitis B is a viral infection caused by the hepatitis B virus (HBV). The hepatitis B virus attacks the healthy liver cells causing swelling and inflammation which can progress to fibrosis, cirrhosis, and liver cancer also known as hepatocellular carcinoma or even death. Hepatitis B is transmitted through parenteral exposure to infective blood, semen or other body fluids (Hou, Liu, and Gu 2005). HBV can also be transmitted from an infected mother to a newborn at the time of birth, through sexual contact with HBV infected individual, use of non-sterile medical equipment and through sharing contaminated needles or syringes among the drug users (Harpaz et al. 1996, Shepard et al. 2006). Unlike other forms of hepatitis viruses, HBV is not spread through contaminated food or water, or by sharing cooking utensils and casual contact with infected HBV individual. Moreover, the virus can survive outside the human body for seven days and can transmit infection to healthy unvaccinated individual (Bond et al. 1981, Lok and McMahon 2009).

1.2.1 Discovery of HBV

Hepatitis B virus (HBV) is a member of the family of viruses referred as Hepadnaviridae, which is divided into two genera; Orthohepadnavirus (which infect mammals) and the Avihepadnavirus (which infect birds). In 1963, Bauch Blumberg identified an antigen protein in blood samples of an Australian aborigine and named it Australian antigen (AuAg) (Blumberg 1964). Few years later in 1967, the Australian antigen was officially recognized as a protein that causes Hepatitis B infection. Soon after, Blumberg and his colleague, Iriving Millman developed the blood test to detect the virus and invented the first hepatitis B vaccine in 1969 (Blumberg 1977).

1.2.2 Structure and Molecular virology

Hepatitis B virus also known as Dane particle was first observed by Dan et al in 1970 (Dane, Cameron, and Briggs 1970). HBV is a small, spherical virus particle with a diameter of 42 nm, consisting of an outer lipoprotein coat (also called envelope; 4 nm) with lipids and HBsAg, and the inner nucleocapsid (core; 27 nm) enclosing the viral DNA and DNA polymerase (Almeida, Rubenstein, and Stott 1971, Jokelainen et al. 1970). HBV genome is circular, partially double-stranded DNA molecule of about 3.2 kb in length. It contains 3020-3320 nucleotide long negative strand DNA and a 1700-2800 nucleotides long positive strand DNA (Delius et al. 1983).

HBV genome encodes four highly overlapping open reading frames (ORF) which are called C, P, S and X, respectively (Kay and Zoulim 2007, Kidd-Ljunggren, Miyakawa, and Kidd 2002). The longest ORF encodes the viral polymerase (Pol), the second ORF C encodes the nucleocapsid protein HBcAg and the non-structural precore protein HBeAg. The third

ORF S encodes the viral surface envelope proteins (HBsAg), containing three-in frame start (ATG) codons that divides the gene into three sections, preS1, preS2 and S producing polypeptides of three different sizes called large, middle and small envelope proteins. The fourth ORF encodes for the non-structural protein (HBxAg) with multiple functions. HBx is known to be involved in several cellular processes such as signal transduction, transcriptional activation, DNA repair and protein degradation inhibition (Ali et al. 2014).

1.2.3 HBV life cycle

The life cycle of Hepatitis B virus is complex (Figure 1.1). The hepatitis B virus gains entry into the host cell when pre-S1 of large envelope protein binds to sodium taurocholate co-transporting polypeptide (NTCP) receptors. NTCP are mainly expressed in liver transport bile salts and act as HBV entrance receptors (Ni et al. 2014, Yan et al. 2012). After binding to the receptors, the virus enters hepatocytes either by endocytosis or direct fusion of the virus envelope into the plasma membrane (Urban et al. 2010). The viral nucleocapsid containing partial double stranded relaxed circular DNA (ds-rcDNA) covalently bounded to the HBV polymerase, is released into the cytoplasm and transported to the nucleus via nuclear pore complex (NPC) (Blondot, Bruss, and Kann 2016, Ezzikouri et al. 2014). The viral polymerase completes the positive chain of rcDNA and both HBV DNA chains are converted into covalently closed circular DNA (cccDNA) by host DNA repair proteins. Viral cccDNA exists as stable mini-chromosomes that serve as transcriptional template to perform all viral RNA synthesis (i.e pre-genomic RNA and sub-genomic RNA) (Doo and Ghany 2010, Nassal 2015). The pre-genomic RNA (pgRNA) is involved in the synthesis of polymerase and core proteins, whereas subgenomic RNA encodes for the viral envelope proteins (HBsAg). The pgRNA are encapsidated into viral capsids with HBV polymerase and transported into the cytoplasm. Inside the capsid, pgRNA is reverse transcribed into cccDNA or occasionally dslDNA. These mature encapsulated rcDNA either exits the cell as infectious virions or are redirected to the nucleus where rcDNA is transcribed into cccDNA using intracellular conversion pathway. The dslDNA integrates randomly into the host cell genome and drives the expression of envelope proteins (Bill and Summers 2004).

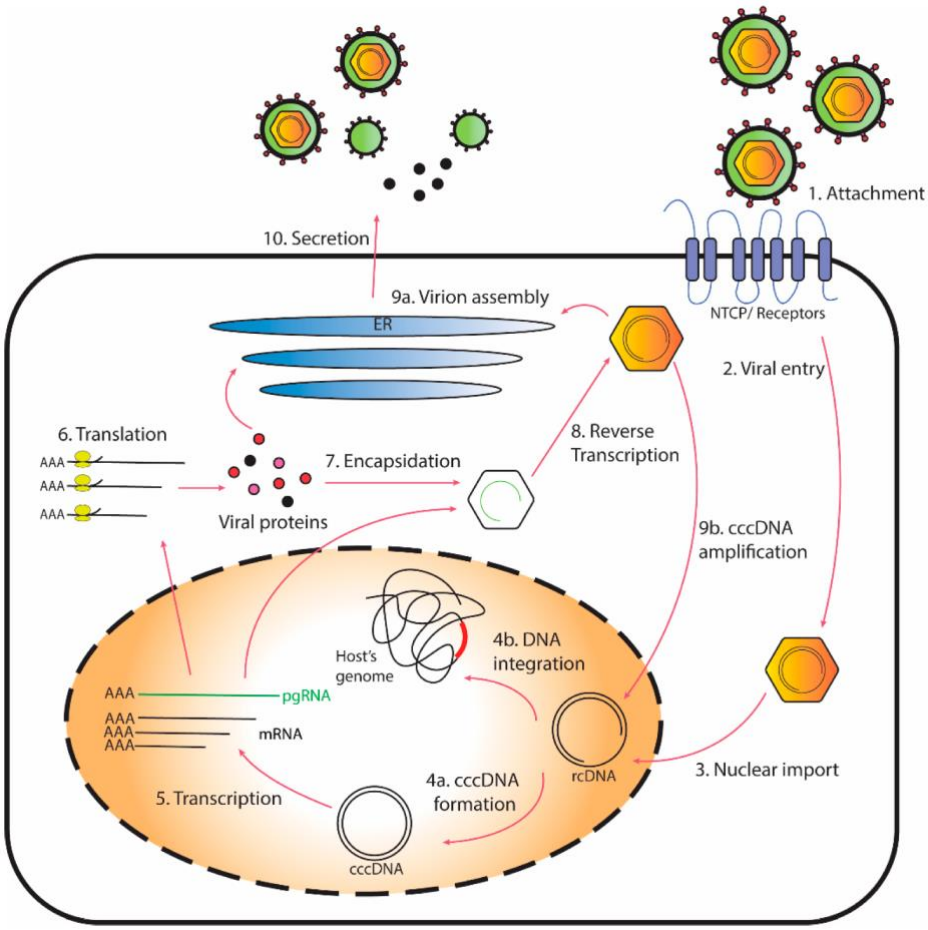

Figure 1.1: *Schematic representation of the Hepatitis B virus. The entire life cycle of HBV (1) attachment to the surface receptor NTCP, (2) viral entry into hepatocytes, (3) uncoating and the nuclear import of genome, (4a) formation of covalently closed circular DNA (cccDNA), (4b) DNA integration into host genome, (5) transcription of viral mRNAs (pregenomic RNA), (6) translation of HBV mRNAs into the large (L), middle (M), and small (S) surface, precore, core, polymerase (pol), and HBx proteins, (7) encapsidation of pgRNA and pol, (8) reverse transcription of viral DNA, (9) Assembly and (10) secretion of HBV virions from hepatocytes. (adapted from (Mohd-Ismail et al. 2019)*

1.2.4 Epidemiology

Hepatitis B infection is a serious health problem resulting in major chronic liver disease. According to the world health organization more than 290 million individuals are chronically infected with hepatitis B worldwide, resulting in 900,000 deaths each year (Polaris Observatory 2018). Hepatitis B virus has been classified into nine known genotypes (A-I) (Deny 2006), defined by >8% genomic divergence in the nucleotide sequence. These genotypes are further classified into 35 sub-genotypes with nucleotide divergence by 4% (Kramvis 2014, Pourkarim et al. 2014). Recently, a tenth putative genotype (J) isolated from a Japanese individual have been proposed (Tatematsu et al. 2009).

The HBV genotypes and sub-genotypes are associated with distinct geographical distributions; depending on the differences in their clinical consequences including the natural course of infection, transmission mode and treatment response (Kramvis 2016, Sunbul 2014). Genotype B, C and I are commonly found in Asia and Africa as well as other parts of developing countries with higher rates of vertical transmission (from mother to child) (Lin and Kao 2015). Genotype A, D and G are highly prevalent in Africa, Europe and Asia are associated with horizontal transmission via sexual contact or percutaneously by injection-drug use (Kramvis 2014, Liu and Kao 2013).

Based on the pathogenesis, the individuals infected with genotype A, C, D and F show faster disease progression to liver cirrhosis and hepatocellular carcinoma (HCC) (Kramvis and Kew 2007, Lin and Kao 2015, Liu and Kao 2013). Whereas, all other genotypes respond to interferon-alpha treatment, genotype A and B show an increased virological response and a higher anti-HBe seroconversion than other HBV genotypes (Lin and Kao 2015, Liu and Kao 2013). This treatment is not recommended for patients who are negative for the HBVe antigen (HBeAg) and infected with genotype D or E. Moreover, different stopping points are proposed for patients infected with genotypes A–D who have not responded to the therapy (Terrault et al. 2018).

1.2.5 Animal models

HBV infection is well documented for humans and chimpanzees. Chimpanzees are the only non-human primates that are immunocompetent and fully susceptible to HBV infection (Barker et al. 1973, Maynard et al. 1972) and result in both acute and chronic hepatitis B infection with immune response similar to humans (Wieland 2015). In addition to HBV, species-specific hepadnaviruses have been found in different animals such as eastern woodchucks (Marmota monax) (Summers, Smolec, and Snyder 1978), Beechey ground squirrel (Marion et al. 1986), and domestic ducks and goose species (Barker et al. 1975, Cova et al. 1990, Marion 1988). These viruses are known as woodchuck hepatitis virus (WHV), ground squirrel hepatitis virus (GSHV), and duck hepatitis virus (DHBV) respectively.

Among the non-primate species, rodents are the most commonly used animal model organism in medical research. Additionally, tree shrew (Tupaia belangeri, order Scandentia) are small mammals showing more sequence similarity to humans than rodents have been employed as experimental model to study viral hepatitis and hepatocellular carcinoma (Walter et al. 1996, Xiao, Liu, and Chen 2017). Similarly, WHV infection in Chinese woodchuck species (Marmota himalayana, Marmota baibacina, and Marmota bobak) strongly resemble human HBV infection, therefore, these Chinese Marmota species are

used as animal models to develop new antiviral vaccines in treating chronic hepadnaviral infections (Wang et al. 2014, Wang et al. 2011).

Furthermore, in vivo studies involving infection of duck hepatocytes with duck hepatitis virus (DHBV) has contributed substantially to our understanding in elucidating the mechanism of hepadnavirus replication and antiviral therapy (Mason 2015).

1.2.6 Hepatitis B infection

Hepatitis B virus (HBV) infection that attacks the liver of healthy individual causes wide range of clinical manifestations which result in either an acute infection or a chronic infection. In acute infection, most individuals do not experience any symptoms and recover from the illness within 6 months after eradication of virus in the presence of the immunoglobin antibodies. However, some individuals experience symptoms such as nausea, vomiting, jaundice and abdominal pain. In chronic infection, the individuals continue to remain positive for the hepatitis B virus for more than 6 months of infection. These individuals are at higher risk of developing chronic liver inflammation and fibrosis, which may lead to liver cirrhosis and hepatocellular carcinoma (HCC).

1.2.7 Antiviral immune response

Most pathogens upon entry into the host cell trigger first set of immune response known as innate immune response. Innate immunity activates set of cellular receptors such as natural killer cells (NK) which secrete a wide variety of antiviral cytokines (interferon-γ (IFN-γ), tumour necrosis factor-α (TNF-α) to destroy infected host cells and limit the infection from spreading. However, if the first line of defence does not clear the viral infection, the adaptive immune response involves activation of B lymphocytes and plasma cell which produce antibodies (CD+4 helper T cells, CD8+ cytotoxic T cells) against the viral envelope antigens leading to the clearance of circulating viral particles. This adaptive immune response has a memory component that protects the host from subsequent risk of reinfection.

In vivo studies with experimentally infected chimpanzees have been used to study host immune response to HBV infection. The liver biopsies have showed that HBV does not induce an interferon-stimulated gene (IFN/ISG) expression early in the infection as it spreads through the liver (Cheng et al. 2017, Wieland et al. 2004). On the contrary, several studies have reported that HBV infection activates innate immunity and triggers the production of type I and III interferons. (Sato et al. 2015, Shlomai et al. 2014).

Experimental studies have shown that HBV virions interfere with immune signalling pathways which leads to inhibition of interferons despite being recognized by the pattern recognition receptors (PRRs) (Cui et al. 2016, Luangsay et al. 2015, Liu et al. 2015, Kumar et al. 2011). According to one study, in early stages of HBV infection the hepatocytes

suppress IFN -β induction in response to stimulation with poly(I/C) or Sendai virus (SeV) infection (Luangsay et al. 2015). Despite this substantial evidence acquired in understanding the pathogenesis of HBV, our knowledge of the immune mechanisms of HBV are hampered due to technical and ethical constraints limited to HBV host range i.e. humans and chimpanzees.

1.2.8 Viral Quasispecies

Most RNA viruses exhibit very high replication rates (HBV has a high mutation rate 10^{-5} to 10^{-4} base substitutions/site/year) and lack the proof-reading ability of viral polymerase during reverse transcription. (Osiowy et al. 2006, Zhou and Holmes 2007). The accumulation of mutations results in genetic diversity in the form of genotypes, subtypes and haplotypes also referred as quasispecies (QS). Quasispecies theory was initially formulated by Manfred Eigen and Peter Schuster to explain the origin of life (Eigen and Schuster 1977, Eigen 1971). Quasispecies are swarm of closely related but genetically non-identical viral variants capable of surviving, rapidly adapting and propagating in a complex mutagenic environment (Domingo and Gomez 2007).

Viral quasispecies mutate and evolve continuously by establishing a balance of mutations and selection pressure (Figure 1.2) (Comas, Moya, and Gonzalez-Candelas 2005). Variants with varying fitness experience both positive selection and negative (purifying) selection (Domingo and Holland 1997). In positive selection, the variants with high fitness becomes dominant in the viral population overtime. On contrary, negative selection is a process by which deleterious variants are removed from the viral population. Also, negative selection allows maintenance of low fitness variants that might out-compete higher fitness variants provided that these high fitness variants only constitute minor proportion of the entire viral population (de la Torre and Holland 1990, Domingo, Sheldon, and Perales 2012). For a schematic distribution of a viral quasispecies population, see Figure 1.2.

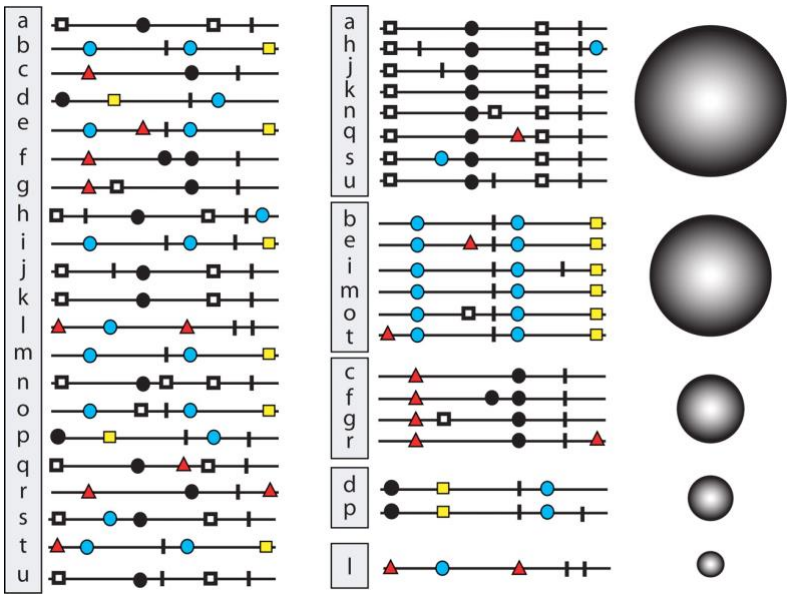

Figure 1.2: *Schematic quasispecies distribution. The left panel shows a schematic of a complete viral population with symbols representing mutations. The middle panel reveals the quasispecies distribution, after sorting all sequences by their shared, species defining, mutations. Some variants do have additional, individual mutations or sequencing errors. The right side visualises the proportion each species takes up from the overall population. (adapted from (Domingo and Gomez 2007)).*

Over the years, several bioinformatics tools have been developed to reconstruct viral quasispecies from high throughput sequencing data. In general, quasispecies are reconstructed either locally, in a genomic region which is completely covered by the average read length (Ezzikouri et al. 2014, Macalalad et al. 2012, Zagordi et al. 2010), or globally, over a large gene of interest or across the entire genome such that size of genomic region exceeds the read length (Salzberg et al. 2012). Eriksson and his coworkers proposed one of the earliest methods for viral population estimation using next-generation sequencing (NGS) (Eriksson et al. 2008). Since then, several methods have been developed for reconstruction of viral quasispecies which includes a probabilistic clustering algorithm ShoRAH (Zagordi et al. 2011), a combinatorial algorithm QuRe (Prosperi and Salemi 2012), a hidden markov model based scheme QuasiRecomb (Topfer et al. 2013), a novel reference assisted denovo assembly algorithm ViQuaS (Jayasundara et al. 2015), a probabilistic mixture model PredictHaplo (Prabhakaran et al. 2014) and an iterative maximal clique enumeration approach HaploClique (Topfer et al. 2014). Recently, low diversity (≤5%) quasispecies are constructed using sequential Bayesian inference method called aBayesQR (Ahn and Vikalo 2018).

The information about genetically heterogeneous quasispecies play a vital role in understanding the host-viral interaction processes, anti-viral drug resistance and immune escape mutations (Mese et al. 2013, Rodriguez et al. 2013, Zhang et al. 2016). Hence, an

important aspect in the quasispecies reconstruction is to distinguish real mutations from the technical errors resulting during the sample preparation and sequencing. It is a vital factor in understanding many aspects of viral evolution, therefore, in recent years several bioinformatic tools have been developed by different sequencing technologies to overcome sequencing limitations (Prosperi and Salemi 2012, Topfer et al. 2013, Zagordi et al. 2011).

1.2.9 Treatment of Hepatitis B

There is no specific treatment to cure acute hepatitis B infection. However, acute hepatitis B patients are advised to rest well and maintain a healthy and nutritionally balanced diet by increasing fluid intake and avoiding alcohol consumption, smoking, and fatty/greasy foods that will prevent the development of fatty liver and its progression to chronic liver disease. Additionally, it is recommended to avoid unnecessary medications, which include use of vitamins, pain-killers (paracetamol or ibuprofen) and acetaminophen against vomiting.

Chronic hepatitis B (CHB) treatment is limited to oral antiviral agents, either interferon (standard or pegylated INF) or Nucleos(t)die analogues (NUCs). In 1991, interferon-alpha (IFN-α) was approved as antiviral treatment for CHB, later substituted with pegylated interferon (Peg-INF-α) with improved pharmacokinetic properties and greater antiviral efficacy than the regular IFN (Craxi and Cooksley 2003). The two major forms of Peg-IFN (Peg-IFN- α 2a and Peg-IFN- α2b) are used to treat both HBeAg-positive and HBeAg-negative chronic hepatitis B patients (Cooksley 2005, Woo, Kwok, and Ahmed 2017).

In addition, five nucleoside analogues (NUCs) effectively used to treat chronic hepatitis B infection include lamivudine (LMV), telbivudine (TBV), adefovir dipivoxil (ADV), tenofovir (disoproxil fumarate, TDF; alafenamide, TAF) and entecavir (ETV) (Jordheim et al. 2013, Lampertico and Liaw 2012). Nucleoside analogues suppress the HBV replication by inhibiting the reverse transcriptase activity of HBV DNA polymerase during the antiviral therapy, with little or no risk of antiviral resistance (Gish et al. 2012, Zoulim and Locarnini 2009). Several studies have reported that the mutations in reverse transcriptase region can help in establishing a correlation between the evolutionary dynamics of HBV quasispecies resulting in response to NUC antiviral therapy and the clinical outcomes of the treatment (Liu et al. 2011, Yang, Huang, et al. 2015).

Recently, Nucleic acid polymers (NAPs) are characterized as the newest member of antiviral polymer family of compounds with broad spectrum antiviral activity against several enveloped viruses including hepatitis B and hepatitis D virus (Vaillant 2016). In HBV infection, NAPs have a unique ability to block the release of the hepatitis B virus surface antigen (HBsAg) from the HBV-infected hepatocytes in ducks (Noordeen et al. 2015, Noordeen, Vaillant, and Jilbert 2013b, a). The effective inhibition of HBsAg assembly and release was confirmed in vitro in HepG2.2.15 cells (Guillot et al. 2017). The clearance of

HBsAg during NAP therapy is accompanied by substantial clearance or reduction of HBV DNA from the blood circulation. These effects were evaluated and validated in several clinical trials in hepatitis B e antigen (HBeAg)-positive, HBeAg-negative, HBV mono-infected, and HBV/HDV (hepatitis D virus) coinfected patients (Al-Mahtab, Bazinet, and Vaillant 2016, Bazinet et al. 2017). In recent studies, NAPs (REP 2055 and REP 2139) antiviral performance have been investigated in HBV infected ducks, woodchucks chronically infected with WHV and other rodent models to improve our understanding on NAPs mode of action in hepadnaviruses (Quinet et al. 2018, Schoneweis et al. 2018).

1.2.10 Vaccines against HBV

Hepatitis B infection can only be prevented if the individual is vaccinated. In 1981, the first commercial vaccine "Heptavax" approved by FDA was developed from an inactivated purified HBsAg extracted from the plasma of chronically infected HBV individuals (Vitaliti et al. 2013). Later, in 1986, an approved concept of genetically engineered hepatitis B vaccines replaced the conventional vaccines. These vaccines were synthetically prepared and did not contain any of the blood products.

The most effective way to prevent hepatitis B infection is to get vaccinated and immunized during infancy. Thus, it is highly recommended that all newborn infants receive first dose of HepB vaccine at the beginning of birth, followed by second and third dose given at 1-2 months and 6-18 months of age. In 2017, a new hepatitis B vaccine “Heplisav-B“(Dynavax) approved by FDA is commercially used for adults aged 18 and older. The vaccine is administered in two doses over one month apart and provides immunization against all known subtypes of hepatitis B virus (Splawn et al. 2018).

1.3 Hepatitis D virus

Hepatitis Delta is one of the five known hepatitis viruses caused by the hepatitis delta virus (HDV). HDV is the smallest known RNA virus, unable to infect hepatocytes by itself and requires the presence of helper Hepatitis B virus for its infection and replication. The infection causes inflammation of liver, which eventually leads to liver cirrhosis, fibrosis, liver decompensation and hepatocellular carcinoma.

1.3.1 Discovery of HDV

Hepatitis D virus (HDV), was identified in 1977 by Mario Rizzetto and colleagues as a novel antigen-antibody system (called δ antigen/anti-δ antibodies) in the liver biopsies of individuals infected with hepatitis B virus (HBV) (Rizzetto et al. 1977). Years later, subsequent experiments in chimpanzees confirmed that δ antigen was a structural component of transmissible pathogen (HDV) which requires the presence of HBV to complete its life cycle (Rizzetto et al. 1980). In fact, HDV requires the HBV envelope

proteins (HBsAg; Hepatitis B surface antigen) to act as a helper for the assembly and transmission of HDV virions. For this reason, the HDV is considered as a defective or satellite virus of hepatitis B virus (HBV).

1.3.2 Genomic structure of HDV

HDV is a small, spherical particle of about 35-37 nm in diameter, and is considered the smallest virus known to infect humans (Wang et al. 1986). The viral genome is a single-stranded, negative-sense circular RNA molecule of approximately 1680 nucleotides, with an extremely high degree of intramolecular complementarity (self-annealing) (~74%) that results in an unbranched rod-like structure (Lai 1995, Taylor 2006). The genome expresses single open reading frame (ORF) which encodes the delta antigen (HDAg) in two different forms; the small delta antigen [S-HDAg; 195 amino acids, 24 kDa] essential for replication of RNA genome and large delta antigen [L-HDAg; 214 amino acids, 27 kDa] which arises as a consequence of posttranslational RNA editing event results in additional 19 amino acid at is the C terminus (Casey 2012, Chang et al. 1991) (Figure 1.3).

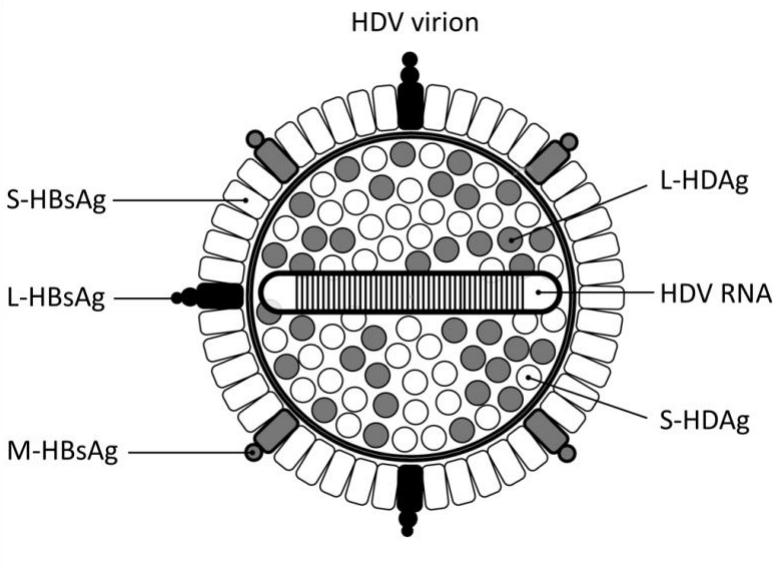

Figure 1.3: *Structural representation of HDV virion*. *Genomic RNA folds into rod-like structure to form ribonucleoprotein (RNP) surrounded by HBsAg envelope proteins, the S-, M- and L-HBsAg modified from (Sureau 2006).*

The delta antigen isoforms (S-HDAg and L-HDAG) consists of three functional domains, coiled-coil sequence also known as dimerization domain (CCD, 12-60 amino acid), nuclear localization signal (NLS, 66-75 amino acids) sequence (Alves, Freitas, and Cunha 2008) and RNA-binding domain (970-146 amino acid) (Lee et al. 1993). In addition, L-HDAg comprises of an additional functional domain, a nuclear export signal (NES, 198-210 amino acid) (Lee et al. 2001). Both S-HDAg and L-HDAg are localized in the nucleus since they carry nuclear localization signals (NLS). However, only L-HDAg is translocated to HBsAg

in the cytoplasm as it contains nuclear export signal (NES) required for the assembly of HDV virions (Lee et al. 2001).

1.3.3 Life cycle of HDV

Although the life cycle of HDV is not clearly understood, it is assumed that HDV exhibits the same mechanism of entry into the hepatocytes as HBV, since both viruses share common envelope protein (HBsAg) (Sureau and Negro 2016). HDV virion enters the hepatocyctes by binding to negatively charged heparin sulphate proteoglycans (HSPGs) present on the surface of host cell membrane (Lamas Longarela et al. 2013). In 2012, Yan and his colleagues identified human sodium taurocholate co-transporting peptide (hNTCP, encoded by SLC10A1), a putative receptor for HDV and HBV entry into hepatocytes (Yan et al. 2012). The myristoylated N-terminal amino acids of Pre-S1 region of L-HBsAg is essential for both HDV and HBV entry into the hepatocytes (Ni et al. 2014, Yan et al. 2012).

Following the attachment and entry of HDV virion into the hepatocytes, the virus uncoats and the viral genome released in the cytoplasm is transported by HDAg to the nucleus where replication takes place. Unlike other RNA viruses, HDV lacks its own RNA-dependent RNA polymerase and requires the host cell DNA-dependent RNA polymerase to replicate its genome (Modahl et al. 2000, Taylor 2009). RNA polymerase I and II interacts with S-HDAg and mediates replication through rolling-circle mechanism and results in synthesis of three different RNA transcripts, circular genomic RNA, circular complementary anti-genomic RNA and a linear anti-genomic RNA (Abeywickrama-Samarakoon et al. 2018, Yamaguchi et al. 2001). During replication, post translational modification of amber/W site on linear anti-genomic RNA by adenosine deaminase 1 (ADAR1), results in extension of an additional 19 amino acid which leads to the production of L-HDAg (Casey 2012).

1.3.4 Animal models of HDV

For many years, researchers considered HDV infection rare and limited to humans only. As mentioned previously, HDV requires helper functions of HBV to propagate HDV infection. Thus, the host range of HDV is limited to species that support the HBV replication. Among the animal models, chimpanzees and eastern woodchucks are known hosts of human HBV and WHV1 and are capable of supporting HDV replication and propagation (Casey and Gerin 2006, Negro et al. 1988, Negro et al. 1989).

Recent studies have demonstrated novel discoveries of HDV-like sequences in animals such as snakes (Boa constrictor) (Hetzel et al. 2018), ducks (Anas species) (Wille et al. 2018), rodents (Proechimys semispinosus) (C. Drosten personal communication) fish, amphibians and invertebrate (termites) (Chang et al. 2019). Most of these viruses have similar genomic features including size, circular and unbranched rod-like structures. The

snake derived HD protein bears 50% with the L-HDAg (Hetzel et al. 2018) whereas the duck associated HDV protein shares homology of 32% (Wille et al. 2018).

These exciting new findings based on different animal models are not only important in understanding the origin of HDV, but also shed light on main characteristics of HDV infection, mechanisms of host immunological response, development of novel treatment and vaccination strategies to cure and control HDV infection.

1.3.5 Epidemiology

The HDV RNA exhibits significant sequence heterogeneity, with an estimated mutation rate of 3×10^{-2} and 3×10^{-3} base substitutions per nucleotide per year (Lee et al. 1992). HDV infection has been reported from around the globe, with variable prevalence rates, clinical features and epidemiological characteristics (Nath et al. 1985). Globally, approximately 15-20 million people are infected with HDV (Alfaiate, Deny, and Durantel 2015, Chen et al. 2018, Rizzetto 2015).

The genetic heterogeneity is different along the HDV genome with highly conserved structural and functional domains (Chao et al. 1990). Initially, the nucleotide sequence analysis of the delta antigen-coding region in the antigenomic strand of HDV RNA classified the HDV isolates into three distinct genotypes (Casey et al. 1993). At present, great number of HDV cases have updated the proposed HDV nomenclature to eight major genotypes, HDV-1 to HDV-8 recognized with distinct geographical distributions and associated clinical features (Alvarado-Mora et al. 2013, Radjef et al. 2004, Le Gal et al. 2006). HDV is highly endemic in populations with a high prevalence of HBV infection, such as in Middle eastern, Amazonian, Mediterranean, and African countries (Husa et al. 2005).

In recent years, several countries in Europe (Italy, Spain, Germany etc.) (Gaeta et al. 2000, Navascues et al. 1995, Sagnelli et al. 1997), South East Asia (China, Taiwan) (Huo et al. 1997, Yang, Yu, et al. 2015) and Turkey (Degertekin, Yalcin, and Yakut 2006) experienced significant decline in HDV prevalence. This is mainly as a result of successive HBV vaccination campaigns, accelerated efforts for hepatitis B prevention and treatment and improved socioeconomic conditions. However, the incidence of HDV infection still persists in western countries due to increased immigration from endemic regions, intravenous drug use, high risk sexual practices and reuse of unsterilized equipment's for medical and dental care (Gish et al. 2013, Ho et al. 2013, Jakubovics, Greenwood, and Meechan 2014, Kucirka et al. 2010).

1.3.6 Pathogenicity

HDV infection is clinically associated with two major modes of infection: HBV/HDV co-infection and HBV/HDV super-infection. A third, rare mode of infection is reported as helper

independent latent infection. Co-infection, is simultaneous infection with both HDV and HBV clinically similar to HBV mono-infection, translates into acute hepatitis which leads to severe liver disease like fulminant hepatitis with 90-95% recovery rate in adults (Yurdaydin et al. 2010). Super-infection develops in individuals with pre-existent or persistent HBV infection, evolves into a chronic infection in 90% of the cases which rapidly progresses to chronic liver disease, cirrhosis and liver cancer commonly known as hepatocellular carcinoma (HCC) (Negro 2014). Helper independent latent infection occurs after liver transplantation, where the new liver graft gets infected with HDV without the presence of HBV. HBV is prevented by administering hepatitis B immunoglobins. The infection may remain asymptomatic only if and when HBV is reactivated with reappearance of HBeAg and high levels of HBV DNA (Ottobrelli et al. 1991).

1.3.7 Treatment

There are currently no vaccine available to prevent Hepatitis D infection. However, prior immunization against hepatitis B greatly reduces the risk of developing hepatitis D infection, since HDV requires the HBV surface antigens (HBsAg) to cause infection. Currently, limited therapeutic options are available to treat hepatitis D infection. A successful treatment of HDV infection involves complete eradication of HDV (i.e. clearance of both HDV RNA and HDAg) accompanied by decline in serum levels of HBsAg.

Since 1980s, Interferon-α (IFN-α) have been used to treat hepatitis D infection. Several studies involving long-term administration of recombinant interferons (IFN-α) have shown antiviral activity in most clinical trials (Farci et al. 1994, Hadziyannis 1991). In 2006, standard IFN-α was replaced by long-lasting pegylayed IFN-aplha (Peg-IFN-α) administered weekly for atleast one year demonstrated better virological response when compared with standard IFN-α (Castelnau et al. 2006, Erhardt et al. 2006, Niro et al. 2006).

Other antiviral drugs such lamivudine, adefovir dipivoxil, famciclovir, entecavair and nucleic acid analogues have displayed promising results in treating chronic hepatitis B individuals. However, these antiviral agents appear ineffective against HDV in either monotherapy or in combination with IFN-alpha (Kabacam et al. 2012, Lau et al. 1999, Niro et al. 2006, Wolters et al. 2000, Yurdaydin et al. 2002).

Recently, new antiviral therapeutic agents are under clinical trials to treat chronic hepatitis D infections. Pegylated interferon lambda (PEG-IFN-λ) is a type III interferon induces interferon stimulated genes (ISG) that promotes cell-mediated immune response against the viral infection. IFN-λ showed comparable antiviral effects in HBV/HDV infected humanized mice (Giersch et al. 2017). IFN-λ is under phase II clinical trials to assess the effectiveness against HDV both in monotherapy and in combination with lonafarnib and ritonavir.

Lonafarnib is a farnesyl-transferase inhibitor (FTI) that prevents the farnesylation of C-terminal Cys211 residue in L-HDAg. The clinical trials revealed decline in HDV RNA serum levels both in vitro and in mouse models (Bordier et al. 2002, Bordier et al. 2003). These reductions are directly associated with decrease in HDV virion assembly and secretion from hepatocyctes (Koh, Canini, et al. 2015, Yurdaydin et al. 2018). Myrcludex B is a myristolated lipopeptide consisting of pre-S1 N-terminal (2-48 amino acids) of L-HDAg and inhibits the entry of HBV and HDV in hepatocytes (Gripon, Cannie, and Urban 2005). This drug targets the bile acid transport function of sodium taurocholate co-transporting polypeptide (NTCP) and inhibits virus uptake in the cell (Blank et al. 2016, Blank et al. 2018, Bogomolov et al. 2016).

Nucleic acid polymer (REP 2139-Ca) are negatively charged oligonucleotides with broad-spectrum antiviral activity against several viruses (HIV, HSV and LCMV) (Vaillant 2016). Although, NAPs mode of action remains debatable, NAPs tested in duck HBV model system has shown to interfere with both the duck HBV virus entry and release of DHBsAG from the hepatocytes (Noordeen et al. 2015, Noordeen, Vaillant, and Jilbert 2013b). The safety and efficacy of NAP REP 2139-Ca were assessed in HBV infected patients and chronic HDV patients (Bazinet et al. 2017). Both trials showed reductions in HBsAg serum levels accompanied by strong reductions of HBV DNA.

1.4 Cloning and nucleotide sequencing

In 1975, Frederick Sanger and his colleagues developed an original DNA sequencing method based on incorporation of dideoxynucleotide (causing chain termination) by DNA polymerase during in-vitro DNA replication (Sanger 1975). This first-generation DNA sequencing method also known as Sanger sequencing or the chain termination method is considered gold standard in clinical research (Liu et al. 2011). However, this method is expensive, time-consuming, labour intensive, and incapable of detecting low frequency mutations. Therefore, Sanger sequencing services are now being discontinued in different clinical laboratories.

Over the past decade, rapid advances in high-throughput sequencing technologies have led to the development of more efficient and cost effective methods to sequence large genomes (Kircher and Kelso 2010). These are also known as second generation sequencing often referred as next generation sequencing (NGS) methods. These technologies allow massive parallel sequencing of long DNA or RNA fragments with increased sensitivity of detecting low-frequency genetic variants in a heterogeneous population. The commercially available second generation NGS technologies include instruments from Illumina, Roche 454 and Life Technologies.

In recent years, “Nanopore sequencing”, also known as “third generation“ sequencing technology is used to sequence long DNA or RNA molecules (up to 2MB reads) without the need for PCR amplification and chemical labelling of the sample. Among the third generation sequencing technologies include Pacific Biosciences (PacBio), Illumina Tru-Seq Synthetic Long-read technology and Oxford Nanopore technologies.

1.4.1 Next Generation sequencing pipeline

The pipeline used to analyse NGS data comprises of two main sections, wet-lab experiments followed by bioinformatics analysis of sequence data. The wet-lab experiments involve sample preparation, and library amplification, whereas, the bioinformatics analysis comprises of customized workflow of integrated tools and methods to analyse biological data. The bioinformatics tools include sequence analysis algorithms, computational and statistical techniques to analyse sequences. In NGS, different platform technologies follow the same eight major steps (See Figure 1.4).

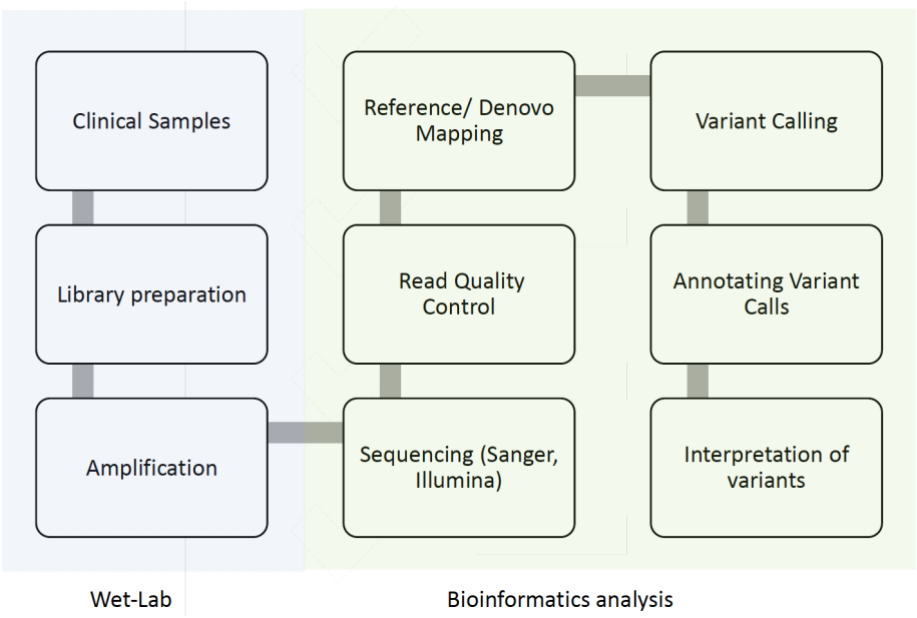

Figure 1.4: *Workflow of NGS data analysis. The steps involved in high-throughput sequencing of biological data (1) library preparation, (2) amplification, (3) sequence reads, (4) quality control/ read filtration, (5) alignment, (6) variant calling, (7) annotating variant calls, (8) interpretation of variants.*

i. **Library preparation**: The first critical step in NGS workflow involves preparing high quality and high-yield sequence library. The genomic isolates (DNA or RNA) are sheared into smaller fragments (150-5,000 bp) depending on the sequencing platform. The desired fragment libraries are generated using two different fragmentation methods, mechanical shearing or enzyme-based fragmentation (Head et al. 2014, Knierim et al. 2011). Mechanical shearing methods include needle-shear (such as tuberculin syringe), nebulisation, acoustic shearing (high frequency energy waves) and sonication (ultrasound).

On the contrary, enzyme-based methods are based on transposons and restriction enzymes (such as endonucleases) (Marine et al. 2011). The small fragments as a result of fragmentation process known as reads, have short overhangs (sticky ends) of 5'-phosphate and 3'-hydroxl groups. These ends are repaired by adenylation at 3' end resulting in adapter ligation that is important for amplification. During the library preparation, individual barcode sequences are added to the DNA fragments (i.e. reads) to facilitate multiple sequencing of various samples in the same run (Dodt et al. 2012).

ii. **Amplification**: After library preparation, the next step in NGS pipeline is clonal amplification to generate hundreds and thousands of copies for each read. The library is loaded onto the flow-cell and the fragments are amplified using clonal amplification methods such as emulsion PCR or bridge amplification. In emulsion PCR, amplification is carried out within water droplet floating on an oil solution (Dressman et al. 2003, Nakano et al. 2003). In bridge amplification, the denatured DNA fragments (ssDNA) is hybridized to the flow-cell's surface bound forward and reverse oligos that are complementary to the library adapters sequences. Hybridized at one end, the singe stranded DNA then folds over to form a bridge and bind to adapter complementary oligos at the other end. DNA polymerase adds nucleotides to amplify DNA, and clonal cluster is generated as the original strand is washed away leaving complementary strands of amplified DNA attached to the flow cell (Pemov et al. 2005).

iii. **Sequencing**: The amplified DNA fragments are sequenced using different NGS technology platforms that are commercially available, such as Roche 454 sequencing, Illumina (Solexa) sequencing, Solid Ion Torrent (Proton / PGM sequencing), Pacific Biosciences and Nanopore technology (Table 1.1).

Roche 454 sequencing is based on sequencing-by-synthesis approach called pyrosequencing. A technique that detects pyrophosphate release (PPi) after the nucleotides are incorporated by polymerase to a new strand of DNA. Roche 454 sequencing produces sequence reads with average read length of 700bp and some reach up to 1000 bp in length. However, 454 pyrosequency technology has been discontinued in 2013.

Illumina (Solexa) sequencing is based on four color sequencing-by-synthesis approach that works by identifying DNA bases (A, T, C or G). Each base emits a unique fluorescent signal as it is added to the nucleic acid chain. Illumina offers sequencing with longer reads ranging between 100 to 300 bp in length. Illumina platforms mainly differ in read length, run time and the amount of DNA sequenced in each run (Table 1.1).

Table 1.1: *Comparison of next generation sequencing platforms. Different attributes and key features of different sequencing platforms include company, sequencing platforms, run time, maximum output, and read length.*

Company	Sequencing Platforms	Run time	Max Output	Read Length
Roche/454	Roche-GS FLX	8 hours	0.7 Gb	700bp
Illumina	iSeq Series	9-17.5 hours	1.2 Gb	2 x150 bp
	MiniSeq Series	4-24 hours	7.5 Gb	2 x 150 bp
	MiSeq Series	4-55 hours	15 Gb	2 x 300 bp
	NextSeq Series	13-20 hours	120 Gb	2 x 150 bp
	HiSeq Series	<1-3.5 days	1500 Gb	2 x 150 bp
	HiSeq X Series	<3 days	1800 Gb	2 x 150 bp
Life Technologies	SOLiD 5500 x1	6-10 days	30 Gb	75 bp
	Ion Torrent PGM	4-7 hours	2Gb	200-400 bp
	Ion Torrent Proton	~2-4 hours	10 Gb	2 x 200bp
Pacific Biosciences	PacBio RS	4 hours	1 Gb	>10-15 kb
	PacBio RS II	~2 hours	10 Gb	>10-20 kb
Oxford Nanopore	Nanopore MinION	1-2 days	30 Gb	>200 kb

SOLiD sequencing uses emulsion PCR and a dye-labelled oligonucleotide ligation approach (such as DNA ligase). The maximum read length of SOLiD sequencing is 75 bases.

Ion Torrent (Proton / PGM sequencing) measures the release of H+ (protons) during the incorporation of nucleotide by DNA polymerase. This technology differs from the previous two methods as it does not measure light intensity. As in other kinds of NGS, the input DNA or RNA is fragmented into ~200 to 400 bp in length.

PacBio sequencing uses SMRT (Single Molecule Real Time) technology where single DNA molecule is sequenced by immobilized DNA polymerase inside a ZMW (Zero-Mode Wavelength). The fluorescent dNTPs releases a small amount of light during nucleotide base incorporation. The light signal is tracked in real time. PacBio reports an average read length of over 10,000 bp and can reach upto 60,000bp.

Nanopore sequencing works by measuring the changes in the electric current as nucleic acids are passed through a protein nanopore. Oxford Nanopore technologies developed a portable nanopore DNA sequencer referred as the MinION. It is highly compact and efficient sequencer that connects to a computer via USB. The handheld device was first tested during Ebola virus outbreak in west Africa (Hoenen 2016, Hoenen et al. 2016). It rapidly sequenced Ebola samples on-site within 15-60 minutes.

All sequencing technologies offers raw sequence data in the form of reads (20 to 1000 bp) stored in FASTQ format which contains both nucleotide sequence and its corresponding quality scores. These reads can be either 'single-ended' or 'paired-ended'. Paired end reads are produced when the fragment size is longer than 500 bp.

iv. **Quality control & Read Filtration**: After sequencing is complete, the read data is in electronic form can be processed to generate whole genome or a specific gene sequence by using bioinformatics NGS pipeline. Although quality control and filtration is the fourth step in the generation of full analysable sequence, it is first step in the bioinformatics NGS pipeline. The quality control is a vital step in assessing the quality of the raw sequenced reads using FastQC (Andrews 2010) and produces a detailed report on the number, quality and coverage of reads. The read filtration involves identifying and removing low confidence and erroneous reads, followed by clipping of adapters and truncating low quality bases from 3' and 5' end of the reads (PHRED threshold <20) using software's such as CutAdapt (Martin 2011), Trimmomatic (Bolger, Lohse, and Usadel 2014) etc.

v. **Alignment**: Once the high read quality data is obtained, millions of raw sequence reads (single-end or paired-end) are mapped and aligned using either a reference based assembly (in which reference sequence is available) or denovo assembly (in the absence of a reference sequence). The sequence reads of variable lengths are aligned using different bioinformatics alignment tools such as BWA (Li 2013), Bowtie (Langmead et al. 2009), and TopHat (Trapnell, Pachter, and Salzberg 2009). These heuristic based aligners allow fast sequence alignment and generate a consensus sequence from the alignment by searching the overlapping portions of the reads and merging them into longer reads in order to construct region of interest i.e. genes or whole genome. The main aim of this step is to generate a consensus sequence from the millions of reads, based on nucleotide base frequency in all mapped reads at each position. A consensus sequence shows the genetic make-up at the time of sample collection. This step marks the completion of sequence generation for part or whole genome. The following steps are important for in depth analysis beyond generation of only a single consensus sequence.

vi. **Variant identification**: Next generation sequencing is not only time efficient, it can also provide the data for an in-depth sequence analysis. Variant analysis uses the read file to determine the conserved and variable nucleotides at specific positions across the length of entire gene of interest. As this process involves statistical calculations spanning over millions of reads, it is both time and computationally intensive process.

The variations within the genomic sequences such as SNPs (single nucleotide polymorphism), SNV (single nucleotide variants) also known as single position diversity estimation, and indels (insertions and deletions) are detected using software's such as

SAMtools (Li et al. 2009), Genome Analysis Toolkit (GATK) (McKenna et al. 2010) and VarScan (Koboldt et al. 2009, Koboldt et al. 2012). Both SAMtools and GATK uses the bayesian probabilistic approach to identify true variants from alignment errors, whereas VarScan uses heuristic approach.

vii. **Annotation**: The detected genetic variants are annotated and interpreted based on the published peer-reviewed literature and number of public disease specific databases. The databases such as Human Gene Mutation Database (HGMD) (Cooper, Ball, and Krawczak 1998, Cooper, Stenson, and Chuzhanova 2006, Krawczak et al. 2000), CliVar (Landrum et al. 2016), and locus-specific databases (LSDBs) (Fokkema, den Dunnen, and Taschner 2005, Fokkema et al. 2011, Lanthaler et al. 2014) interpret genetic variants associated with different diseases. COSMIC, Catalogue of somatic mutations in cancer is a curated database encompassing somatic variants identified in human cancers (Forbes et al. 2011, Sachdev et al. 2013).

viii. **Interpretation of variants**: Lastly, medical professionals will interpret these variants and obtain the patient's clinical history in order to establish a link between the genetic variant and the disease to make an accurate diagnosis. This includes examining different disease pathways and gene network analysis and identifying actual mutations causing the disease.

1.4.2 Applications of NGS in clinical practice

The NGS technologies have several applications in various research fields to solve diverse range of biological problems. Applications of NGS include whole-genome sequencing, gene expression, transcriptome profiling, and epigenetics. NGS has enabled the researches to sequence large segments of the genome (i.e. whole-genome sequencing) and provide insights into identifying and understanding the genetic variants such as single nucleotide variants (SNVs), single nucleotide polymorphisms (SNPs), insertions and deletions of DNA, re-arrangements such as translocation and inversions associated with complex diseases such as cancers for further targeted studies (Peng et al. 2015).

Recently, RNA sequencing (RNASeq) approach is used to uncover genome wide transcriptome characterization and profiling (Koh, Park, et al. 2015). Analysis involving genome wide gene expression (i.e. gene transcription, post-transitional modifications and translation), and molecular pathway analysis provide a deeper understanding of gene regulation in neurological, immunological and other complex diseases. Other applications include studying heritable changes in gene regulation that occur without a change in the DNA sequence. Epigenetics play a significant role in growth, development, and disease progression. The studies on epigenetic changes in cancer provide insight into important tumorigenic pathways (Kinsella et al. 2011, McPherson et al. 2011).

1.5 Phylogenetics

Darwin‘s theory of evolution has been the most widely-accepted model of how set of different species have evolved from a common ancestor. In biology, the study of evolutionary history and relationship among organisms is called phylogenetics (originates from Greek word phyle meaning tribe or race and genesis meaning origin or birth).

Next generation sequencing provides the sequencing depth and coverage for identifying genetic variants and generating distinct clusters of evolutionary related sequences. Phylogenetics can be used to compare two or more sequences, analyse gene families including functional predictions and estimate evolutionary relationships among different organisms or species based on inferences about patterns of characteristics inherited from common ancestor (Lemey et al. 2009). Constructing phylogenetic tree is a two-step process (1) sequence alignment (2) building a tree (Figure 1.5).

1.5.1 Sequence alignment

The first step in constructing a phylogenetic tree involves performing an alignment on set of DNA or protein sequences using multiple sequence alignment programs such as ClustaOmega, T-Coffee, MAFT or MUSCLE. Sequence alignment is the most critical step in any phylogenetic analysis. The outcome of the analysis depends entirely on whether the sequences were accurately aligned.

1.5.2 Building a tree

These alignments are then employed by different phylogeny reconstruction programs to construct phylogenetic trees using different tree building methods. These methods are divided into two major categories: phenetic methods based on genetic distances and cladistic methods based on characters (Figure 1.5).

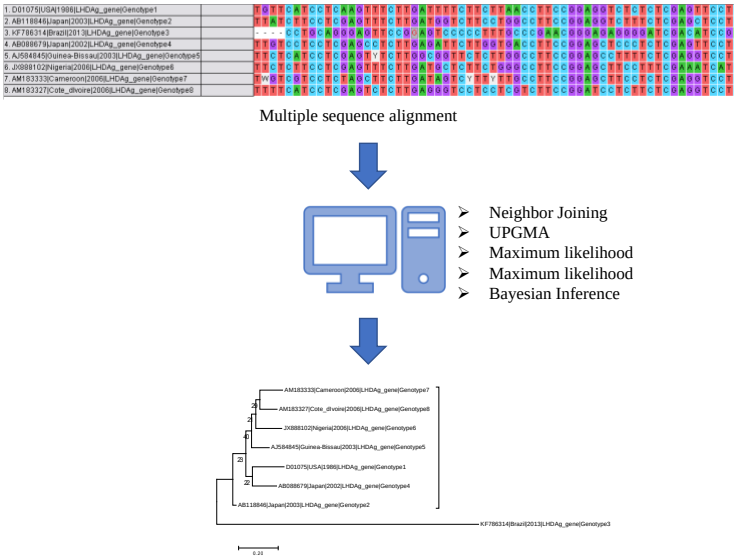

Figure 1.5: *Process in constructing a phylogenetic tree. Step 1 create a multiple sequence alignment, Step 2 select a phylogenetic substitution model, Step 3 tree building and tree visualization.*

1.5.3 Distance-based Methods

Neighbor Joining and unweighted pair group method (UPGMA) are one of the most common distance-based methods. These methods construct the evolutionary tree based on only pre-determined value for observed genetic distance value between the sequences being classified.

1.5.4 Character-based Methods

Character-based method assume that each individual substitutions among the sequences is independent of its neighbors. The most commonly used character-based methods are Maximum parsimony (minimum evolution) Maximum likelihood.

Maximum parsimony searches for the tree with the minimum number of evolutionary changes required to explain (tree) the differences observed in the sequences. It is used to construct phylogenetic trees for similar set of sequences with small amount of variations. Whereas, Maximum likelihood build trees based on mathematical model. These methods find the tree based on calculating probability that best account for large amount of variations in the sequence. Some of the phylogenetic programs using both distance-based and character-based methods include PHYLIP (Felsenstein 1989), Paup (Wilgenbusch and Swofford 2003), Puzzle (Schmidt et al. 2002), and MEGA (Kumar et al. 2008).

1.5.5 Bayesian inference

Bayesian inference refers to a probabilistic method based on Bayes Theorem developed by Reverend Thomas. The Bayesian inferences produce phylogenetic trees using similar approach to the maximum likelihood method. It assumes a prior probability distribution of all the possible trees to generate a phylogenetic tree. Bayesian methods uses Markov chain Monte Carlo (MCMC) algorithms. Some of the tools using Bayesian inference to construct phylogenies include MrBayes (Ronquist et al. 2012), PhyloWGS (Deshwar et al. 2015) and BEAST (Suchard et al. 2018).

1.5.6 Nucleotide substitution models

The use of maximum likelihood algorithm in developing phylogenetic hypotheses requires a model of evolution. Among the most widely used nucleotide substitution models with different combinations of parameters for DNA site substitutions include, Jukes-Cantor (JC): equal base frequencies, all substitutions equally likely (T.H and C.R 1969). Felsenstein 1981 (F81): variable base frequencies, all substitutions equally likely (Felsenstein 1981). Kimura 2-parameter (K80): equal base frequencies, one transition rate and one transversion (Kimura 1980). Hasegawa-Kishino-Yano (HKY): variable base frequencies, one transition rate and one transversion rate (Hasegawa, Kishino, and Yano 1985). Tamura-Nei (TrN): variable base frequencies, equal transversion rates, variable transition rates (Tamura and Nei 1993). Kimura 3-parameter (K3P): variable base frequencies, equal transition rates, two transversion rates (Kimura 1981).

1.5.7 Statistical assessment

Once the phylogenetic tree is constructed. Bootstrapping is a well-known and most frequently used method for estimating the reliability of statistical methods. During the bootstrap method, the authenticity of the resulting tree topology is confirmed by creating a new sequence alignment by random sampling and building a tree based on the new alignment. This process is repeated several times (usually 1000 pseudoreplicates) to generate a probability for particular branch point that occurred in all the trees built (Felsenstein 1985). Higher the bootstrap value (90-100) the more statistically significant are the branch points.

1.5.8 Applications of Phylogenetics

Phylogenetics is widely used to compare genetic sequence of different organisms (plants, animals, bacteria, and viruses) and identify new clusters and estimate their evolutionary relationships. Several phylogenetic methods have been used to study rapidly evolving populations such as viruses causing various infectious diseases. Phylogenetic analysis of viral population also estimates the mutation rate to understand the viral evolution which

might lead to the identification of new species or group referred as quasispecies. The International Committee on the Taxonomy of Viruses (ICTV) uses phylogenetic analysis for the taxonomic classification of various viruses.

1.6 Motivation

Among the viruses, the two viruses i.e., Hepatitis D virus and Hepatitis B virus which belong to group V and VII, are the main focus of this thesis. The first part of thesis is dedicated to Hepatitis B virus with focus on deep sequencing analysis of HBV S region in HBeAg positive chronic patients under REP 2139 antiviral therapy. In this study we addressed the following aims:

I- Characterization of HBV S protein (376bp) quasispecies in patients from different phases of treatment to better understand its outcome.

II- Investigate possible sequence differences among the quasispecies between the Responder and Nonresponder patients, which could help understand mechanism of the drug action.

III- Investigate to what extent the accumulation of mutations in HBsAg region may be responsible for resistance. We need to investigate whether the patients who are low-responder have quasispecies population, as in patients having an early reduction of viremia and HBsAg.

IV- Explore and identify the mutations within the “a” determinant region which results in reduced detection of HBsAg by standard diagnostic assays.

The second part of thesis is dedicated to Hepatitis D virus, revisiting the taxonomical classification of Hepatitis D based on 116 HDV sequences from our cohort together with publicly available data from the GenBank. In particular, we focused on the subtyping of European HDV genotype-1 isolates based on new approach which employs non-redundant HDV sequences to:

I. Identify new subtypes within HDV genotype 1 and other HDV genotypes (2-8)

II. Classify these subtypes with unique evolutionary distances and distinct geographical distribution.

III. Estimate the genetic heterogeneity based on nucleotide differences across genotypes and subtypes.

IV. Identify amino acid specific signatures for the proposed subtypes.

The third part of thesis is dedicated to development of a novel database, Hepatitis Delta Virus database (HDVdb: (http://hdvdb.bio.wzw.tum.de/). The aim of the study was to:

I. Present all the publicly available Hepatitis D virus sequences under one platform i.e. Hepatitis Delta Virus Database (HDVdb).

II. View and retrieve HDV sequences i.e. HDV complete genome, L-HDAg, S-HDAg and partial sequences.
III. Identify genotypes of HDV
IV. Provides sequence analysis services such as building a phylogenetic tree, predicting primers and constructing multiple sequence alignments.
V. Access to external research tools and software's to facilitate sequence analysis.

2 Methods and Materials

The methodology section is divided into three major parts. The first part focuses on the techniques and approaches involved to process and analyse deep sequencing data from chronic HBV patients under REP 2139 therapy using Next generation sequencing technology, whereas, the second part focuses on methods involved in performing phylogenetic analyses to understand evolution of HDV genes. Lastly, the third part focuses on the methods involved in the creation and development of the database for Hepatitis D Virus.

2.1 Longitudinal Evolution of HBV quasispecies

2.1.1 HBV samples

Twelve chronically infected HBV patients with different clinical diagnosis were enrolled in REP 102 study (REP 2139-Ca therapy) at Farabi General Hospital (Dhaka, Bangladesh). For each patient, three samples were withdrawn over the course of treatment with interim times between 35 and 447 days. In addition to these 12 patients, 3 control patients were also included in the study.

2.1.2 Read Statistics

The Illumina sequencing technology generated raw read files in FASTQ format for each sample. A FASTQ file is text-based file contain biological data (nucleotides) along with quality scores. Each sample corresponds to two read files with -R1.fastq.gz and -R2.fastq.gz extensions suggesting paired-end sequencing run. As for the reference sequence, a nucleotide sequence coding only small surface Hepatitis B Antigen protein (157-837bp in length) with accession number NC_003977 was selected for the analysis. The reference sequence NC_003977 belongs to Hepatitis B strain ayw.

2.1.3 Filtering and Trimming

For each sample, quality of the raw sequence reads were analysed using FastQC (version 0.10.1). Upon visualization, we observed a slight quality decline towards the end of the reads, illustrating the need to perform quality filtering and trimming. We, furthermore, filtered all the reads with possible contamination of adapter or primer sequences using Trim_Galore. Trim_Galore uses Cutadapt to remove adapters and performs quality clipping from either the 5’ or 3’ end of the reads with a PHRED threshold of 20. The use of Q20 filtering was addressed to recover sequencing errors and eliminate all false-positive variants generated by PCR errors during the sequencing process so that these sequence reads could be used to detect viral variants in low abundance with great level of confidence.

2.1.4 Read Mapping

After filtration, the sequence reads were mapped to a reference sequence NC_003977 using a read alignment tool known as Burrows-Wheeler Aligner (BWA). At first the reference sequence has to be converted into a format usable by the BWA. For this purpose, bwa provides the tool BwaIndex. In the next step every paired read archive is aligned to the given reference with the Bwa tool Bwa Aligner. The output of the program is a sam-file for every paired read archive that can then be further examined.

2.1.5 Coverage

To get an overview of the distribution of the aligned paired reads, we decided to plot the number of aligned reads per reference position. Therefore, we used the mpileup command of the samtools program version 0.1.18. Beforehand, the sam files created by Bwa had to be sorted and indexed into bam file using Picard software. Once the bam files were created, we checked the mapping information for each of the sample using flagstat command of samtools.

2.1.6 Quasispecies Reconstruction

The reconstruction of viral HBV quasispecies performed with the quasispecies assembly tool QuasiRecomb 1.2 (Topfer et al. 2013). QuasiRecomb implements a hidden markov model (HMM) to infer a viral quasispecies from deep-coverage NGS data using an expectation maximization (EM) algorithm for maximum a posteriori (MAP) parameter estimation. QuasiRecomb requires mapped read bam files as an input. It also implements many option commands allowing flexible analysis. In our study, the flag –conservative was not employed because our interest was to consider minor variants. The region of interest was specified using the -r command (-r 227-602). To reflect the nature of quasispecies variants in the major hydrophilic region (MHR) as accurately as possible in relation to clinical diagnosis; we analysed the viral variants with frequencies as low as 1% in a population.

2.1.6.1 Genetic analysis of HBsAg quasispecies diversity

Quasispecies sequence analysis was performed by estimating quasispecies diversity and complexity. Shannon entropy was used to explain quasispecies complexity by calculating the distribution of different sequences occurring within a population at a particular time point using the following formula(Eriksson et al. 2008):

$$Sn = \sum i \varepsilon_{[A,T,C,G,-]} (p_i \ln p_i) / \ln N$$

Where *pi* is the frequency of nucleotide or deletion (ATGC or -) at a particular position *i, and N* corresponds to number of species analysed in a given population

Moreover, genetic distances (d), ratio of nonsynonymous substitution (dN) to synonymous mutation (dS) were calculated to analyse quasispecies complexity. Distances between pairs

of nucleotide sequences and protein sequences were calculated by using the DNADIST and PRODIST modules in the PHYLIP package version 3.572(Felsenstein 1989). The proportion of nonsynonymous substitution per nonsynonymous site (dN) and synonymous substitution per synonymous site (dS) were computed using Nei-Gojobori method.

2.2 Taxonomic classification of HDV

2.2.1 Dataset

In this study, we used a novel approach of classifying HDV sequences by employing redundancy reduction method such as CD-HIT (Li and Godzik 2006). The data set comprised of sequences from our experimental cohort (i.e. 116 L-HDAg nucleotide sequences) together with the HDV sequences extracted from public repository GenBank (621 nucleotide sequences), were reduced to representative sequences (98 nucleotide sequences) by removing highly identical sequences.

2.2.2 Phylogenetic and Nucleotide sequence analysis

Phylogenetic analysis was performed using Neighbor Joining (NJ) and Maximum Likelihood (ML) methods with Tamura-Nei nucleotide substitution model (Tamura, Nei, and Kumar 2004). The authenticity of the resulting tree topology was confirmed by performing bootstrap method (at 1000 psuedoreplicates) which generated the bootstrap probability for each interior branch in the tree (Felsenstein 1985). A bootstrap value >70% for internal branch node was considered significant. To validate our taxonomic classification, we performed phylogenetic analysis based on Bayesian Inference approach (Mr Bayes) (Ronquist et al. 2012).

To estimate nucleotide diversity across different genotypes and subtypes pairwise genetic distances between the sequences were calculated using Kimura 2 parameter method. Based on the pairwise differences, inter-genotype and inter-subtype differences were estimated.

2.2.3 Specificity determining positions

Amino acid variations specific to HDV genotypes 1, 2 and 4 were predicted using SDPpred method (Kalinina et al. 2004). SDPpred takes multiple sequence alignment of protein sequences and uses mutual information (MI) to calculate the differences between the amino acid distribution at the given multiple sequence alignment.
MI score is given by:

$$MI_p = \sum_{i=1}^{N} \sum_{\alpha=1}^{20} f_p(\alpha, i) \log \frac{f_p(\alpha, i)}{f_p(\alpha) f(i)}$$

Where $f_p(\alpha, i)$ correponds to frequency of an amino acid ***α*** in group ***i*** at position ***p***, $f_p(\alpha)$ is the frequency of an amino acid ***α*** at position ***p*** in the entire alignment, and ***f (i)*** is the fraction of proteins belonging to group ***i.***

Following MI calculation, Z-scores are estimated for each position in the MSA.

$$Z_i^I = \frac{I_{i-(I_i^{exp})}}{\sigma\left(I_i^{exp}\right)}$$

SDPpred sets a threshold using Bernoulli estimator to determine if the Z-score calculated for every position in the multiple sequence alignment is significantly high to indicate a SDP. A high value of Z score indicates that the specific position is most likely to be a SDP.

2.3 HDV database

To facilitate the researchers with the analysis of HDV evolution and variability. We have established a database comprising of all the nucleotide and protein sequences available publicly at GenBank. The database contains all partial, S-HDAg, L-HDAg and complete HDV genome sequences.

2.3.1 Data retrieval

The sequences from GenBank were retrieved using the keyword “hepatitis delta virus” and all entries for taxon “Hepatitis delta virus (taxid :12475)” were selected. In total 2621 HDV sequences deposited at GenBank were downloaded which contained both complete genome sequences and subgenomic sequences (S-HDAG, L-HDAg) and partial cds sequences. Only sequences shorter than 90 bp were discarded from the dataset for the database.

Then, we used a parser script written in Java to extract additional sequence information for the downloaded GenBank entries such as Accession number, genotype. strain name, country, and date of isolation. Moreover, a local BLAST was performed to determine the genotype for sequences that lacked this information.

2.3.2 Database interface

The HDVdb has a static and dynamic part. The static part provides user with general information of HDV whereas the dynamic part allows analysis of software’s and user related queries. Database provides several analytical tools and software’s such as genotype identification, phylogenetic tree reconstruction, multiple sequence alignment generation, primer design prediction. The user can interactively analyse and visualize the phylogenetic trees. In addition, external links to several other database and software’s are also provided.

2.3.3 Database structure

The database was created and runs on PHP Laravel framework. It is hosted using Apache HTTP server. As part of the project, update scripts written in JAVA were used to perform automatic updates of the database. We primarily used Bash scripting to handle these scripts and run HDVdb creation workflow. Also, the database provides a search option for user to search sequences based on accession number, genotype, date, and type of sequence (i.e. protein, nucleotide, complete genome, and partial sequence).

3 Summary of Article 1

Title: Kinetics of hepatitis B surface antigen quasispecies during REP 2139-Ca therapy in HBeAg positive chronic HBV infection

Hepatitis B is a global health problem affecting millions of individuals each year. Although effective vaccines have been available since the 1980s. Over past decade, efforts have been made to treat hepatitis B by developing several new antiviral drugs. However, still there are no effective antiviral drugs available that could cure both acute and chronic HBV infection. Recently, REP2139-Ca which is nucleoside analogue polymer (NAP) have been under clinical trial to treat chronically infected HBV patients.

In this publication, we wanted to contribute to the ongoing research to understand the mechanisms involved in HBV diversity and shed light on dynamics of HBV quasispecies in chronically infected HBV patients under REP2139-Ca treatment. A longitudinal study is effective way to investigate and trace mutational patterns occurring within HBV genome in a patient over different lengths of time. Temporally spaced samples from HBeAg+ chronically infected patients obtained from the Farabi General Hospital (Dhaka, Bangladesh) were analysed to estimate the efficacy of REP 2139-Ca in these patients against HBV.

By means of next generation sequencing (NGS) and the acquisition of viral samples from chronic HBV patients present a great opportunity to understand selective pressures acting on the Hepatitis B virus genome during the infection that influence viral diversity. Our study was conducted to obtain insight into the evolution of immune escape mutations within the major hydrophilic region (MHR) of the “pre-S/S” open reading frame including the “a” determinant in viral populations during the period and after removal of antiviral REP 2139 therapy (Usman et al. 2019).

A very high degree of sequence coverage for a target region of interest achieved by NGS makes it possible to identify several viral variants that make up only a small fraction of the entire virus population. Thus, we are able to trace back the genetic mutations occurring within the viral population, providing a distinct explanation to quasispecies dynamics and the fitness associated to different variants in response to environmental treatment.

In this study, we estimate the genetic variation (i.e. quasispecies complexity and genetic diversity) of the HBsAg in the viral populations of chronically infected patients to evaluate the long term behaviour of parenterally administered REP 2139. We were able to identify several mutations within the MHR of HbSAg in both responder and non-responders. However, no accumulation of mutations were observed within the “a” determinant region which may interfere with the detection of HBsAg by diagnostic assays. Additionally,

quasispecies dynamics among the responders suggest selective pressure on the viral populations of HBsAg within the MHR.

Authors Contribution

Zainab Usman designed the NGS workflow, performed the bioinformatics and statistical analyses. Zainab Usman analysed the data and interpreted the results. Dmitrij Frishman and Michael Roggendorf supervised the project. Zainab Usman and Michael Roggendorf wrote the manuscript. All author contributed to final version of the manuscript.

.

4 Summary of Article 2

Title: Genetic diversity of hepatitis D virus genotype-1 in Europe allows classification into subtypes

Hepatitis D is a rare form of viral hepatitis that causes liver infection in individuals already infected with HBV. Over the past few years, the discovery of new viral sequences presents an opportunity to revisit the taxonomy of Hepatitis Delta Virus. In this publication, we re-evaluate the sequence classification of HDV by employing sequences from our experimental cohort (116 L-HDAg and 13 complete genome sequences) together with publicly available sequences (i.e. 621 L-HDAg) from GenBank NCBI into existing known eight genotypes (HDV-1 to -8).

In this study, we performed phylogenetic analyses using Neighbor joining, Maximum likelihood and Bayesian approach and identified five new subtypes within the HDV-1 genotype and 2 new subtypes for both HDV-2 and HDV-4. Furthermore, we performed pairwise sequence comparisons between HDV complete genome, Large Hepatitis delta Antigen protein (L-HDAg) and Small Hepatitis Delta Antigen protein (S-HDAg) and characterized the sequence differences within and across the known HDV genotypes and subtypes. We estimated that the genotypes exhibit an inter-genotypic difference ≥10%, whereas, the subtypes within a known genotype show an inter-subtype difference ≥3% to <10% (Karimzadeh et al. 2019b) .

Additionally, we identified sequence variations in the nucleotide sequences of L-HDAg for different genotypes and subtypes. These sequence variations include site-specific amino acid preferences associated with genotypes and subtypes distributed across different geographic regions. These specificity determining positions appear within the functional domains of L-HDAg which might be helpful in understanding the adaptation processes to different hosts, genotype or subtype-related pathogenesis, clinical symptoms, response to antiviral therapy, replication capacity and transmission efficiency.

Authors Contribution

Zainab Usman designed and created the computational framework of the study. Zainab Usman extracted the raw sequence data, performed bioinformatics analysis and analysed the data. Dmitrij Frishman and Michael Roggendorf supervised the project. Zainab Usman and Michael Roggendorf interpreted the results and wrote the manuscript. All authors contributed to the final manuscript.

5 Summary of Article 3

Title: HDVdb: a comprehensive hepatitis D virus database

In the past few years, massive amount of HDV sequences have been deposited in publicly available resources such as GenBank, DDBJ, EMBL, Uniprot etc. This presents a growing need to develop an exclusive webserver for Hepatitis delta virus, enabling the researchers to view, retrieve as well as perform various sequence analysis tasks much more efficiently and conveniently.

In this publication, we have developed a comprehensive Hepatitis delta virus database HDVdb Webserver (http://hdvdb.bio.wzw.tum.de/hdvdb/) hosted on the server of Department of Bioinformatics, Technical University Munich. The database comprises of all complete genome, L-HDAg, S-HDAg and partial sequences available at GenBank, together with 116 L-HDAg and 13 complete genome sequences from our previous study (Karimzadeh et al. 2019b). The user can easily download the data of interest from the database. The database also provides a range of analytical tools for sequence analysis, such as identifying HDV genotypes, predicting primers, constructing multiple sequence alignments and phylogenetic trees. In addition, links to several external tools and resources are also available (Usman et al. 2020).

The main objective of HDVdb is to provide a service to researchers to facilitate new discoveries about hepatitis D virus virulence, host range and pathogenesis. This could eventually help to develop effective anti-retroviral vaccines and new therapeutic approaches to treat chronic HDV infections.

Authors Contribution

Zainab Usman retrieved the sequence data from the GenBank, created the webservices, and developed the database. Zainab Usman and Stoyan Velkov implemented software services. Zainab Usman performed database updates. Zainab Usman and Hadi Karimzadeh wrote the manuscript. All authors contributed to the final version of the manuscript.

6 General Discussion

This discussion covers all the three main dissertation topics including a review of relevant literature. The first part of the section focuses on Hepatitis B. The second part focuses on the taxonomic classification of Hepatitis D and third part covers the importance of HDV webserver.

Despite the availability of Hepatitis B vaccines and mass immunization schemes, hepatitis B still remains global health problem with more than 291 million people infected around the globe (Polaris Observatory 2018), some of them are at a risk of developing liver cirrhosis and hepatocellular carcinoma (HCC). Over the past few years, several anti-retroviral drugs have been under clinical trials for the treatment of hepatitis B. The main goal of discovering new anti-retroviral drugs is to improve the drug safety, adherence, and its efficacy to cure the disease. The ultimate cure of hepatitis B infection depends on the clearance of Hepatitis B surface antigen (HBsAg) from the blood stream of infected individual.

With the increasing awareness and understanding of chronic hepatitis B, several novel drug therapies have shown improvements in the disease prognosis. Among the most commonly used antiviral agents include immunomodulatory agents (pegylated interferons-alpha) (Micco et al. 2013, Sun et al. 2014) and long-term oral nucleos(t)ide analgoues (NUCs) (Buster et al. 2008, Lau et al. 2005). The exact mechanism by which interferons interfere with the replication mechanism of hepatitis B virus is unclear. However, nucleoside analogues have been known to exert its antiviral effect by inhibiting the HBV polymerase activity which results in decrease in the viral replication (Gish et al. 2013, Zoulim and Locarnini 2009). Currently, five approved NUCs are used in the treatment of chronic HBV infection. These include three nucleos(t)ide analogues (lamivudine, telbivudine and entecavir) and two nucleotide analogues (adefovir and tenofovir) (Yuen and Lai 2011). These agents have shown great progress in reducing disease progression to cirrhosis, liver failure and hepatocellular carcinoma (Terrault et al. 2018, Terrault et al. 2016).

Recently, Nucleic acid polymers (NAP) which are nucleic acid based antiviral compounds developed for the treatment of chronic hepatitis B and HBV/HDV co-infection with REP 2139 in combination with PegIFN have shown promising results in clinical trials (Al-Mahtab, Bazinet, and Vaillant 2016, Bazinet et al. 2017, Vaillant 2016). A significantly reduced levels or complete loss of surface proteins (HBsAg), followed by seroconversion to anti-HBs and decline in HBV DNA levels were observed. Hence, indicating functional control of chronic HBV infection mediated by humoral immune response as a result of reduction in HBsAg. Elevated levels of HBsAg are known to have immunosuppressive effects in HBV infection and elimination of HBsAg from the liver and the circulation re-establishes the host immune response (Aillot et al. 2018, Dembek, Protzer, and Roggendorf 2018, Kondo et al. 2013, Lebosse et al. 2017). Several in-vivo studies have reported that NAPs REP 2139 inhibits

the HBV subviral particle (SVP) assembly and secretion (Noordeen et al. 2015, Quinet et al. 2018), however the molecular mechanisms underlying SVP assembly and release is not clearly understood and the process by which NAP inhibits SVP assembly and secretion is currently being investigated.

The genetic heterogeneity of HBV quasispecies is known to influence the clinical outcomes and prognosis of HBV infection. (Chevaliez, Rodriguez, and Pawlotsky 2012). In the present study, we used NGS technology to examine whether evolutionary patterns of HBV quasispecies within the MHR correlates with the antiviral response to REP 2139 therapy among the responders and non-responders. In order to compare the complexity and diversity of HBV quasispecies during the course of REP 2139, 3 control samples, 3 non-responder samples and 12 responder samples were collected, and various statistical tests were performed. The dynamics of HBV quasispecies in the responders and non-responders, showed that the responders experienced high genetic complexity and diversity as compared to non-responders. This change may suggest REP 2139 exerts selective targeting of viral quasispecies by host immune response during profound declines in HBV DNA, HBsAg levels and detection of anti-HBs.

During the course of HBV infection, mutations causing amino acid change continuously accumulate within the ‘a’-determinant of HBsAg (Cao 2009), altering the antigenicity of HBsAg protein which leads to viral neutralization failure by anti-HBs and may cause the virus to escape the host immune response (Cooreman, Leroux-Roels, and Paulij 2001, Fortuin et al. 1994, Kay and Zoulim 2007). In our study, no mutations were observed among the responders and non-responders that occurred within the ‘a’-determinant region (i.e. 76-123 aa; 148-200 aa) during the REP-2139 therapy which could impair the detection of HBsAg. No discernible differences in HBsAg mutation patterns were observed in responders and non-responders that would suggest direct interaction of REP 2139 with HBsAg proteins. According to recent studies, REP 2139 doesn’t not directly target HBsAg clearance from the blood circulation (Bazinet et al. 2017, Beilstein et al. 2018). However, it is believed that REP 2139 interacts with the cellular proteins of the host involved in the assembly and release of viral particles.

The dynamics of HBV quasispecies and the mutations observed within HBsAg during the course of REP 2139 needs to be explored further by conducting clinical studies on larger population with similar clinical characteristics, disease severity and underlying medical conditions to understand clinical response and establish methods to determine which patients will respond to the treatment. This will help in understanding the underlying mechanism via which NAPs REP2139 interacts and inhibits the release of HBsAg.

The HDV genome is a small RNA molecule (1.6kb) which requires surface antigen protein of HBV (HBsAg) to cause an infection (Rizzetto et al. 1977). Decades after the discovery of HDV in 1970, hepatitis D still remains a major health problem in regions where HBV is most prevalent. To date, the HDV has been classified into eight different genotypes (Le Gal et al. 2006), each genotype associated with distinct geographical distribution (Hughes, Wedemeyer, and Harrison 2011, Wedemeyer and Manns 2010).

Overall, HDV-1 represents a major genotype and is found worldwide, whereas HDV-2 and HDV-4 are found in North Asia (Russia) and East Asia (Taiwan and Japan) (Ivaniushina et al. 2001, Lee et al. 1996, Marino et al. 2019, Sakugawa et al. 1999). HDV-3 is found in South America (Brazil, Venezuela, Peru, Colombia, Argentina and Ecuador) (Casey et al. 1993, di Filippo et al. 2012, Gomes-Gouvea et al. 2009, Quintero et al. 2001, Scarponi et al. 2019), and HDV-5 to HDV-8 clades are prevalent in Africa. (Makuwa et al. 2008, Radjef et al. 2004) However, in a recent study HDV-8 isolates found in state of Maranhão (Brazil) are thought to be as a result of migration process during 16-18 centuries (Santos et al. 2016). In 1990s, the incidence of HDV significantly reduced in European countries as a result of increased HBV vaccination regimen (Gaeta et al. 2000). However, in the last decade there has been increasing number of HDV incidence because of influx of people migrating from HDV and HBV endemic countries (Wedemeyer 2011, Wedemeyer and Manns 2010).

Recently, an increasing number of HDV isolates have been identified and sequenced. In our study, we revisited the scheme of HDV classification by performing HDV nucleotide sequence comparisons and phylogenetic analysis of sequences available at GenBank together with HDV isolates (116 L-HDAg, 13 HDV complete genome) from our European cohort. To our knowledge, for the first time a comprehensive bioinformatics analysis was performed where the sequence dataset was subjected to redundancy reduction using CD-HIT (Li and Godzik 2006). As a result, highly identical sequences were reduced to a set of unique sequences representing the entire dataset.

The phylogenetic analysis was performed using three methods; Neighbor-Joining, Maximum Likelihood (MEGA) (Li and Godzik 2006) and Bayesian Inference (Mr Bayes) (Kumar et al. 2008). These analyses, involving both complete genome sequences and L-HDAg confirmed the known eight genotypes and allowed us to propose five new subtypes for HDV-1, and two subtypes each for HDV-2 and HDV-4 respectively. Recently, several studies have proposed different HDV classification models based on the genetic heterogeneity of HDV and geographical distributions (Delfino et al. 2018, Le Gal et al. 2006, Miao et al. 2018).

Based on the nucleotide sequences analysis we propose an inter-genotype difference (≥10%) and inter-subtype difference (<10%) for genotypes and subtypes respectively. In

support of the new proposed HDV subtypes, amino acid analyses involving identification of specific amino acid signatures confirmed the existence of new subtypes. Sequence analysis of HDV genome reveals highly conserved nucleotide and protein sequences indicating a strong selection pressure and evolutionary constraints on conserving the gene function and the structure of the protein. Despite the high sequence conservation, the delta proteins (S-HDAg and L-HDAg) experience several amino acid substitutions across different genotypes. These amino acid substitutions appear within the functional domains of HDV such as RNA binding domain, coiled coil domain (CCD), nuclear localization signal (NLS), helix-loop helix motif (HLH) and proline/glycine rich region (PGRS) and nuclear export signal (NES) (Alfaiate, Deny, and Durantel 2015, Pascarella and Negro 2011). According to a recent study, amino acid heterogeneity was explored across the HDV proteins, and the presence of amino acid at position 202 in genotype HDV-1 was shown to reflect geographical background (Le Gal et al. 2017). The amino acid “Serine” was found at high concentrations (90%) in isolates belonging to African strains as compared to non-African isolates having ‘Alanine’ (>70%) at position 202. Similarly, amino acid substitution of Proline at position 205 with Alanine or Arginine located within NES of L-HDAg have shown varying efficiencies of HDV virion assembly and secretion across different genotypes (Shih et al. 2010). These amino acids specificities across different genotypes also suggest how HDV virus have evolved and adapted to new host environments.

Several studies have shown that different HDV genotypes exhibit different clinical course of infection and disease outcome. HDV-1 is associated with severe form of hepatitis, whereas genotype HDV-2 and HDV-4 causes mild course of liver infection. HDV genotype 3 causes severe form of fulminant hepatitis (Hughes, Wedemeyer, and Harrison 2011). However, the clinical outcomes of genotype 5-8 are not clearly understood. Thus, we believe that the identification of new HDV subtypes may be important in understanding different virological responses and resistance to different anti-retroviral treatment. In addition, a correlation can be established between various subtypes and specific clinical conditions and outcomes. For instance, several studies indicate that different HBV genotype responds differently to Interferon treatment, HBV genotypes A and B response to IFN treatment is greater than genotype C and D, whereas genotype D exhibits the worst response (Boglione et al. 2013, Moura et al. 2013, Shi 2012). In another study, HBV subtype adw showed significantly higher lamivudine resistance as compared to subtype ayw (Kao 2002). A study focusing on HCV subtype distribution across Germany showed a notable variation in the distribution of subtypes 1a and 1b in different age groups. The subtype 1b was more frequently prevalent in individuals older than 50 years. Over a period of 20 years, subtype 1a gradually substituted the subtype 1b (Ross et al. 2000). Hence, the genetic shifts in subtype

prevalence within a population may help in understanding the future epidemiology and clinical pattern of HDV infection.

In recent years, the advancements in the sequencing technologies have resulted in enormous amount of sequence data deposited in public repositories such as EMBL, DDBJ and GenBank (Tateno 2008). The flood of raw sequence data in these repositories has led to the development of various small specialized databases. These databases facilitate research as they provide researchers the access to all the relevant information and tools under one platform to perform various sequence analysis. Several established databases have been known for known viruses, such as HBV (Hayer et al. 2013), HIV (Kuiken, Korber, and Shafer 2003) , HCV (Kuiken et al. 2008).

Over the past decade, HDV genomic data both nucleotide and protein sequences have increased exponentially. Given this situation, we have developed a comprehensive database HDVdb webserver (http://hdvdb.bio.wzw.tum.de/hdvdb/), for HDV genes, genomes and proteins. The webserver is managed and hosted on a server at Department of Bioinformatics, Technical University Munich. The database offers an efficient retrieval of sequence data together with various user-friendly analytical tools for performing comparative sequence analysis, such as identifying HDV genotype, performing multiple sequence alignments, constructing phylogenetic trees, and predicting primers. HDVdb webserver aims to be a resource for scientists' community working on HDV genomics, pathogenicity and drug design. It allows the researchers to study HDV genetic variability by facilitating retrieval and performing comparative sequence analysis using all the publicly available HDV sequences.

The HDVdb webserver was made public in May 2020. The database contains the HDV sequences by downloading all the submitted GenBank entries along with additional information such as GenBank accession numbers, country, sampling year, genotypes. The entries lacking the genotype details were manually assigned by performing BLAST searches and phylogenetic analysis. In addition, we also provide this opportunity to integrate sequence data from various collaborators worldwide.

In future, we plan to develop a predictor for identifying HDV subtypes. In recent years, several studies have reported new subtypes emerging within the known eight genotypes (Delfino et al. 2018, Le Gal et al. 2017, Miao et al. 2018). The studies by Le Gal and Miao not only confirmed the eight major HDV genotypes but also identified several new subtypes. In accordance with these studies, we have also proposed new subtypes for genotype 1 , 2 and 4 (Karimzadeh et al. 2019a). The identification of new HDV subtypes is a crucial step towards understanding the variations associated with disease prevalence geographically and identifying key factors affecting the treatment response across populations.

7 References

Abeywickrama-Samarakoon, N., J. C. Cortay, C. Sureau, D. Alfaiate, M. Levrero, and P. Deny. 2018. "[Hepatitis delta virus replication and the role of the small hepatitis delta protein S-HDAg]." *Med Sci (Paris)* 34 (10):833-841. doi: 10.1051/medsci/2018209.

Ahn, S., and H. Vikalo. 2018. "aBayesQR: A Bayesian Method for Reconstruction of Viral Populations Characterized by Low Diversity." *J Comput Biol* 25 (7):637-648. doi: 10.1089/cmb.2017.0249.

Aillot, L., M. Bonnin, M. Ait-Goughoulte, N. Bendriss-Vermare, S. Maadadi, L. Dimier, M. Subic, C. Scholtes, I. Najera, F. Zoulim, J. Lucifora, and D. Durantel. 2018. "Interaction between Toll-Like Receptor 9-CpG Oligodeoxynucleotides and Hepatitis B Virus Virions Leads to Entry Inhibition in Hepatocytes and Reduction of Alpha Interferon Production by Plasmacytoid Dendritic Cells." *Antimicrob Agents Chemother* 62 (4). doi: 10.1128/AAC.01741-17.

Al-Mahtab, M., M. Bazinet, and A. Vaillant. 2016. "Safety and Efficacy of Nucleic Acid Polymers in Monotherapy and Combined with Immunotherapy in Treatment-Naive Bangladeshi Patients with HBeAg+ Chronic Hepatitis B Infection." *PLoS One* 11 (6):e0156667. doi: 10.1371/journal.pone.0156667.

Alfaiate, D., P. Deny, and D. Durantel. 2015. "Hepatitis delta virus: From biological and medical aspects to current and investigational therapeutic options." *Antiviral Res* 122:112-29. doi: 10.1016/j.antiviral.2015.08.009.

Ali, A., H. Abdel-Hafiz, M. Suhail, A. Al-Mars, M. K. Zakaria, K. Fatima, S. Ahmad, E. Azhar, A. Chaudhary, and I. Qadri. 2014. "Hepatitis B virus, HBx mutants and their role in hepatocellular carcinoma." *World J Gastroenterol* 20 (30):10238-48. doi: 10.3748/wjg.v20.i30.10238.

Almeida, J. D., D. Rubenstein, and E. J. Stott. 1971. "New antigen-antibody system in Australia-antigen-positive hepatitis." *Lancet* 2 (7736):1225-7. doi: 10.1016/s0140-6736(71)90543-5.

Alvarado-Mora, M. V., S. Locarnini, M. Rizzetto, and J. R. Pinho. 2013. "An update on HDV: virology, pathogenesis and treatment." *Antivir Ther* 18 (3 Pt B):541-8. doi: 10.3851/IMP2598.

Alves, C., N. Freitas, and C. Cunha. 2008. "Characterization of the nuclear localization signal of the hepatitis delta virus antigen." *Virology* 370 (1):12-21. doi: 10.1016/j.virol.2007.07.034.

Andrews, S. 2010. "FastQC: a quality control tool for high throughput sequence data.".

Baltimore, D. 1971. "Expression of animal virus genomes." *Bacteriol Rev* 35 (3):235-41.

Barker, L. F., F. V. Chisari, P. P. McGrath, D. W. Dalgard, R. L. Kirschstein, J. D. Almeida, T. S. Edington, D. G. Sharp, and M. R. Peterson. 1973. "Transmission of type B viral hepatitis to chimpanzees." *J Infect Dis* 127 (6):648-62. doi: 10.1093/infdis/127.6.648.

Barker, L. F., J. E. Maynard, R. H. Purcell, J. H. Hoofnagle, K. R. Berquist, and W. T. London. 1975. "Viral hepatitis, type B, in experimental animals." *Am J Med Sci* 270 (1):189-95.

Bazinet, M., V. Pantea, V. Cebotarescu, L. Cojuhari, P. Jimbei, J. Albrecht, P. Schmid, F. Le Gal, E. Gordien, A. Krawczyk, H. Mijocevic, H. Karimzadeh, M. Roggendorf, and A. Vaillant. 2017. "Safety and efficacy of REP 2139 and pegylated interferon alfa-2a for treatment-naive patients with chronic hepatitis B virus and hepatitis D virus co-infection (REP 301 and REP 301-LTF): a non-randomised, open-label, phase 2 trial." *Lancet Gastroenterol Hepatol* 2 (12):877-889. doi: 10.1016/S2468-1253(17)30288-1.

Beilstein, F., M. Blanchet, A. Vaillant, and C. Sureau. 2018. "Nucleic Acid Polymers Are Active against Hepatitis Delta Virus Infection In Vitro." *J Virol* 92 (4). doi: 10.1128/JVI.01416-17.

Bill, C. A., and J. Summers. 2004. "Genomic DNA double-strand breaks are targets for hepadnaviral DNA integration." *Proc Natl Acad Sci U S A* 101 (30):11135-40. doi: 10.1073/pnas.0403925101.

Blank, A., C. Markert, N. Hohmann, A. Carls, G. Mikus, T. Lehr, A. Alexandrov, M. Haag, M. Schwab, S. Urban, and W. E. Haefeli. 2016. "First-in-human application of the novel hepatitis B and hepatitis D virus entry inhibitor myrcludex B." *J Hepatol* 65 (3):483-9. doi: 10.1016/j.jhep.2016.04.013.

Blank, A., K. Meier, S. Urban, W. E. Haefeli, and J. Weiss. 2018. "Drug-drug interaction potential of the HBV and HDV entry inhibitor myrcludex B assessed in vitro." *Antivir Ther* 23 (3):267-275. doi: 10.3851/IMP3206.

Blondot, M. L., V. Bruss, and M. Kann. 2016. "Intracellular transport and egress of hepatitis B virus." *J Hepatol* 64 (1 Suppl):S49-S59. doi: 10.1016/j.jhep.2016.02.008.

Blumberg, B. S. 1964. "Polymorphisms of the Serum Proteins and the Development of Iso-Precipitins in Transfused Patients." *Bull N Y Acad Med* 40:377-86.

Blumberg, B. S. 1977. "Australia antigen and the biology of hepatitis B." *Science* 197 (4298):17-25. doi: 10.1126/science.325649.

Boglione, L., A. D'Avolio, G. Cariti, M. G. Milia, M. Simiele, A. De Nicolo, V. Ghisetti, and G. Di Perri. 2013. "Sequential therapy with entecavir and PEG-INF in patients affected by chronic hepatitis B and high levels of HBV-DNA with non-D genotypes." *J Viral Hepat* 20 (4):e11-9. doi: 10.1111/jvh.12018.

Bogomolov, P., A. Alexandrov, N. Voronkova, M. Macievich, K. Kokina, M. Petrachenkova, T. Lehr, F. A. Lempp, H. Wedemeyer, M. Haag, M. Schwab, W. E. Haefeli, A. Blank, and S. Urban. 2016. "Treatment of chronic hepatitis D with the entry inhibitor myrcludex B: First results of a phase Ib/IIa study." *J Hepatol* 65 (3):490-8. doi: 10.1016/j.jhep.2016.04.016.

Bolger, A. M., M. Lohse, and B. Usadel. 2014. "Trimmomatic: a flexible trimmer for Illumina sequence data." *Bioinformatics* 30 (15):2114-20. doi: 10.1093/bioinformatics/btu170.

Bond, W. W., M. S. Favero, N. J. Petersen, C. R. Gravelle, J. W. Ebert, and J. E. Maynard. 1981. "Survival of hepatitis B virus after drying and storage for one week." *Lancet* 1 (8219):550-1. doi: 10.1016/s0140-6736(81)92877-4.

Bordier, B. B., P. L. Marion, K. Ohashi, M. A. Kay, H. B. Greenberg, J. L. Casey, and J. S. Glenn. 2002. "A prenylation inhibitor prevents production of infectious hepatitis delta virus particles." *J Virol* 76 (20):10465-72. doi: 10.1128/jvi.76.20.10465-10472.2002.

Bordier, B. B., J. Ohkanda, P. Liu, S. Y. Lee, F. H. Salazar, P. L. Marion, K. Ohashi, L. Meuse, M. A. Kay, J. L. Casey, S. M. Sebti, A. D. Hamilton, and J. S. Glenn. 2003. "In vivo antiviral efficacy of prenylation inhibitors against hepatitis delta virus." *J Clin Invest* 112 (3):407-14. doi: 10.1172/JCI17704.

Buster, E. H., H. J. Flink, Y. Cakaloglu, K. Simon, J. Trojan, F. Tabak, T. M. So, S. V. Feinman, T. Mach, U. S. Akarca, M. Schutten, W. Tielemans, A. J. van Vuuren, B. E. Hansen, and H. L. Janssen. 2008. "Sustained HBeAg and HBsAg loss after long-term follow-up of HBeAg-positive patients treated with peginterferon alpha-2b." *Gastroenterology* 135 (2):459-67. doi: 10.1053/j.gastro.2008.05.031.

Cao, G. W. 2009. "Clinical relevance and public health significance of hepatitis B virus genomic variations." *World J Gastroenterol* 15 (46):5761-9. doi: 10.3748/wjg.15.5761.

Carlson, C. J., C. M. Zipfel, R. Garnier, and S. Bansal. 2019. "Global estimates of mammalian viral diversity accounting for host sharing." *Nat Ecol Evol* 3 (7):1070-1075. doi: 10.1038/s41559-019-0910-6.

Casey, J. L. 2012. "Control of ADAR1 editing of hepatitis delta virus RNAs." *Curr Top Microbiol Immunol* 353:123-43. doi: 10.1007/82_2011_146.

Casey, J. L., T. L. Brown, E. J. Colan, F. S. Wignall, and J. L. Gerin. 1993. "A genotype of hepatitis D virus that occurs in northern South America." *Proc Natl Acad Sci U S A* 90 (19):9016-20.

Casey, J. L., and J. L. Gerin. 2006. "The woodchuck model of HDV infection." *Curr Top Microbiol Immunol* 307:211-25.

Castelnau, C., F. Le Gal, M. P. Ripault, E. Gordien, M. Martinot-Peignoux, N. Boyer, B. N. Pham, S. Maylin, P. Bedossa, P. Deny, P. Marcellin, and E. Gault. 2006. "Efficacy of peginterferon alpha-2b in chronic hepatitis delta: relevance of quantitative RT-PCR for follow-up." *Hepatology* 44 (3):728-35. doi: 10.1002/hep.21325.

Chang, F. L., P. J. Chen, S. J. Tu, C. J. Wang, and D. S. Chen. 1991. "The large form of hepatitis delta antigen is crucial for assembly of hepatitis delta virus." *Proc Natl Acad Sci U S A* 88 (19):8490-4.

Chang, W. -S , J. H. Pettersson, C. Le Lay, M. Shi, N. Lo, M. Wille, J. -S. Eden, and E. Holmes. 2019. "Novel hepatitis D-like agents in vertebrates and invertebrates." *BioRxiv Preprint.* doi: 10.1101/539924.

Chao, Y. C., M. F. Chang, I. Gust, and M. M. Lai. 1990. "Sequence conservation and divergence of hepatitis delta virus RNA." *Virology* 178 (2):384-92.

Chen, H. Y., D. T. Shen, D. Z. Ji, P. C. Han, W. M. Zhang, J. F. Ma, W. S. Chen, H. Goyal, S. Pan, and H. G. Xu. 2018. "Prevalence and burden of hepatitis D virus infection in the global population: a systematic review and meta-analysis." *Gut*. doi: 10.1136/gutjnl-2018-316601.

Cheng, X., Y. Xia, E. Serti, P. D. Block, M. Chung, K. Chayama, B. Rehermann, and T. J. Liang. 2017. "Hepatitis B virus evades innate immunity of hepatocytes but activates cytokine production by macrophages." *Hepatology* 66 (6):1779-1793. doi: 10.1002/hep.29348.

Chevaliez, S., C. Rodriguez, and J. M. Pawlotsky. 2012. "New virologic tools for management of chronic hepatitis B and C." *Gastroenterology* 142 (6):1303-1313 e1. doi: 10.1053/j.gastro.2012.02.027.

Comas, I., A. Moya, and F. Gonzalez-Candelas. 2005. "Validating viral quasispecies with digital organisms: a re-examination of the critical mutation rate." *BMC Evol Biol* 5:5. doi: 10.1186/1471-2148-5-5.

Cooksley, W. G. 2005. "Peginterferon-alpha 2a for the treatment of hepatitis B infection." *Expert Opin Pharmacother* 6 (8):1373-80. doi: 10.1517/14656566.6.8.1373.

Cooper, D. N., E. V. Ball, and M. Krawczak. 1998. "The human gene mutation database." *Nucleic Acids Res* 26 (1):285-7. doi: 10.1093/nar/26.1.285.

Cooper, D. N., P. D. Stenson, and N. A. Chuzhanova. 2006. "The Human Gene Mutation Database (HGMD) and its exploitation in the study of mutational mechanisms." *Curr Protoc Bioinformatics* Chapter 1:Unit 1 13. doi: 10.1002/0471250953.bi0113s12.

Cooreman, M. P., G. Leroux-Roels, and W. P. Paulij. 2001. "Vaccine- and hepatitis B immune globulin-induced escape mutations of hepatitis B virus surface antigen." *J Biomed Sci* 8 (3):237-47. doi: 10.1007/BF02256597.

Cova, L., C. P. Wild, R. Mehrotra, V. Turusov, T. Shirai, V. Lambert, C. Jacquet, L. Tomatis, C. Trepo, and R. Montesano. 1990. "Contribution of aflatoxin B1 and hepatitis B virus infection in the induction of liver tumors in ducks." *Cancer Res* 50 (7):2156-63.

Craxi, A., and W. G. Cooksley. 2003. "Pegylated interferons for chronic hepatitis B." *Antiviral Res* 60 (2):87-9.

Cui, X., D. N. Clark, K. Liu, X. D. Xu, J. T. Guo, and J. Hu. 2016. "Viral DNA-Dependent Induction of Innate Immune Response to Hepatitis B Virus in Immortalized Mouse Hepatocytes." *J Virol* 90 (1):486-96. doi: 10.1128/JVI.01263-15.

Dane, D. S., C. H. Cameron, and M. Briggs. 1970. "Virus-like particles in serum of patients with Australia-antigen-associated hepatitis." *Lancet* 1 (7649):695-8. doi: 10.1016/s0140-6736(70)90926-8.

de la Torre, J. C., and J. J. Holland. 1990. "RNA virus quasispecies populations can suppress vastly superior mutant progeny." *J Virol* 64 (12):6278-81.

Degertekin, H., K. Yalcin, and M. Yakut. 2006. "The prevalence of hepatitis delta virus infection in acute and chronic liver diseases in Turkey: an analysis of clinical studies." *Turk J Gastroenterol* 17 (1):25-34.

Delfino, C. M., C. S. Cerrudo, M. Biglione, J. R. Oubina, P. D. Ghiringhelli, and V. L. Mathet. 2018. "A comprehensive bioinformatic analysis of hepatitis D virus full-length genomes." *J Viral Hepat* 25 (7):860-869. doi: 10.1111/jvh.12876.

Delius, H., N. M. Gough, C. H. Cameron, and K. Murray. 1983. "Structure of the hepatitis B virus genome." *J Virol* 47 (2):337-43.
Dembek, C., U. Protzer, and M. Roggendorf. 2018. "Overcoming immune tolerance in chronic hepatitis B by therapeutic vaccination." *Curr Opin Virol* 30:58-67. doi: 10.1016/j.coviro.2018.04.003.
Deny, P. 2006. "Hepatitis delta virus genetic variability: from genotypes I, II, III to eight major clades?" *Curr Top Microbiol Immunol* 307:151-71.
Deshwar, A. G., S. Vembu, C. K. Yung, G. H. Jang, L. Stein, and Q. Morris. 2015. "PhyloWGS: reconstructing subclonal composition and evolution from whole-genome sequencing of tumors." *Genome Biol* 16:35. doi: 10.1186/s13059-015-0602-8.
di Filippo, D., F. Cortes-Mancera, M. Beltran, M. P. Arbelaez, S. Jaramillo, J. C. Restrepo, G. Correa, and M. C. Navas. 2012. "Molecular characterization of hepatitis c virus in multi-transfused Colombian patients." *Virol J* 9:242. doi: 10.1186/1743-422X-9-242.
Dodt, M., J. T. Roehr, R. Ahmed, and C. Dieterich. 2012. "FLEXBAR-Flexible Barcode and Adapter Processing for Next-Generation Sequencing Platforms." *Biology (Basel)* 1 (3):895-905. doi: 10.3390/biology1030895.
Domingo, E., and J. Gomez. 2007. "Quasispecies and its impact on viral hepatitis." *Virus Res* 127 (2):131-50. doi: 10.1016/j.virusres.2007.02.001.
Domingo, E., and J. J. Holland. 1997. "RNA virus mutations and fitness for survival." *Annu Rev Microbiol* 51:151-78. doi: 10.1146/annurev.micro.51.1.151.
Domingo, E., J. Sheldon, and C. Perales. 2012. "Viral quasispecies evolution." *Microbiol Mol Biol Rev* 76 (2):159-216. doi: 10.1128/MMBR.05023-11.
Doo, E. C., and M. G. Ghany. 2010. "Hepatitis B virology for clinicians." *Clin Liver Dis* 14 (3):397-408. doi: 10.1016/j.cld.2010.05.001.
Dressman, D., H. Yan, G. Traverso, K. W. Kinzler, and B. Vogelstein. 2003. "Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations." *Proc Natl Acad Sci U S A* 100 (15):8817-22. doi: 10.1073/pnas.1133470100.
Eigen, M. 1971. "Selforganization of matter and the evolution of biological macromolecules." *Naturwissenschaften* 58 (10):465-523. doi: 10.1007/bf00623322.
Eigen, M., and P. Schuster. 1977. "The hypercycle. A principle of natural self-organization. Part A: Emergence of the hypercycle." *Naturwissenschaften* 64 (11):541-65. doi: 10.1007/bf00450633.
Erhardt, A., W. Gerlich, C. Starke, U. Wend, A. Donner, A. Sagir, T. Heintges, and D. Haussinger. 2006. "Treatment of chronic hepatitis delta with pegylated interferon-alpha2b." *Liver Int* 26 (7):805-10. doi: 10.1111/j.1478-3231.2006.01279.x.
Eriksson, N., L. Pachter, Y. Mitsuya, S. Y. Rhee, C. Wang, B. Gharizadeh, M. Ronaghi, R. W. Shafer, and N. Beerenwinkel. 2008. "Viral population estimation using pyrosequencing." *PLoS Comput Biol* 4 (4):e1000074. doi: 10.1371/journal.pcbi.1000074.
Ezzikouri, S., M. Ozawa, M. Kohara, N. Elmdaghri, S. Benjelloun, and K. Tsukiyama-Kohara. 2014. "Recent insights into hepatitis B virus-host interactions." *J Med Virol* 86 (6):925-32. doi: 10.1002/jmv.23916.
Farci, P., A. Mandas, A. Coiana, M. E. Lai, V. Desmet, P. Van Eyken, Y. Gibo, L. Caruso, S. Scaccabarozzi, D. Criscuolo, and et al. 1994. "Treatment of chronic hepatitis D with interferon alfa-2a." *N Engl J Med* 330 (2):88-94. doi: 10.1056/NEJM199401133300202.
Felsenstein, J. 1989. "PHYLIP -- Phylogeny Inference Package version 3.2." *Cladistics* 5:164-165.
Felsenstein, J. 1981. "Evolutionary trees from DNA sequences: a maximum likelihood approach." *J Mol Evol* 17 (6):368-76. doi: 10.1007/bf01734359.
Felsenstein, J. 1985. "Confidence Limits on Phylogenies: An Approach Using the Bootstrap." *Evolution* 39 (4):783-791. doi: 10.1111/j.1558-5646.1985.tb00420.x.

Fokkema, I. F., J. T. den Dunnen, and P. E. Taschner. 2005. "LOVD: easy creation of a locus-specific sequence variation database using an "LSDB-in-a-box" approach." *Hum Mutat* 26 (2):63-8. doi: 10.1002/humu.20201.

Fokkema, I. F., P. E. Taschner, G. C. Schaafsma, J. Celli, J. F. Laros, and J. T. den Dunnen. 2011. "LOVD v.2.0: the next generation in gene variant databases." *Hum Mutat* 32 (5):557-63. doi: 10.1002/humu.21438.

Forbes, S. A., N. Bindal, S. Bamford, C. Cole, C. Y. Kok, D. Beare, M. Jia, R. Shepherd, K. Leung, A. Menzies, J. W. Teague, P. J. Campbell, M. R. Stratton, and P. A. Futreal. 2011. "COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer." *Nucleic Acids Res* 39 (Database issue):D945-50. doi: 10.1093/nar/gkq929.

Fortuin, M., V. Karthigesu, L. Allison, C. Howard, S. Hoare, M. Mendy, and H. C. Whittle. 1994. "Breakthrough infections and identification of a viral variant in Gambian children immunized with hepatitis B vaccine." *J Infect Dis* 169 (6):1374-6. doi: 10.1093/infdis/169.6.1374.

Gaeta, G. B., T. Stroffolini, M. Chiaramonte, T. Ascione, G. Stornaiuolo, S. Lobello, E. Sagnelli, M. R. Brunetto, and M. Rizzetto. 2000. "Chronic hepatitis D: a vanishing Disease? An Italian multicenter study." *Hepatology* 32 (4 Pt 1):824-7. doi: 10.1053/jhep.2000.17711.

Giersch, K., M. Homs, T. Volz, M. Helbig, L. Allweiss, A. W. Lohse, J. Petersen, M. Buti, T. Pollicino, C. Sureau, M. Dandri, and M. Lutgehetmann. 2017. "Both interferon alpha and lambda can reduce all intrahepatic HDV infection markers in HBV/HDV infected humanized mice." *Sci Rep* 7 (1):3757. doi: 10.1038/s41598-017-03946-9.

Gish, R. G., D. H. Yi, S. Kane, M. Clark, M. Mangahas, S. Baqai, M. A. Winters, J. Proudfoot, and J. S. Glenn. 2013. "Coinfection with hepatitis B and D: epidemiology, prevalence and disease in patients in Northern California." *J Gastroenterol Hepatol* 28 (9):1521-5. doi: 10.1111/jgh.12217.

Gish, R., J. D. Jia, S. Locarnini, and F. Zoulim. 2012. "Selection of chronic hepatitis B therapy with high barrier to resistance." *Lancet Infect Dis* 12 (4):341-53. doi: 10.1016/S1473-3099(11)70314-0.

Gomes-Gouvea, M. S., M. C. Soares, G. Bensabath, I. M. de Carvalho-Mello, E. M. Brito, O. S. Souza, A. T. Queiroz, F. J. Carrilho, and J. R. Pinho. 2009. "Hepatitis B virus and hepatitis delta virus genotypes in outbreaks of fulminant hepatitis (Labrea black fever) in the western Brazilian Amazon region." *J Gen Virol* 90 (Pt 11):2638-43. doi: 10.1099/vir.0.013615-0.

Gripon, P., I. Cannie, and S. Urban. 2005. "Efficient inhibition of hepatitis B virus infection by acylated peptides derived from the large viral surface protein." *J Virol* 79 (3):1613-22. doi: 10.1128/JVI.79.3.1613-1622.2005.

Guillot, C., N. Martel, F. Berby, I. Bordes, O. Hantz, M. Blanchet, C. Sureau, A. Vaillant, and I. Chemin. 2017. "Inhibition of hepatitis B viral entry by nucleic acid polymers in HepaRG cells and primary human hepatocytes." *PLoS One* 12 (6):e0179697. doi: 10.1371/journal.pone.0179697.

Hadziyannis, S. J. 1991. "Use of alpha-interferon in the treatment of chronic delta hepatitis." *J Hepatol* 13 Suppl 1:S21-6.

Harpaz, R., L. Von Seidlein, F. M. Averhoff, M. P. Tormey, S. D. Sinha, K. Kotsopoulou, S. B. Lambert, B. H. Robertson, J. D. Cherry, and C. N. Shapiro. 1996. "Transmission of hepatitis B virus to multiple patients from a surgeon without evidence of inadequate infection control." *N Engl J Med* 334 (9):549-54. doi: 10.1056/NEJM199602293340901.

Hasegawa, M., H. Kishino, and T. Yano. 1985. "Dating of the human-ape splitting by a molecular clock of mitochondrial DNA." *J Mol Evol* 22 (2):160-74. doi: 10.1007/bf02101694.

Hayer, J., F. Jadeau, G. Deleage, A. Kay, F. Zoulim, and C. Combet. 2013. "HBVdb: a knowledge database for Hepatitis B Virus." *Nucleic Acids Res* 41 (Database issue):D566-70. doi: 10.1093/nar/gks1022.

Head, S. R., H. K. Komori, S. A. LaMere, T. Whisenant, F. Van Nieuwerburgh, D. R. Salomon, and P. Ordoukhanian. 2014. "Library construction for next-generation sequencing: overviews and challenges." *Biotechniques* 56 (2):61-4, 66, 68, passim. doi: 10.2144/000114133.

Hetzel, U., L. Szirovicza, T. Smura, B. Prähauser, O. Vapalahti, A. Kipar, and J. Hepojoki. 2018. "Identification of a novel deltavirus in boa constrictor." *BioRxiv Preprint.* doi: 10.1101/429753.

Ho, E., P. Deltenre, M. Nkuize, J. Delwaide, I. Colle, P. Michielsen, and Liver Belgian Association for the Study of the. 2013. "Coinfection of hepatitis B and hepatitis delta virus in Belgium: a multicenter BASL study. Prospective epidemiology and comparison with HBV mono-infection." *J Med Virol* 85 (9):1513-7. doi: 10.1002/jmv.23653.

Hoenen, T. 2016. "Sequencing of Ebola Virus Genomes Using Nanopore Technology." *Bio Protoc* 6 (21). doi: 10.21769/BioProtoc.1998.

Hoenen, T., A. Groseth, K. Rosenke, R. J. Fischer, A. Hoenen, S. D. Judson, C. Martellaro, D. Falzarano, A. Marzi, R. B. Squires, K. R. Wollenberg, E. de Wit, J. Prescott, D. Safronetz, N. van Doremalen, T. Bushmaker, F. Feldmann, K. McNally, F. K. Bolay, B. Fields, T. Sealy, M. Rayfield, S. T. Nichol, K. C. Zoon, M. Massaquoi, V. J. Munster, and H. Feldmann. 2016. "Nanopore Sequencing as a Rapidly Deployable Ebola Outbreak Tool." *Emerg Infect Dis* 22 (2):331-4. doi: 10.3201/eid2202.151796.

Hou, J., Z. Liu, and F. Gu. 2005. "Epidemiology and Prevention of Hepatitis B Virus Infection." *Int J Med Sci* 2 (1):50-57. doi: 10.7150/ijms.2.50.

Hughes, S. A., H. Wedemeyer, and P. M. Harrison. 2011. "Hepatitis delta virus." *Lancet* 378 (9785):73-85. doi: 10.1016/S0140-6736(10)61931-9.

Huo, T. I., J. C. Wu, R. Y. Lin, W. Y. Sheng, F. Y. Chang, and S. D. Lee. 1997. "Decreasing hepatitis D virus infection in Taiwan: an analysis of contributory factors." *J Gastroenterol Hepatol* 12 (11):747-51.

Husa, P., A. Linhartova, V. Nemecek, and L. Husova. 2005. "Hepatitis D." *Acta Virol* 49 (4):219-25.

Ivaniushina, V., N. Radjef, M. Alexeeva, E. Gault, S. Semenov, M. Salhi, O. Kiselev, and P. Deny. 2001. "Hepatitis delta virus genotypes I and II cocirculate in an endemic area of Yakutia, Russia." *J Gen Virol* 82 (Pt 11):2709-18. doi: 10.1099/0022-1317-82-11-2709.

Jakubovics, N., M. Greenwood, and J. G. Meechan. 2014. "General medicine and surgery for dental practitioners: part 4. Infections and infection control." *Br Dent J* 217 (2):73-7. doi: 10.1038/sj.bdj.2014.593.

Jayasundara, D., I. Saeed, S. Maheswararajah, B. C. Chang, S. L. Tang, and S. K. Halgamuge. 2015. "ViQuaS: an improved reconstruction pipeline for viral quasispecies spectra generated by next-generation sequencing." *Bioinformatics* 31 (6):886-96. doi: 10.1093/bioinformatics/btu754.

Jokelainen, P. T., K. Krohn, A. M. Prince, and N. D. Finlayson. 1970. "Electrn microscopic observations on virus-like particles associated with SH antigen." *J Virol* 6 (5):685-9.

Jordheim, L. P., D. Durantel, F. Zoulim, and C. Dumontet. 2013. "Advances in the development of nucleoside and nucleotide analogues for cancer and viral diseases." *Nat Rev Drug Discov* 12 (6):447-64. doi: 10.1038/nrd4010.

Kabacam, G., F. O. Onder, M. Yakut, G. Seven, S. C. Karatayli, E. Karatayli, B. Savas, R. Idilman, A. M. Bozdayi, and C. Yurdaydin. 2012. "Entecavir treatment of chronic hepatitis D." *Clin Infect Dis* 55 (5):645-50. doi: 10.1093/cid/cis459.

Kalinina, O. V., P. S. Novichkov, A. A. Mironov, M. S. Gelfand, and A. B. Rakhmaninova. 2004. "SDPpred: a tool for prediction of amino acid residues that determine differences in functional specificity of homologous proteins." *Nucleic Acids Res* 32 (Web Server issue):W424-8. doi: 10.1093/nar/gkh391.

Kao, J. H. 2002. "Hepatitis B viral genotypes: clinical relevance and molecular characteristics." *J Gastroenterol Hepatol* 17 (6):643-50. doi: 10.1046/j.1440-1746.2002.02737.x.

Karimzadeh, H., Z. Usman, D. Frishman, and M. Roggendorf. 2019a. "Genetic diversity of hepatitis D virus genotype-1 in Europe allows classification into subtypes." *J Viral Hepat*. doi: 10.1111/jvh.13086.

Karimzadeh, H., Z. Usman, D. Frishman, and M. Roggendorf. 2019b. "Genetic diversity of hepatitis D virus genotype-1 in Europe allows classification into subtypes." *J Viral Hepat* 26 (7):900-910. doi: 10.1111/jvh.13086.

Kay, A., and F. Zoulim. 2007. "Hepatitis B virus genetic variability and evolution." *Virus Res* 127 (2):164-76. doi: 10.1016/j.virusres.2007.02.021.

Kidd-Ljunggren, K., Y. Miyakawa, and A. H. Kidd. 2002. "Genetic variability in hepatitis B viruses." *J Gen Virol* 83 (Pt 6):1267-80. doi: 10.1099/0022-1317-83-6-1267.

Kimura, M. 1980. "A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences." *J Mol Evol* 16 (2):111-20.

Kimura, M. 1981. "Estimation of evolutionary distances between homologous nucleotide sequences." *Proc Natl Acad Sci U S A* 78 (1):454-8. doi: 10.1073/pnas.78.1.454.

Kinsella, M., O. Harismendy, M. Nakano, K. A. Frazer, and V. Bafna. 2011. "Sensitive gene fusion detection using ambiguously mapping RNA-Seq read pairs." *Bioinformatics* 27 (8):1068-75. doi: 10.1093/bioinformatics/btr085.

Kircher, M., and J. Kelso. 2010. "High-throughput DNA sequencing--concepts and limitations." *Bioessays* 32 (6):524-36. doi: 10.1002/bies.200900181.

Knierim, E., B. Lucke, J. M. Schwarz, M. Schuelke, and D. Seelow. 2011. "Systematic comparison of three methods for fragmentation of long-range PCR products for next generation sequencing." *PLoS One* 6 (11):e28240. doi: 10.1371/journal.pone.0028240.

Koboldt, D. C., K. Chen, T. Wylie, D. E. Larson, M. D. McLellan, E. R. Mardis, G. M. Weinstock, R. K. Wilson, and L. Ding. 2009. "VarScan: variant detection in massively parallel sequencing of individual and pooled samples." *Bioinformatics* 25 (17):2283-5. doi: 10.1093/bioinformatics/btp373.

Koboldt, D. C., Q. Zhang, D. E. Larson, D. Shen, M. D. McLellan, L. Lin, C. A. Miller, E. R. Mardis, L. Ding, and R. K. Wilson. 2012. "VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing." *Genome Res* 22 (3):568-76. doi: 10.1101/gr.129684.111.

Koh, C., L. Canini, H. Dahari, X. Zhao, S. L. Uprichard, V. Haynes-Williams, M. A. Winters, G. Subramanya, S. L. Cooper, P. Pinto, E. F. Wolff, R. Bishop, M. Ai Thanda Han, S. J. Cotler, D. E. Kleiner, O. Keskin, R. Idilman, C. Yurdaydin, J. S. Glenn, and T. Heller. 2015. "Oral prenylation inhibition with lonafarnib in chronic hepatitis D infection: a proof-of-concept randomised, double-blind, placebo-controlled phase 2A trial." *Lancet Infect Dis* 15 (10):1167-1174. doi: 10.1016/S1473-3099(15)00074-2.

Koh, Y., I. Park, C. H. Sun, S. Lee, H. Yun, C. K. Park, S. H. Park, J. K. Park, and S. H. Lee. 2015. "Detection of a Distinctive Genomic Signature in Rhabdoid Glioblastoma, A Rare Disease Entity Identified by Whole Exome Sequencing and Whole Transcriptome Sequencing." *Transl Oncol* 8 (4):279-87. doi: 10.1016/j.tranon.2015.05.003.

Kondo, Y., M. Ninomiya, E. Kakazu, O. Kimura, and T. Shimosegawa. 2013. "Hepatitis B surface antigen could contribute to the immunopathogenesis of hepatitis B virus infection." *ISRN Gastroenterol* 2013:935295. doi: 10.1155/2013/935295.

Kramvis, A. 2014. "Genotypes and genetic variability of hepatitis B virus." *Intervirology* 57 (3-4):141-50. doi: 10.1159/000360947.

Kramvis, A. 2016. "The clinical implications of hepatitis B virus genotypes and HBeAg in pediatrics." *Rev Med Virol* 26 (4):285-303. doi: 10.1002/rmv.1885.

Kramvis, A., and M. C. Kew. 2007. "Molecular characterization of subgenotype A1 (subgroup Aa) of hepatitis B virus." *Hepatol Res* 37 (s1):S27-32. doi: 10.1111/j.1872-034X.2007.00100.x.

Krawczak, M., E. V. Ball, I. Fenton, P. D. Stenson, S. Abeysinghe, N. Thomas, and D. N. Cooper. 2000. "Human gene mutation database-a biomedical information and

research resource." *Hum Mutat* 15 (1):45-51. doi: 10.1002/(SICI)1098-1004(200001)15:1<45::AID-HUMU10>3.0.CO;2-T.

Kucirka, L. M., H. Farzadegan, J. J. Feld, S. H. Mehta, M. Winters, J. S. Glenn, G. D. Kirk, D. L. Segev, K. E. Nelson, M. Marks, T. Heller, and E. T. Golub. 2010. "Prevalence, correlates, and viral dynamics of hepatitis delta among injection drug users." *J Infect Dis* 202 (6):845-52. doi: 10.1086/655808.

Kuiken, C., P. Hraber, J. Thurmond, and K. Yusim. 2008. "The hepatitis C sequence database in Los Alamos." *Nucleic Acids Res* 36 (Database issue):D512-6. doi: 10.1093/nar/gkm962.

Kuiken, C., B. Korber, and R. W. Shafer. 2003. "HIV sequence databases." *AIDS Rev* 5 (1):52-61.

Kumar, M., S. Y. Jung, A. J. Hodgson, C. R. Madden, J. Qin, and B. L. Slagle. 2011. "Hepatitis B virus regulatory HBx protein binds to adaptor protein IPS-1 and inhibits the activation of beta interferon." *J Virol* 85 (2):987-95. doi: 10.1128/JVI.01825-10.

Kumar, S., M. Nei, J. Dudley, and K. Tamura. 2008. "MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences." *Brief Bioinform* 9 (4):299-306. doi: 10.1093/bib/bbn017.

Lai, M. M. 1995. "The molecular biology of hepatitis delta virus." *Annu Rev Biochem* 64:259-86. doi: 10.1146/annurev.bi.64.070195.001355.

Lamas Longarela, O., T. T. Schmidt, K. Schoneweis, R. Romeo, H. Wedemeyer, S. Urban, and A. Schulze. 2013. "Proteoglycans act as cellular hepatitis delta virus attachment receptors." *PLoS One* 8 (3):e58340. doi: 10.1371/journal.pone.0058340.

Lampertico, P., and Y. F. Liaw. 2012. "New perspectives in the therapy of chronic hepatitis B." *Gut* 61 Suppl 1:i18-24. doi: 10.1136/gutjnl-2012-302085.

Landrum, M. J., J. M. Lee, M. Benson, G. Brown, C. Chao, S. Chitipiralla, B. Gu, J. Hart, D. Hoffman, J. Hoover, W. Jang, K. Katz, M. Ovetsky, G. Riley, A. Sethi, R. Tully, R. Villamarin-Salomon, W. Rubinstein, and D. R. Maglott. 2016. "ClinVar: public archive of interpretations of clinically relevant variants." *Nucleic Acids Res* 44 (D1):D862-8. doi: 10.1093/nar/gkv1222.

Langmead, B., C. Trapnell, M. Pop, and S. L. Salzberg. 2009. "Ultrafast and memory-efficient alignment of short DNA sequences to the human genome." *Genome Biol* 10 (3):R25. doi: 10.1186/gb-2009-10-3-r25.

Lanthaler, B., S. Wieser, A. Deutschmann, A. Schossig, C. Fauth, J. Zschocke, and M. Witsch-Baumgartner. 2014. "Genotype-based databases for variants causing rare diseases." *Gene* 550 (1):136-40. doi: 10.1016/j.gene.2014.08.016.

Lau, D. T., E. Doo, Y. Park, D. E. Kleiner, P. Schmid, M. C. Kuhns, and J. H. Hoofnagle. 1999. "Lamivudine for chronic delta hepatitis." *Hepatology* 30 (2):546-9. doi: 10.1002/hep.510300217.

Lau, G. K., T. Piratvisuth, K. X. Luo, P. Marcellin, S. Thongsawat, G. Cooksley, E. Gane, M. W. Fried, W. C. Chow, S. W. Paik, W. Y. Chang, T. Berg, R. Flisiak, P. McCloud, N. Pluck, and HBeAg-Positive Chronic Hepatitis B. Study Group Peginterferon Alfa-2a. 2005. "Peginterferon Alfa-2a, lamivudine, and the combination for HBeAg-positive chronic hepatitis B." *N Engl J Med* 352 (26):2682-95. doi: 10.1056/NEJMoa043470.

Le Gal, F., S. Brichler, T. Drugan, C. Alloui, D. Roulot, J. M. Pawlotsky, P. Deny, and E. Gordien. 2017. "Genetic diversity and worldwide distribution of the deltavirus genus: A study of 2,152 clinical strains." *Hepatology* 66 (6):1826-1841. doi: 10.1002/hep.29574.

Le Gal, F., E. Gault, M. P. Ripault, J. Serpaggi, J. C. Trinchet, E. Gordien, and P. Deny. 2006. "Eighth major clade for hepatitis delta virus." *Emerg Infect Dis* 12 (9):1447-50. doi: 10.3201/eid1209.060112.

Lebosse, F., B. Testoni, J. Fresquet, F. Facchetti, E. Galmozzi, M. Fournier, V. Hervieu, P. Berthillon, F. Berby, I. Bordes, D. Durantel, M. Levrero, P. Lampertico, and F. Zoulim. 2017. "Intrahepatic innate immune response pathways are downregulated in untreated chronic hepatitis B." *J Hepatol* 66 (5):897-909. doi: 10.1016/j.jhep.2016.12.024.

Lee, C. H., S. C. Chang, C. H. Wu, and M. F. Chang. 2001. "A novel chromosome region maintenance 1-independent nuclear export signal of the large form of hepatitis delta antigen that is required for the viral assembly." *J Biol Chem* 276 (11):8142-8. doi: 10.1074/jbc.M004477200.

Lee, C. M., F. Y. Bih, Y. C. Chao, S. Govindarajan, and M. M. Lai. 1992. "Evolution of hepatitis delta virus RNA during chronic infection." *Virology* 188 (1):265-73.

Lee, C. M., C. S. Changchien, J. C. Chung, and Y. F. Liaw. 1996. "Characterization of a new genotype II hepatitis delta virus from Taiwan." *J Med Virol* 49 (2):145-54. doi: 10.1002/(SICI)1096-9071(199606)49:2<145::AID-JMV12>3.0.CO;2-D.

Lee, C. Z., J. H. Lin, M. Chao, K. McKnight, and M. M. Lai. 1993. "RNA-binding activity of hepatitis delta antigen involves two arginine-rich motifs and is required for hepatitis delta virus RNA replication." *J Virol* 67 (4):2221-7.

Lemey, P., A. Rambaut, A. J. Drummond, and M. A. Suchard. 2009. "Bayesian phylogeography finds its roots." *PLoS Comput Biol* 5 (9):e1000520. doi: 10.1371/journal.pcbi.1000520.

Li, H. 2013. " Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.". doi: arXiv:1303.3997 [q-bio.GN].

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and Subgroup Genome Project Data Processing. 2009. "The Sequence Alignment/Map format and SAMtools." *Bioinformatics* 25 (16):2078-9. doi: 10.1093/bioinformatics/btp352.

Li, W., and A. Godzik. 2006. "Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences." *Bioinformatics* 22 (13):1658-9. doi: 10.1093/bioinformatics/btl158.

Lin, C. L., and J. H. Kao. 2015. "Hepatitis B virus genotypes and variants." *Cold Spring Harb Perspect Med* 5 (5):a021436. doi: 10.1101/cshperspect.a021436.

Liu, C. J., and J. H. Kao. 2013. "Global perspective on the natural history of chronic hepatitis B: role of hepatitis B virus genotypes A to J." *Semin Liver Dis* 33 (2):97-102. doi: 10.1055/s-0033-1345716.

Liu, F., L. Chen, D. M. Yu, L. Deng, R. Chen, Y. Jiang, L. Chen, S. Y. Huang, J. L. Yu, Q. M. Gong, and X. X. Zhang. 2011. "Evolutionary patterns of hepatitis B virus quasispecies under different selective pressures: correlation with antiviral efficacy." *Gut* 60 (9):1269-77. doi: 10.1136/gut.2010.226225.

Liu, Y., J. Li, J. Chen, Y. Li, W. Wang, X. Du, W. Song, W. Zhang, L. Lin, and Z. Yuan. 2015. "Hepatitis B virus polymerase disrupts K63-linked ubiquitination of STING to block innate cytosolic DNA-sensing pathways." *J Virol* 89 (4):2287-300. doi: 10.1128/JVI.02760-14.

Lok, A. S., and B. J. McMahon. 2009. "Chronic hepatitis B: update 2009." *Hepatology* 50 (3):661-2. doi: 10.1002/hep.23190.

Luangsay, S., M. Gruffaz, N. Isorce, B. Testoni, M. Michelet, S. Faure-Dupuy, S. Maadadi, M. Ait-Goughoulte, R. Parent, M. Rivoire, H. Javanbakht, J. Lucifora, D. Durantel, and F. Zoulim. 2015. "Early inhibition of hepatocyte innate responses by hepatitis B virus." *J Hepatol* 63 (6):1314-22. doi: 10.1016/j.jhep.2015.07.014.

Macalalad, A. R., M. C. Zody, P. Charlebois, N. J. Lennon, R. M. Newman, C. M. Malboeuf, E. M. Ryan, C. L. Boutwell, K. A. Power, D. E. Brackney, K. N. Pesko, J. Z. Levin, G. D. Ebel, T. M. Allen, B. W. Birren, and M. R. Henn. 2012. "Highly sensitive and specific detection of rare variants in mixed viral populations from massively parallel sequence data." *PLoS Comput Biol* 8 (3):e1002417. doi: 10.1371/journal.pcbi.1002417.

Makuwa, M., M. Caron, S. Souquiere, G. Malonga-Mouelet, A. Mahe, and M. Kazanji. 2008. "Prevalence and genetic diversity of hepatitis B and delta viruses in pregnant women in Gabon: molecular evidence that hepatitis delta virus clade 8 originates from and is endemic in central Africa." *J Clin Microbiol* 46 (2):754-6. doi: 10.1128/JCM.02142-07.

Marine, R., S. W. Polson, J. Ravel, G. Hatfull, D. Russell, M. Sullivan, F. Syed, M. Dumas, and K. E. Wommack. 2011. "Evaluation of a transposase protocol for rapid

generation of shotgun high-throughput sequencing libraries from nanogram quantities of DNA." *Appl Environ Microbiol* 77 (22):8071-9. doi: 10.1128/AEM.05610-11.

Marino, Q., M. Cisse, A. Gerber, O. Dolo, S. Sayon, A. Ba, S. Brichler, F. T. Traore, E. Gordien, J. Togo, V. Calvez, A. G. Marcelin, A. I. Maiga, and E. Todesco. 2019. "Low hepatitis D seroprevalence in blood donors of Bamako, Mali." *Infect Dis (Lond)* 51 (8):622-624. doi: 10.1080/23744235.2019.1620963.

Marion, P. L. 1988. "Use of animal models to study hepatitis B virus." *Prog Med Virol* 35:43-75.

Marion, P. L., M. J. Van Davelaar, S. S. Knight, F. H. Salazar, G. Garcia, H. Popper, and W. S. Robinson. 1986. "Hepatocellular carcinoma in ground squirrels persistently infected with ground squirrel hepatitis virus." *Proc Natl Acad Sci U S A* 83 (12):4543-6. doi: 10.1073/pnas.83.12.4543.

Martin, M. 2011. "Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads." *EMBnet Journal* 17:10-12.

Mason, W. S. 2015. "Animal models and the molecular biology of hepadnavirus infection." *Cold Spring Harb Perspect Med* 5 (4). doi: 10.1101/cshperspect.a021352.

Maynard, J. E., K. R. Berquist, D. H. Krushak, and R. H. Purcell. 1972. "Experimental infection of chimpanzees with the virus of hepatitis B." *Nature* 237 (5357):514-5. doi: 10.1038/237514a0.

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, and M. A. DePristo. 2010. "The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data." *Genome Res* 20 (9):1297-303. doi: 10.1101/gr.107524.110.

McPherson, A., F. Hormozdiari, A. Zayed, R. Giuliany, G. Ha, M. G. Sun, M. Griffith, A. Heravi Moussavi, J. Senz, N. Melnyk, M. Pacheco, M. A. Marra, M. Hirst, T. O. Nielsen, S. C. Sahinalp, D. Huntsman, and S. P. Shah. 2011. "deFuse: an algorithm for gene fusion discovery in tumor RNA-Seq data." *PLoS Comput Biol* 7 (5):e1001138. doi: 10.1371/journal.pcbi.1001138.

Mese, S., M. Arikan, A. Cakiris, N. Abaci, E. Gumus, O. Kursun, D. Onel, D. Ustek, S. Kaymakoglu, S. Badur, O. S. Yenen, and E. Bozkaya. 2013. "Role of the line probe assay INNO-LiPA HBV DR and ultradeep pyrosequencing in detecting resistance mutations to nucleoside/nucleotide analogues in viral samples isolated from chronic hepatitis B patients." *J Gen Virol* 94 (Pt 12):2729-38. doi: 10.1099/vir.0.053041-0.

Miao, Z., S. Zhang, Z. Ma, M. S. Hakim, W. Wang, M. P. Peppelenbosch, and Q. Pan. 2018. "Recombinant identification, molecular classification and proposed reference genomes for hepatitis delta virus." *J Viral Hepat*. doi: 10.1111/jvh.13010.

Micco, L., D. Peppa, E. Loggi, A. Schurich, L. Jefferson, C. Cursaro, A. M. Panno, M. Bernardi, C. Brander, F. Bihl, P. Andreone, and M. K. Maini. 2013. "Differential boosting of innate and adaptive antiviral responses during pegylated-interferon-alpha therapy of chronic hepatitis B." *J Hepatol* 58 (2):225-33. doi: 10.1016/j.jhep.2012.09.029.

Modahl, L. E., T. B. Macnaughton, N. Zhu, D. L. Johnson, and M. M. Lai. 2000. "RNA-Dependent replication and transcription of hepatitis delta virus RNA involve distinct cellular RNA polymerases." *Mol Cell Biol* 20 (16):6030-9. doi: 10.1128/mcb.20.16.6030-6039.2000.

Mohd-Ismail, N. K., Z. Lim, J. Gunaratne, and Y. J. Tan. 2019. "Mapping the Interactions of HBV cccDNA with Host Factors." *Int J Mol Sci* 20 (17). doi: 10.3390/ijms20174276.

Moura, I. F., E. P. Lopes, M. V. Alvarado-Mora, J. R. Pinho, and F. J. Carrilho. 2013. "Phylogenetic analysis and subgenotypic distribution of the hepatitis B virus in Recife, Brazil." *Infect Genet Evol* 14:195-9. doi: 10.1016/j.meegid.2012.11.022.

Nakano, M., J. Komatsu, S. Matsuura, K. Takashima, S. Katsura, and A. Mizuno. 2003. "Single-molecule PCR using water-in-oil emulsion." *J Biotechnol* 102 (2):117-24.

Nassal, M. 2015. "HBV cccDNA: viral persistence reservoir and key obstacle for a cure of chronic hepatitis B." *Gut* 64 (12):1972-84. doi: 10.1136/gutjnl-2015-309809.

Nath, N., I. K. Mushahwar, C. T. Fang, H. Berberian, and R. Y. Dodd. 1985. "Antibodies to delta antigen in asymptomatic hepatitis B surface antigen-reactive blood donors in the United States and their association with other markers of hepatitis B virus." *Am J Epidemiol* 122 (2):218-25.

Navascues, C. A., M. Rodriguez, N. G. Sotorrio, P. Sala, A. Linares, A. Suarez, and L. Rodrigo. 1995. "Epidemiology of hepatitis D virus infection: changes in the last 14 years." *Am J Gastroenterol* 90 (11):1981-4.

Negro, F. 2014. "Hepatitis D virus coinfection and superinfection." *Cold Spring Harb Perspect Med* 4 (11):a021550. doi: 10.1101/cshperspect.a021550.

Negro, F., K. F. Bergmann, B. M. Baroudy, W. C. Satterfield, H. Popper, R. H. Purcell, and J. L. Gerin. 1988. "Chronic hepatitis D virus (HDV) infection in hepatitis B virus carrier chimpanzees experimentally superinfected with HDV." *J Infect Dis* 158 (1):151-9. doi: 10.1093/infdis/158.1.151.

Negro, F., B. E. Korba, B. Forzani, B. M. Baroudy, T. L. Brown, J. L. Gerin, and A. Ponzetto. 1989. "Hepatitis delta virus (HDV) and woodchuck hepatitis virus (WHV) nucleic acids in tissues of HDV-infected chronic WHV carrier woodchucks." *J Virol* 63 (4):1612-8.

Ni, Y., F. A. Lempp, S. Mehrle, S. Nkongolo, C. Kaufman, M. Falth, J. Stindt, C. Koniger, M. Nassal, R. Kubitz, H. Sultmann, and S. Urban. 2014. "Hepatitis B and D viruses exploit sodium taurocholate co-transporting polypeptide for species-specific entry into hepatocytes." *Gastroenterology* 146 (4):1070-83. doi: 10.1053/j.gastro.2013.12.024.

Niro, G. A., A. Ciancio, G. B. Gaeta, A. Smedile, A. Marrone, A. Olivero, M. Stanzione, E. David, G. Brancaccio, R. Fontana, F. Perri, A. Andriulli, and M. Rizzetto. 2006. "Pegylated interferon alpha-2b as monotherapy or in combination with ribavirin in chronic hepatitis delta." *Hepatology* 44 (3):713-20. doi: 10.1002/hep.21296.

Noordeen, F., C. A. Scougall, A. Grosse, Q. Qiao, B. B. Ajilian, G. Reaiche-Miller, J. Finnie, M. Werner, R. Broering, J. F. Schlaak, A. Vaillant, and A. R. Jilbert. 2015. "Therapeutic Antiviral Effect of the Nucleic Acid Polymer REP 2055 against Persistent Duck Hepatitis B Virus Infection." *PLoS One* 10 (11):e0140909. doi: 10.1371/journal.pone.0140909.

Noordeen, F., A. Vaillant, and A. R. Jilbert. 2013a. "Nucleic acid polymers inhibit duck hepatitis B virus infection in vitro." *Antimicrob Agents Chemother* 57 (11):5291-8. doi: 10.1128/AAC.01003-13.

Noordeen, F., A. Vaillant, and A. R. Jilbert. 2013b. "Nucleic acid polymers prevent the establishment of duck hepatitis B virus infection in vivo." *Antimicrob Agents Chemother* 57 (11):5299-306. doi: 10.1128/AAC.01005-13.

Osiowy, C., E. Giles, Y. Tanaka, M. Mizokami, and G. Y. Minuk. 2006. "Molecular evolution of hepatitis B virus over 25 years." *J Virol* 80 (21):10307-14. doi: 10.1128/JVI.00996-06.

Ottobrelli, A., A. Marzano, A. Smedile, S. Recchia, M. Salizzoni, C. Cornu, M. E. Lamy, J. B. Otte, B. De Hemptinne, A. Geubel, and et al. 1991. "Patterns of hepatitis delta virus reinfection and disease in liver transplantation." *Gastroenterology* 101 (6):1649-55.

Pascarella, S., and F. Negro. 2011. "Hepatitis D virus: an update." *Liver Int* 31 (1):7-21. doi: 10.1111/j.1478-3231.2010.02320.x.

Pemov, A., H. Modi, D. P. Chandler, and S. Bavykin. 2005. "DNA analysis with multiplex microarray-enhanced PCR." *Nucleic Acids Res* 33 (2):e11. doi: 10.1093/nar/gnh184.

Peng, L., X. W. Bian, D. K. Li, C. Xu, G. M. Wang, Q. Y. Xia, and Q. Xiong. 2015. "Large-scale RNA-Seq Transcriptome Analysis of 4043 Cancers and 548 Normal Tissue Controls across 12 TCGA Cancer Types." *Sci Rep* 5:13413. doi: 10.1038/srep13413.

Polaris Observatory, Collaborators. 2018. "Global prevalence, treatment, and prevention of hepatitis B virus infection in 2016: a modelling study." *Lancet Gastroenterol Hepatol* 3 (6):383-403. doi: 10.1016/S2468-1253(18)30056-6.

Pourkarim, M. R., S. Amini-Bavil-Olyaee, F. Kurbanov, M. Van Ranst, and F. Tacke. 2014. "Molecular identification of hepatitis B virus genotypes/subgenotypes: revised classification hurdles and updated resolutions." *World J Gastroenterol* 20 (23):7152-68. doi: 10.3748/wjg.v20.i23.7152.

Prabhakaran, S., M. Rey, O. Zagordi, N. Beerenwinkel, and V. Roth. 2014. "HIV Haplotype Inference Using a Propagating Dirichlet Process Mixture Model." *IEEE/ACM Trans Comput Biol Bioinform* 11 (1):182-91. doi: 10.1109/TCBB.2013.145.

Prosperi, M. C., and M. Salemi. 2012. "QuRe: software for viral quasispecies reconstruction from next-generation sequencing data." *Bioinformatics* 28 (1):132-3. doi: 10.1093/bioinformatics/btr627.

Quinet, J., C. Jamard, M. Burtin, M. Lemasson, S. Guerret, C. Sureau, A. Vaillant, and L. Cova. 2018. "Nucleic acid polymer REP 2139 and nucleos(T)ide analogues act synergistically against chronic hepadnaviral infection in vivo in Pekin ducks." *Hepatology* 67 (6):2127-2140. doi: 10.1002/hep.29737.

Quintero, A., N. Uzcategui, C. L. Loureiro, L. Villegas, X. Illarramendi, M. E. Guevara, J. E. Ludert, L. Blitz, F. Liprandi, and F. H. Pujol. 2001. "Hepatitis delta virus genotypes I and III circulate associated with hepatitis B virus genotype F In Venezuela." *J Med Virol* 64 (3):356-9.

Radjef, N., E. Gordien, V. Ivaniushina, E. Gault, P. Anais, T. Drugan, J. C. Trinchet, D. Roulot, M. Tamby, M. C. Milinkovitch, and P. Deny. 2004. "Molecular phylogenetic analyses indicate a wide and ancient radiation of African hepatitis delta virus, suggesting a deltavirus genus of at least seven major clades." *J Virol* 78 (5):2537-44.

Rizzetto, M. 2015. "Hepatitis D Virus: Introduction and Epidemiology." *Cold Spring Harb Perspect Med* 5 (7):a021576. doi: 10.1101/cshperspect.a021576.

Rizzetto, M., M. G. Canese, S. Arico, O. Crivelli, C. Trepo, F. Bonino, and G. Verme. 1977. "Immunofluorescence detection of new antigen-antibody system (delta/anti-delta) associated to hepatitis B virus in liver and in serum of HBsAg carriers." *Gut* 18 (12):997-1003.

Rizzetto, M., M. G. Canese, J. L. Gerin, W. T. London, D. L. Sly, and R. H. Purcell. 1980. "Transmission of the hepatitis B virus-associated delta antigen to chimpanzees." *J Infect Dis* 141 (5):590-602.

Rodriguez, C., S. Chevaliez, P. Bensadoun, and J. M. Pawlotsky. 2013. "Characterization of the dynamics of hepatitis B virus resistance to adefovir by ultra-deep pyrosequencing." *Hepatology* 58 (3):890-901. doi: 10.1002/hep.26383.

Ronquist, F., M. Teslenko, P. van der Mark, D. L. Ayres, A. Darling, S. Hohna, B. Larget, L. Liu, M. A. Suchard, and J. P. Huelsenbeck. 2012. "MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space." *Syst Biol* 61 (3):539-42. doi: 10.1093/sysbio/sys029.

Ross, R. S., S. Viazov, K. Renzing-Kohler, and M. Roggendorf. 2000. "Changes in the epidemiology of hepatitis C infection in Germany: shift in the predominance of hepatitis C subtypes." *J Med Virol* 60 (2):122-5.

Sachdev, P. S., D. M. Lipnicki, N. A. Kochan, J. D. Crawford, K. Rockwood, S. Xiao, J. Li, X. Li, C. Brayne, F. E. Matthews, B. C. Stephan, R. B. Lipton, M. J. Katz, K. Ritchie, I. Carriere, M. L. Ancelin, S. Seshadri, R. Au, A. S. Beiser, L. C. Lam, C. H. Wong, A. W. Fung, K. W. Kim, J. W. Han, T. H. Kim, R. C. Petersen, R. O. Roberts, M. M. Mielke, M. Ganguli, H. H. Dodge, T. Hughes, K. J. Anstey, N. Cherbuin, P. Butterworth, T. P. Ng, Q. Gao, S. Reppermund, H. Brodaty, K. Meguro, N. Schupf, J. Manly, Y. Stern, A. Lobo, R. Lopez-Anton, J. Santabarbara, and Cosmic. 2013. "COSMIC (Cohort Studies of Memory in an International Consortium): an international consortium to identify risk and protective factors and biomarkers of cognitive ageing and dementia in diverse ethnic and sociocultural groups." *BMC Neurol* 13:165. doi: 10.1186/1471-2377-13-165.

Sagnelli, E., T. Stroffolini, A. Ascione, M. Chiaramonte, A. Craxi, G. Giusti, and F. Piccinino. 1997. "Decrease in HDV endemicity in Italy." *J Hepatol* 26 (1):20-4.

Sakugawa, H., H. Nakasone, T. Nakayoshi, Y. Kawakami, S. Miyazato, F. Kinjo, A. Saito, S. P. Ma, H. Hotta, and M. Kinoshita. 1999. "Hepatitis delta virus genotype IIb predominates in an endemic area, Okinawa, Japan." *J Med Virol* 58 (4):366-72.

Salzberg, S. L., A. M. Phillippy, A. Zimin, D. Puiu, T. Magoc, S. Koren, T. J. Treangen, M. C. Schatz, A. L. Delcher, M. Roberts, G. Marcais, M. Pop, and J. A. Yorke. 2012. "GAGE: A critical evaluation of genome assemblies and assembly algorithms." *Genome Res* 22 (3):557-67. doi: 10.1101/gr.131383.111.

Sanger, F. 1975. "The Croonian Lecture, 1975. Nucleotide sequences in DNA." *Proc R Soc Lond B Biol Sci* 191 (1104):317-33. doi: 10.1098/rspb.1975.0131.

Santos, M. D., M. S. Gomes-Gouvea, J. D. Nunes, L. M. Barros, F. J. Carrilho, S. Ferreira Ade, and J. R. Pinho. 2016. "The hepatitis delta genotype 8 in Northeast Brazil: The North Atlantic slave trade as the potential route for infection." *Virus Res* 224:6-11. doi: 10.1016/j.virusres.2016.08.003.

Sato, S., K. Li, T. Kameyama, T. Hayashi, Y. Ishida, S. Murakami, T. Watanabe, S. Iijima, Y. Sakurai, K. Watashi, S. Tsutsumi, Y. Sato, H. Akita, T. Wakita, C. M. Rice, H. Harashima, M. Kohara, Y. Tanaka, and A. Takaoka. 2015. "The RNA sensor RIG-I dually functions as an innate sensor and direct antiviral factor for hepatitis B virus." *Immunity* 42 (1):123-32. doi: 10.1016/j.immuni.2014.12.016.

Scarponi, C. F. O., Rdnd Silva, J. A. Souza Filho, M. R. L. Guerra, M. A. F. Pedrosa, and M. P. G. Mol. 2019. "Hepatitis Delta Prevalence in South America: A Systematic Review and Meta-Analysis." *Rev Soc Bras Med Trop* 52:e20180289. doi: 10.1590/0037-8682-0289-2018.

Schmidt, H. A., K. Strimmer, M. Vingron, and A. von Haeseler. 2002. "TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing." *Bioinformatics* 18 (3):502-4. doi: 10.1093/bioinformatics/18.3.502.

Schoneweis, K., N. Motter, P. L. Roppert, M. Lu, B. Wang, I. Roehl, D. Glebe, D. Yang, J. D. Morrey, M. Roggendorf, and A. Vaillant. 2018. "Activity of nucleic acid polymers in rodent models of HBV infection." *Antiviral Res* 149:26-33. doi: 10.1016/j.antiviral.2017.10.022.

Shepard, C. W., E. P. Simard, L. Finelli, A. E. Fiore, and B. P. Bell. 2006. "Hepatitis B virus infection: epidemiology and vaccination." *Epidemiol Rev* 28:112-25. doi: 10.1093/epirev/mxj009.

Shi, Y. H. 2012. "Correlation between hepatitis B virus genotypes and clinical outcomes." *Jpn J Infect Dis* 65 (6):476-82. doi: 10.7883/yoken.65.476.

Shih, H. H., C. Shih, H. W. Wang, C. W. Su, I. J. Sheen, and J. C. Wu. 2010. "Pro-205 of large hepatitis delta antigen and Pro-62 of major hepatitis B surface antigen influence the assembly of different genotypes of hepatitis D virus." *J Gen Virol* 91 (Pt 4):1004-12. doi: 10.1099/vir.0.017541-0.

Shlomai, A., R. E. Schwartz, V. Ramanan, A. Bhatta, Y. P. de Jong, S. N. Bhatia, and C. M. Rice. 2014. "Modeling host interactions with hepatitis B virus using primary and induced pluripotent stem cell-derived hepatocellular systems." *Proc Natl Acad Sci U S A* 111 (33):12193-8. doi: 10.1073/pnas.1412631111.

Splawn, L. M., C. A. Bailey, J. P. Medina, and J. C. Cho. 2018. "Heplisav-B vaccination for the prevention of hepatitis B virus infection in adults in the United States." *Drugs Today (Barc)* 54 (7):399-405. doi: 10.1358/dot.2018.54.7.2833984.

Suchard, M. A., P. Lemey, G. Baele, D. L. Ayres, A. J. Drummond, and A. Rambaut. 2018. "Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10." *Virus Evol* 4 (1):vey016. doi: 10.1093/ve/vey016.

Summers, J., J. M. Smolec, and R. Snyder. 1978. "A virus similar to human hepatitis B virus associated with hepatitis and hepatoma in woodchucks." *Proc Natl Acad Sci U S A* 75 (9):4533-7. doi: 10.1073/pnas.75.9.4533.

Sun, J., Q. Xie, D. Tan, Q. Ning, J. Niu, X. Bai, R. Fan, S. Chen, J. Cheng, Y. Yu, H. Wang, M. Xu, G. Shi, M. Wan, X. Chen, H. Tang, J. Sheng, X. Dou, J. Shi, H. Ren, M. Wang, H. Zhang, Z. Gao, C. Chen, H. Ma, J. Jia, and J. Hou. 2014. "The 104-week efficacy and safety of telbivudine-based optimization strategy in chronic hepatitis B

patients: a randomized, controlled study." *Hepatology* 59 (4):1283-92. doi: 10.1002/hep.26885.
Sunbul, M. 2014. "Hepatitis B virus genotypes: global distribution and clinical importance." *World J Gastroenterol* 20 (18):5427-34. doi: 10.3748/wjg.v20.i18.5427.
Sureau, C. 2006. "The role of the HBV envelope proteins in the HDV replication cycle." *Curr Top Microbiol Immunol* 307:113-31. doi: 10.1007/3-540-29802-9_6.
Sureau, C., and F. Negro. 2016. "The hepatitis delta virus: Replication and pathogenesis." *J Hepatol* 64 (1 Suppl):S102-S116. doi: 10.1016/j.jhep.2016.02.013.
T.H, Jukes, and Cantor C.R. 1969. *Evolution of protein molecules*. Edited by H.N. Munro Ed. New York: American Press.
Tamura, K., and M. Nei. 1993. "Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees." *Mol Biol Evol* 10 (3):512-26. doi: 10.1093/oxfordjournals.molbev.a040023.
Tamura, K., M. Nei, and S. Kumar. 2004. "Prospects for inferring very large phylogenies by using the neighbor-joining method." *Proc Natl Acad Sci U S A* 101 (30):11030-5. doi: 10.1073/pnas.0404206101.
Tatematsu, K., Y. Tanaka, F. Kurbanov, F. Sugauchi, S. Mano, T. Maeshiro, T. Nakayoshi, M. Wakuta, Y. Miyakawa, and M. Mizokami. 2009. "A genetic variant of hepatitis B virus divergent from known human and ape genotypes isolated from a Japanese patient and provisionally assigned to new genotype J." *J Virol* 83 (20):10538-47. doi: 10.1128/JVI.00462-09.
Tateno, Y. 2008. "[International collaboration among DDBJ, EMBL Bank and GenBank]." *Tanpakushitsu Kakusan Koso* 53 (2):182-9.
Taylor, J. M. 2006. "Structure and replication of hepatitis delta virus RNA." *Curr Top Microbiol Immunol* 307:1-23.
Taylor, J. M. 2009. "Chapter 3. Replication of the hepatitis delta virus RNA genome." *Adv Virus Res* 74:103-21. doi: 10.1016/S0065-3527(09)74003-5.
Terrault, N. A., N. H. Bzowej, K. M. Chang, J. P. Hwang, M. M. Jonas, M. H. Murad, and Diseases American Association for the Study of Liver. 2016. "AASLD guidelines for treatment of chronic hepatitis B." *Hepatology* 63 (1):261-83. doi: 10.1002/hep.28156.
Terrault, N. A., A. S. F. Lok, B. J. McMahon, K. M. Chang, J. P. Hwang, M. M. Jonas, R. S. Brown, Jr., N. H. Bzowej, and J. B. Wong. 2018. "Update on prevention, diagnosis, and treatment of chronic hepatitis B: AASLD 2018 hepatitis B guidance." *Hepatology* 67 (4):1560-1599. doi: 10.1002/hep.29800.
Topfer, A., T. Marschall, R. A. Bull, F. Luciani, A. Schonhuth, and N. Beerenwinkel. 2014. "Viral quasispecies assembly via maximal clique enumeration." *PLoS Comput Biol* 10 (3):e1003515. doi: 10.1371/journal.pcbi.1003515.
Topfer, A., O. Zagordi, S. Prabhakaran, V. Roth, E. Halperin, and N. Beerenwinkel. 2013. "Probabilistic inference of viral quasispecies subject to recombination." *J Comput Biol* 20 (2):113-23. doi: 10.1089/cmb.2012.0232.
Trapnell, C., L. Pachter, and S. L. Salzberg. 2009. "TopHat: discovering splice junctions with RNA-Seq." *Bioinformatics* 25 (9):1105-11. doi: 10.1093/bioinformatics/btp120.
Urban, S., A. Schulze, M. Dandri, and J. Petersen. 2010. "The replication cycle of hepatitis B virus." *J Hepatol* 52 (2):282-4. doi: 10.1016/j.jhep.2009.10.031.
Usman, Z., H. Mijocevic, H. Karimzadeh, M. Daumer, M. Al-Mathab, M. Bazinet, D. Frishman, A. Vaillant, and M. Roggendorf. 2019. "Kinetics of hepatitis B surface antigen quasispecies during REP 2139-Ca therapy in HBeAg-positive chronic HBV infection." *J Viral Hepat* 26 (12):1454-1464. doi: 10.1111/jvh.13180.
Usman, Z., S. Velkov, U. Protzer, M. Roggendorf, D. Frishman, and H. Karimzadeh. 2020. "HDVdb: A Comprehensive Hepatitis D Virus Database." *Viruses* 12 (5). doi: 10.3390/v12050538.
Vaillant, A. 2016. "Nucleic acid polymers: Broad spectrum antiviral activity, antiviral mechanisms and optimization for the treatment of hepatitis B and hepatitis D infection." *Antiviral Res* 133:32-40. doi: 10.1016/j.antiviral.2016.07.004.

Vitaliti, G., A. D. Pratico, C. Cimino, G. Di Dio, E. Lionetti, M. La Rosa, and S. Leonardi. 2013. "Hepatitis B vaccine in celiac disease: yesterday, today and tomorrow." *World J Gastroenterol* 19 (6):838-45. doi: 10.3748/wjg.v19.i6.838.

Walter, E., R. Keist, B. Niederost, I. Pult, and H. E. Blum. 1996. "Hepatitis B virus infection of tupaia hepatocytes in vitro and in vivo." *Hepatology* 24 (1):1-5. doi: 10.1002/hep.510240101.

Wang, B. J., Y. J. Tian, Z. J. Meng, M. Jiang, B. Q. Wei, Y. Q. Tao, W. Fan, A. Y. Li, J. J. Bao, X. Y. Li, Z. M. Zhang, Z. D. Wang, H. Wang, M. Roggendorf, M. J. Lu, and D. L. Yang. 2011. "Establishing a new animal model for hepadnaviral infection: susceptibility of Chinese Marmota-species to woodchuck hepatitis virus infection." *J Gen Virol* 92 (Pt 3):681-91. doi: 10.1099/vir.0.025023-0.

Wang, B., Z. Zhu, B. Zhu, J. Wang, Z. Song, S. Huang, W. Fan, Y. Tao, Z. Wang, H. Wang, M. Lu, and D. Yang. 2014. "Nucleoside analogues alone or combined with vaccination prevent hepadnavirus viremia and induce protective immunity: alternative strategy for hepatitis B virus post-exposure prophylaxis." *Antiviral Res* 105:118-25. doi: 10.1016/j.antiviral.2014.02.016.

Wang, K. S., Q. L. Choo, A. J. Weiner, J. H. Ou, R. C. Najarian, R. M. Thayer, G. T. Mullenbach, K. J. Denniston, J. L. Gerin, and M. Houghton. 1986. "Structure, sequence and expression of the hepatitis delta (delta) viral genome." *Nature* 323 (6088):508-14. doi: 10.1038/323508a0.

Wedemeyer, H. 2011. "Hepatitis D revival." *Liver Int* 31 Suppl 1:140-4. doi: 10.1111/j.1478-3231.2010.02408.x.

Wedemeyer, H., and M. P. Manns. 2010. "Epidemiology, pathogenesis and management of hepatitis D: update and challenges ahead." *Nat Rev Gastroenterol Hepatol* 7 (1):31-40. doi: 10.1038/nrgastro.2009.205.

Wieland, S. F. 2015. "The chimpanzee model for hepatitis B virus infection." *Cold Spring Harb Perspect Med* 5 (6). doi: 10.1101/cshperspect.a021469.

Wieland, S., R. Thimme, R. H. Purcell, and F. V. Chisari. 2004. "Genomic analysis of the host response to hepatitis B virus infection." *Proc Natl Acad Sci U S A* 101 (17):6669-74. doi: 10.1073/pnas.0401771101.

Wilgenbusch, J. C., and D. Swofford. 2003. "Inferring evolutionary trees with PAUP*." *Curr Protoc Bioinformatics* Chapter 6:Unit 6 4. doi: 10.1002/0471250953.bi0604s00.

Wille, M., H. J. Netter, M. Littlejohn, L. Yuen, M. Shi, J. S. Eden, M. Klaassen, E. C. Holmes, and A. C. Hurt. 2018. "A Divergent Hepatitis D-Like Agent in Birds." *Viruses* 10 (12). doi: 10.3390/v10120720.

Wolters, L. M., A. B. van Nunen, P. Honkoop, A. C. Vossen, H. G. Niesters, P. E. Zondervan, and R. A. de Man. 2000. "Lamivudine-high dose interferon combination therapy for chronic hepatitis B patients co-infected with the hepatitis D virus." *J Viral Hepat* 7 (6):428-34.

Woo, A. S. J., R. Kwok, and T. Ahmed. 2017. "Alpha-interferon treatment in hepatitis B." *Ann Transl Med* 5 (7):159. doi: 10.21037/atm.2017.03.69.

Xiao, J., R. Liu, and C. S. Chen. 2017. "Tree shrew (Tupaia belangeri) as a novel laboratory disease animal model." *Zool Res* 38 (3):127-137. doi: 10.24272/j.issn.2095-8137.2017.033.

Yamaguchi, Y., J. Filipovska, K. Yano, A. Furuya, N. Inukai, T. Narita, T. Wada, S. Sugimoto, M. M. Konarska, and H. Handa. 2001. "Stimulation of RNA polymerase II elongation by hepatitis delta antigen." *Science* 293 (5527):124-7. doi: 10.1126/science.1057925.

Yan, H., G. Zhong, G. Xu, W. He, Z. Jing, Z. Gao, Y. Huang, Y. Qi, B. Peng, H. Wang, L. Fu, M. Song, P. Chen, W. Gao, B. Ren, Y. Sun, T. Cai, X. Feng, J. Sui, and W. Li. 2012. "Sodium taurocholate cotransporting polypeptide is a functional receptor for human hepatitis B and D virus." *Elife* 1:e00049. doi: 10.7554/eLife.00049.

Yang, S., C. Yu, P. Chen, M. Deng, Q. Cao, Y. Li, J. Ren, K. Xu, J. Yao, T. Xie, C. Wang, Y. Cui, C. Ding, G. Tian, B. Wang, X. Zhang, B. Ruan, and L. Li. 2015. "Protective immune barrier against hepatitis B is needed in individuals born before infant HBV vaccination program in China." *Sci Rep* 5:18334. doi: 10.1038/srep18334.

Yang, Z. T., S. Y. Huang, L. Chen, F. Liu, X. H. Cai, Y. F. Guo, M. J. Wang, Y. Han, D. M. Yu, J. H. Jiang, D. H. Zhang, Q. M. Gong, G. Q. Zhang, G. Q. Zang, Z. H. Lu, L. H. Huang, and X. X. Zhang. 2015. "Characterization of Full-Length Genomes of Hepatitis B Virus Quasispecies in Sera of Patients at Different Phases of Infection." *J Clin Microbiol* 53 (7):2203-14. doi: 10.1128/JCM.00068-15.

Yuen, M. F., and C. L. Lai. 2011. "Treatment of chronic hepatitis B: Evolution over two decades." *J Gastroenterol Hepatol* 26 Suppl 1:138-43. doi: 10.1111/j.1440-1746.2010.06545.x.

Yurdaydin, C., H. Bozkaya, S. Gurel, H. L. Tillmann, N. Aslan, A. Okcu-Heper, E. Erden, K. Yalcin, N. Iliman, O. Uzunalimoglu, M. P. Manns, and A. M. Bozdayi. 2002. "Famciclovir treatment of chronic delta hepatitis." *J Hepatol* 37 (2):266-71.

Yurdaydin, C., R. Idilman, H. Bozkaya, and A. M. Bozdayi. 2010. "Natural history and treatment of chronic delta hepatitis." *J Viral Hepat* 17 (11):749-56. doi: 10.1111/j.1365-2893.2010.01353.x.

Yurdaydin, C., O. Keskin, C. Kalkan, F. Karakaya, A. Caliskan, E. Karatayli, S. Karatayli, A. M. Bozdayi, C. Koh, T. Heller, R. Idilman, and J. S. Glenn. 2018. "Optimizing lonafarnib treatment for the management of chronic delta hepatitis: The LOWR HDV-1 study." *Hepatology* 67 (4):1224-1236. doi: 10.1002/hep.29658.

Zagordi, O., A. Bhattacharya, N. Eriksson, and N. Beerenwinkel. 2011. "ShoRAH: estimating the genetic diversity of a mixed sample from next-generation sequencing data." *BMC Bioinformatics* 12:119. doi: 10.1186/1471-2105-12-119.

Zagordi, O., L. Geyrhofer, V. Roth, and N. Beerenwinkel. 2010. "Deep sequencing of a genetically heterogeneous sample: local haplotype reconstruction and read error correction." *J Comput Biol* 17 (3):417-28. doi: 10.1089/cmb.2009.0164.

Zhang, X. X., M. R. Li, Y. Cao, R. W. Zhang, Y. Zhang, F. Li, H. L. Xi, and X. Y. Xu. 2016. "Dynamics of Genotypic Mutations of the Hepatitis B Virus Associated With Long-Term Entecavir Treatment Determined With Ultradeep Pyrosequencing: A Retrospective Observational Study." *Medicine (Baltimore)* 95 (4):e2614. doi: 10.1097/MD.0000000000002614.

Zhou, Y., and E. C. Holmes. 2007. "Bayesian estimates of the evolutionary rate and age of hepatitis B virus." *J Mol Evol* 65 (2):197-205. doi: 10.1007/s00239-007-0054-1.

Zoulim, F., and S. Locarnini. 2009. "Hepatitis B virus resistance to nucleos(t)ide analogues." *Gastroenterology* 137 (5):1593-608 e1-2. doi: 10.1053/j.gastro.2009.08.063.

8 Appendix

A Research Article I

Received: 7 February 2019 | Accepted: 22 June 2019
DOI: 10.1111/jvh.13180

ORIGINAL ARTICLE

JVH WILEY

Kinetics of hepatitis B surface antigen quasispecies during REP 2139-Ca therapy in HBeAg-positive chronic HBV infection

Zainab Usman[1] | Hrvoje Mijočević[2] | Hadi Karimzadeh[2,3] | Martin Däumer[4] | Mamun Al-Mathab[5] | Michel Bazinet[6] | Dmitrij Frishman[1,7] | Andrew Vaillant[6] | Michael Roggendorf[2]

[1]Department of Bioinformatics, Technische Universität München, Wissenschaftszentrum Weihenstephan, Freising, Germany
[2]Institute of Virology, Technische Universität München, Munich, Germany
[3]Department of Medicine II, University Hospital Munich-Grosshadern, Munich, Germany
[4]Institute of Immunology and Genetics, Kaiserslautern, Germany
[5]Bangabandhu Sheikh Mujib Medical University, Dhaka, Bangladesh
[6]Replicor Inc., QC, Canada
[7]Laboratory of Bioinformatics, RASA research center, St Petersburg State Polytechnical University, Saint Petersburg, Russia

Correspondence
Michael Roggendorf, Institute of Virology, Technische Universität München, Munich, Germany.
Email: michael.roggendorf@tum.de
and
Andrew Vaillant, Replicor Inc., QC, Canada.
Email: availlant@replicor.com

Funding information
Deutscher Akademischer Austauschdienst (DAAD); Replicor Inc.

Abstract

Chronic HBV infection results in various clinical manifestations due to different levels of immune response. In recent years, hepatitis B treatment has improved by long-term administration of nucleos(t)ide analogues (NUCs) and peg-interferon. Nucleic acid polymers (NAPs; REP 2139-Ca and REP 2139-Mg) are new antiviral drugs that block the assembly of subviral particles, thus preventing the release of HBsAg and allowing its clearance and restoration of functional control of infection when combined with various immunotherapies. In the REP 102 study (NCT02646189), 9 of 12 patients showed substantial reduction of HBsAg and seroconversion to anti-HBs in response to REP 2139-Ca, whereas 3 of 12 patients did not show responses (>1 log reduction of HBsAg and HBV DNA from baseline). We characterized the dynamic changes of HBV quasispecies (QS) within the major hydrophilic region (MHR) of the 'pre-S/S' open reading frame including the 'a' determinant in responders and nonresponders of the REP 102 study and four untreated matched controls. HBV QS complexity at baseline varied slightly between responders and nonresponders (P = .28). However, these responders showed significant decline in viral complexity (P = .001) as REP 2139-Ca therapy progressed but no significant change in complexity was observed among the nonresponders (P = .99). The MHR mutations were more frequently observed in responders than in nonresponders and matched controls. No mutations were observed in 'a' determinant of major QS population which may interfere with the detection of HBsAg by diagnostic assays. No specific mutations were found within the MHR which could explain patients' poor HBsAg response during REP 2139-Ca therapy.

KEYWORDS
hepatitis B surface antigen, major hydrophilic region, nucleic acid polymer, quasispecies

Abbreviations: HBsAg, hepatitis B surface antigen; HBV, hepatitis B virus; HCC, hepatocellular carcinoma; MHR, major hydrophilic region; NAPs, nucleic acid polymers; NR, nonresponders; QS, quasispecies; Sn, shannon entropy.

1 | INTRODUCTION

Over the last 30 years preventive hepatitis B vaccination has efficiently decreased the incidence of hepatitis B virus (HBV) infections worldwide. However, a large pool of chronically infected individuals still remain at risk of developing liver cirrhosis and hepatocellular carcinoma (HCC) and serves as a reservoir for HBV transmission. An estimated 291 million people worldwide have been affected with chronic HBV infection leading to 890 000 deaths each year from the complications of liver disease and HCC.[1,2]

The lack of proofreading functionality of HBV reverse transcriptase in chronic HBV patients gives rise to genetically similar variants known as quasispecies (QS) which gradually increase in complexity during the progression of chronic HBV infection with the accumulation of mutations and may play a significant role in pathogenesis, disease progression and drug resistance during chronic infection.[3-6] The number and complexity of quasispecies reflects also the adaptation of the virus to the host.

The hepatitis B surface antigen (HBsAg) is the most abundant circulating viral protein and acts as a major target for viral neutralization by the host immune response, either naturally or by vaccine-induced antibodies.[7,8] HBsAg contains an exposed major hydrophilic region (MHR; aa 99-169) which includes the 'a' determinant (aa 124-147) containing major epitopes for antibody recognition.[9,10] Neutralizing antibodies produced in response to HBV infection or after HBV vaccination are directed against a cluster of B-cell epitopes within the 'a' determinant which are shared among all HBV subtypes.[11] Antibodies against these epitopes are also used in diagnostic assays to detect HBsAg levels in acute or chronic HBV infection. As a consequence, the 'a' determinant is under constant selective pressure from the host immune response. The mutations occurring within the MHR, particularly in the 'a' determinant can allow HBsAg to escape the host immune response or vaccine-induced immunity or passive immunoglobulin therapy.[12,13]

Nucleic acid polymers (NAPs) are a new class of antiviral agent [14] which block the assembly and release of HBV subviral particles into the blood.[15,16] Recently, clinical studies have demonstrated that treatment with REP 2139 leads to the clearance of HBsAg in HBV mono-infection and HBV/HDV co-infection.[17,18] REP 2139-mediated HBsAg clearance when used in concert with immunotherapy has been shown to restore host immune control over HBV and HDV infection in a significant proportion of patients after withdrawal of therapy.[17,18] However, the molecular mechanisms of how NAPs interfere with subviral particle assembly and release are not yet well understood.

The number of quasispecies and their complexity reflects the error of reverse transcription during replication of HBV DNA and evolution of the virus due to the host immune response over time. We hypothesized that quasispecies diversity may influence the response to REP 2139-Ca. In addition, the aa sequence in the MHR of HBsAg is critical for the detection of HBsAg by various diagnostic assays, thus the deep sequencing analysis of the MHR in this study provides a more sensitive analysis of fluctuations in HBV QS during REP 2139-Ca therapy which might influence detection of HBsAg.

Deep sequencing analysis of the HBsAg fragment within the MHR of HBsAg spanning aa 76-200 was performed and genetic diversity of the HBsAg MHR was explored (a) to determine whether QS prevalence during REP 2139-Ca therapy is correlated with treatment response (HBsAg and HBV DNA reduction >1 log from baseline), (b) to investigate whether mutations in the 'a' determinant of HBsAg could play a role in escaping the host immune response or HBsAg escape from detection with diagnostic assays and (c) to investigate whether QS kinetics during REP 2139-Ca treatment could influence viral replication competence and subsequent treatment outcomes.

2 | MATERIAL AND METHODS

2.1 | Patients

Serum samples of all 12 patients with HBeAg-positive chronic HBV infection in the REP 102 trial [17] were obtained from the Farabi General Hospital (Dhaka, Bangladesh). For each patient, samples that spanned the entire course of REP 2139-Ca treatment; at the beginning (baseline/early REP 2139-Ca monotherapy), in the middle (end of REP 2139-Ca monotherapy) and towards the end of REP 2139-Ca therapy with immunotherapy) with sufficient HBV DNA levels were selected for deep sequencing analysis (see Table 1). Serum samples from four untreated patients (i.e., 3 HBeAg (+) and 1 HBeAg (−)) who match the demographics of patients with respect to age, sex and genotypes in the REP 102 trial signed the informed consent and were obtained from University Hospital Munich-Grosshadern (Munich, Germany) were used as matched controls. The pre-treatment clinical profiles of the patients included in this study are summarized in Table 2. All samples were stored and preserved at −80°C until use. All procedures performed in the REP 102 study were in accordance with the ethical principles of the responsible institutional ethics committee on human experimentation in Dhaka, Bangladesh (Viral Hepatitis Foundation of Bangladesh) and with the Helsinki Declaration of 1975, as revised in 2008.

2.2 | Extraction and amplification of HBV DNA

HBV DNA was extracted using the mSample® Preparation System DNA kit (Abbott) on m24sp extraction device (Abbott). For samples with high viral load, a single PCR using primers targeting MHR was performed. For samples with low viral load, a nested PCR using specific primers targeting the L-HBsAg gene was performed in first amplification,[19] followed by a second amplification with the MHR-specific primers (5′-GTCTGCGGCGTTTTATC-3′ and 5′-CAATACCACATCATCCATAT-3′). The PCR was carried out in a final volume of 50 μL containing 10 μL of 5 × Q5 Reaction Buffer, 2 μL of 5 mmol/L dNTP, 2.5 μL of each primer (10 μmol/L), 0.3 μL of Q5 Hot Start High-Fidelity DNA Polymerase (New England Biolabs), and 5 μL of extracted DNA for the single PCR and first round of nested PCR, or 3 μL of the first round PCR product for the second round of nested PCR. PCR amplification was performed on

TABLE 1 The weekly administration and pattern of HBV DNA and HBsAg responses to REP 2139-Ca therapy to patients under REP 102 study

Group	Patient (GT)	Sampling Point[a]		Therapy completed at sampling time point	HBV DNA (copies/mL)	HBsAg (IU/mL)
Non responders	1 (C)	1.	wk 1	Baseline	9.98E + 08	1.68E + 05
		2.	wk 27	26 wk of REP 2139-Ca	7.18E + 08	2.39E + 04
		3.	wk 40	39 wk of REP 2139-Ca including 1 wk of thymosin alpha 1	5.90E + 08	1.19E + 04
	5 (D)	1.	wk 2	1 wk of REP 2139-Ca	3.15E + 07	5.25E + 03
		2.	wk 23	22 wk of REP 2139-Ca	5.47E + 06	8.26E + 03
		3.	wk 34	33 wk of REP 2139-Ca + 2 wk REP 2139/ thymosin alpha 1	1.11E + 07	8.07E + 03
	10 (C)	1.	wk -2	baseline	9.89E + 08	>1.25E + 05
		2.	wk 11	10 wk of REP 2139-Ca	9.89E + 08	>1.25E + 05
		3.	wk 25	24 wk of REP 2139-Ca + 4 wk REP 2139/ pegIFN alpha	9.89E + 08	>1.25E + 05
Responders	2 (D)	1.	wk -2	baseline	6.23E + 08	5.23E + 04
		2.	wk 5	4 wk of REP 2139-Ca	4.48E + 07	3.02E + 04
		3.	wk 11	10 wk of REP 2139-Ca	4.04E + 04	1.12E + 03
	3 (C)	1.	wk -3	baseline	1.01E + 08	1.30E + 04
		2.	wk 6	5 wk of REP 2139-Ca	1.69E + 07	1.66E + 03
		3.	wk 17	16 wk of REP 2139-Ca	2.18E + 04	9.00E -02
	4 (A)	1.	wk-3	baseline	1.18E + 08	3.65E + 03
		2.	wk 13	12 wk of REP 2139-Ca	8.15E + 03	2.65E + 00
		3.	wk 27	26 wk of REP 2139-Ca	8.71E + 03	2.59E + 00
	6 (C)	1.	wk 4	baseline	1.29E + 08	4.76E + 04
		2.	wk 12	13 wk of REP 2139-Ca	6.96E + 05	3.13E + 03
		3.	wk 30	29 wk of REP 2139-Ca	1.19E + 04	6.88E + 01
	7 (C)	1.	wk 7	6 wk of REP 2139-Ca	9.89E + 08	1.68E + 04
		2.	wk 37	36 wk of REP 2139-Ca including 5 wk of thymosin alpha 1	5.74E + 06	2.50E + 02
		3.	wk 76[b]	58 wk of REP 2139-Ca including 13 wk of thymosin alpha 1 followed by 13 wk of pegIFN alpha 2a 17 wk of follow-up	4.57E + 03	5.76E + 00
	8 (D)	1	wk 2	1 wk of REP 2139-Ca	9.89E + 08	9.99E + 05
		2.	wk 11	10 wk of REP 2139-Ca	8.38E + 04	6.22E + 04
		3.	wk 20	19 wk of REP 2139-Ca	2.75E + 03	2.01E + 03
	9 (D)	1.	wk 2	1 wk of REP 2139-Ca	7.25E + 05	2.29E + 04
		2.	wk 25	24 wk of REP 2139-Ca	1.28E + 04	9.60E + 00
		3.	wk 80[b]	48 wk of REP 2139-Ca including 24 wk of pegIFN alpha 2a 22 wk of pegIFN alpha 2a 10 wk of follow-up	8.90E + 03	3.26E + 01
	11 (C)	1.	wk -2	baseline	9.89E + 08	8.57E + 04
		2.	wk 8	7 wk of REP 2139-Ca	3.55E + 07	1.60E + 02
		3.	wk 13	12 wk of REP 2139-Ca	~1.00E + 04	8.00E -02
	12 (C)	1.	wk -1	baseline	9.09E + 06	1.29E + 03
		2.	wk 12	11 wk of REP 2139-Ca	8.77E + 03	0.01E + 00
		3.	wk 78[b]	33 wk of REP 2139-Ca including 13 wk of pegIFN alpha 2a 43 wk of follow-up	2.15E + 03	7.24E + 00

Note: GT corresponds to patient genotype.
[a]week in relation to the start of REP 2139-Ca therapy (treatment was given after blood collection for serum samples).
[b]taken during follow-up after removal of all therapy.

GeneAmp® PCR System 9700 (Applied Biosystems). The amplification products were analysed by electrophoresis on a 2% agarose gel stained with ethidium bromide and detected using ultraviolet light.

2.3 | Deep sequencing of HBV DNA

All amplicons were purified using the Agencourt AMPure®XP system on a BioMek NX workstation (Beckman Coulter) and quantified on a FluoStar Optima (BMG Labtech) using Quant-iT Picogreen dsDNA reagent (Life technologies). Library preparation for Illumina deep sequencing was done using TruSeq Nano DNA Library kit (Illumina Inc.) according to the manufacturer's instructions. Resulting libraries were normalized and pooled for subsequent sequencing on an Illumina MiSeq platform using the 2 × 250 cycle paired-end sequencing protocol.

2.4 | Short read sequence accession numbers

The raw reads were submitted to NCBI as Short Read Archive (SRA ID: SRP154287) and are accessible through NCBI BioProject PRJNA478149.

2.5 | Sequence read mapping and data analysis

Illumina deep sequencing results were subjected to quality control and data trimming using FastQC v0.10.1[20]and CutAdapt v1.9.1[21] to remove adapters and perform quality clipping from either 5′ or 3′ end of the reads (PHRED threshold <20). The reads shorter than 20 bp and with three or more base calls (N) were discarded. The remaining reads were mapped against the reference sequence (GenBank accession no. NC_003977.2) of genotype D strain ayw from the NCBI reference sequence (Refseq) database[22] using the BWA-mem (Burrows-Wheeler Aligner) algorithm.[23] The alignment file was converted, sorted and indexed using Samtools v0.1.19.[24] Coverage and sequencing depth were assessed using Bedtools v2.17.0.[25] For each sample, consensus sequences were generated using mpileup Samtools v0.1.19.[24]

For each patient, genotype source was identified by performing sequence similarity search using BLAST (Basic local alignment search tool) [26] against the NCBI nonredundant sequence database.[22] The genotype-specific amino acid substitutions observed between reference D ayw strain (GenBank accession no. NC_003977.2) against the isolates of genotype A1 (GenBank accession no. KP234050),[27] C1 (GenBank accession no. KM999990) [28] and D2 (GenBank accession no. GQ477452) [29] were estimated by performing pairwise alignment (Figure S1).

2.6 | Quasispecies reconstruction

Viral quasispecies were reconstructed using QuasiRecomb 1.2 [30] with flag '–r 227-602' to specify the region of interest. The flag '–conservative' was not employed because our interest was to consider minor variants. To reflect the nature of quasispecies variants in the MHR in vivo as accurately as possible in relation to clinical diagnosis by distinguishing the true quasispecies variants from sequencing errors, we analysed the variants detected in ≥1% of the viral population. A similar cut-off was applied in previous studies.[31,32]

2.7 | HBsAg quasispecies sequence analysis

Viral quasispecies heterogeneity was evaluated using two parameters (complexity and diversity). Quasispecies complexity can be defined in terms of probabilities of different sequences that can appear at a given time point within a population and is expressed as normalized Shannon entropy (Sn). Quasispecies complexity was assessed at nucleotide level using the formula:

$$Sn = \sum i\varepsilon_{[A,T,C,G,-]}(p_i/\ln p_i)/\ln N$$

where p_i represents the frequency of a particular nucleotide or deletion i at this position, and N is the total number of species within a population.[3,33] Sn varied from 0 (no diversity) to 1 (maximum diversity).

Quasispecies diversity was evaluated at nucleotide and amino acid level by three parameters: the mean genetic distances (d), the number of synonymous substitutions per synonymous site (dS) and the number of nonsynonymous substitutions per nonsynonymous site (dN). Estimation of d was conducted using program DNADIST (Kimura-2-parameter setting) and PROTDIST (Jones-Thornton-Taylor setting) in PHYLIP Package v3.572.[34] The modified Nei-Gojobori model (Jukes-Cantor correction) was applied for dS and dN calculation. Additionally, nucleotide changes among the quasispecies belonging to responder, nonresponder and control were recorded and mutation type (transition or transversion) were determined using Kimura 2-parameter.

2.8 | Statistical analysis

Differences between the groups are expressed as the mean ± SD using a t- test for the variables showing a normal distribution. A P-value of < .05 was considered statistically significant.

3 | RESULTS

3.1 | Patient demographics

In the REP 102 study, twelve HBeAg-positive chronic hepatitis B patients received REP 2139-Ca therapy for a period of 32-58 weeks (Table 1). Nine patients who experienced >2 log HBsAg and HBV DNA reductions at baseline during REP 2139-Ca monotherapy were eligible to receive add-on immunotherapy. Three patients experienced HBV DNA and HBsAg reductions <1 log during 24-40 weeks of REP 2139-Ca monotherapy and were considered nonresponders and entered into follow-up with entecavir without receiving immunotherapy. The demographic and clinical features of REP 2139-Ca treated patients and matched control patients are shown in Table 2.

TABLE 2 Baseline characteristics and comparison of genetic diversity in hepatitis B virus infected patients in the REP 102 protocol and matched controls

Factor	Responders (n = 9)	Nonresponders (n = 3)	Matched controls (n = 4)	*P*-value Responders vs Nonresponders
Baseline characteristics				
Age in y (Mean ± SD)	22.22 ± 3.19	23.33 ± 4.61	25.75 ± 3.77	0.64
Sex (M: F)	7:2	3:0	3:1	0.41
Genotype	C (5), D (3), A (1)	C (2), D (1)	C (1), D (3)	
AST (Mean ± SD) (IU/L)	43.77 ± 12.04	34.33 ± 16.29	-	0.25
ALT (mean ± SD) (IU/L)	72.67 ± 32.09	40.33 ± 8.38	-	0.12
HBsAg (Mean ± SD) (IU/L)	138 026.67 ± 324 012.91	99 733.33 ± 84 724.41	10 037 ± 4483	0.84
HBV DNA (mean ± SD) (log copies/mL)	8.09 ± 1.07	8.50 ± 0.86	3.72 ± 3.39	0.48
Coverage (Mean ± SD)	201 022 ± 114 613	278 307 ± 69 252	186 188 ± 16 651	0.31
Quasispecies diversity				
No of Quasispecies	1-10	1-6	1-5	0.292
Transitions (Mean ± SD)	2.11 ± 1.45	0.33 ± 0.577	0.5 ± 0.57	0.474
Transversions (Mean ± SD)	2.77 ± 1.71	2.33 ± 2.30	1.5 ± 0.57	0.447
HBV DNA decline during therapy (log10 copies/mL)	4.16 ± 1.10	1.89 ± 2.89	0.56 ± 0.96[a]	0.062
Genetic distance *d* (Mean ± SD)				
Nucleotide level	0.0485 ± 0.071	0.0138 ± 0.013	0.0181 ± 0.0317	0.191
Amino acid level	0.1392 ± 0.215	0.0416 ± 0.040	0.0198 ± 0.0397	0.251
dN	0.00613 ± 0.008	0.00319 ± 0.0001	0.0023 ± 0.003	0.352
dS	0.00125 ± 0.003	0.00033 ± 0.00057	0.0156 ± 0.016	0.588

[a]HBV DNA decline observed in corresponding two samples in control patients.

In total, 43 samples from 16 patients (12 REP 102 patients and four matched control patients) were sequenced and analysed. The average mean read lengths after pre-processing procedures were 246 ± 1.6 bp (mean ± sd), ranging from 20-249 bases. A total of 16 189 811 reads were aligned against the HBV reference (GenBank Accession no. NC_003977.2) with a mean coverage depth per nucleotide after passing quality control (Q20) of 201 022 for responders, 278 307 for nonresponders and 179 226 for matched controls (Table 2). Several subgenotype specific amino acids identified for genotype A1,[27] C1 [28] and D2 [29] were not considered as actual mutations (Figure S1).

3.2 | Deep sequencing analysis of the MHR of HBsAg

QuasiRecomb detected different viral QS occurring at frequencies lower than <1% among responders and nonresponders (Table S1). To characterize the HBV population in vivo as accurately as possible, only quasispecies (QS) with prevalence of ≥1% (i.e., eliminating sequencing errors) in responders, nonresponders and controls across different time points were used for analysis (Figure 1, Table S1).

3.3 | HBV quasispecies complexity

The kinetics of QS complexity during REP 2139-Ca therapy in the REP 102 study was determined for responders, nonresponders and matched controls. Quasispecies complexity for all patients between time points 1 to 3 in REP 102 patients (mean 273 days) and time points 1 to 2 in matched control patients (mean 557 was determined by Shannon entropy (Table 2). At baseline, HBV QS complexity varied slightly between responders and nonresponders (Figure 2) but the difference was not statistically significant (0.0162 vs 0.0120, P = .28). A statistically significant decline in viral complexity during treatment in responders was observed (0.0162-0.0081, P = .001, Figure 2) but no significant change in complexity was observed among nonresponders (0.0120 vs 0.0122, P = .99, Figure 2).

In responder patients, QS complexity decreased over time except for two patients with transient increases in QS complexity at time point 2 which declined to below baseline by time point 3 (Figure 1A). Based on the observed numbers of quasispecies at baseline among the nine responders, we arbitrary created two groups. Among the five responder patients with 4-10 quasispecies at baseline, the number of quasispecies increased in three patients at the second time point analysed (during REP 2139-Ca monotherapy) and decreased as REP 2139-Ca treatment progressed to additional immunotherapy (Figure 1A). On the contrary, 4/9 responders with less than four quasispecies at baseline experienced no increase in the number of quasispecies during REP 2139-Ca treatment (Figure 1A).

Additionally, we observed that the QS frequencies in responders fluctuate at different time points exhibiting three different trends

(Figure 3A-I). (a) In most cases, a dominant quasispecies present before the onset of treatment decreased in its frequency in the second sample during treatment with an emergence of minor and intermediate quasispecies (5%-20%). These dominant quasispecies increased again in its frequency (>88%-100%) at the longest exposure of REP 2139-Ca (Figure 3A,3,3,3,G and I). (b) In some cases, the elimination of quasispecies from the viral population occurred concomitantly with rise of new quasispecies (Figure 3B,E), and (c) in one patient, no change in the viral population was observed suggesting fixation of that quasispecies (Figure 3H).

Quasispecies complexity increased over time in 1/3 nonresponders where as the remaining two nonresponders remained stationary throughout the course of therapy (Figure 1B, Figure 3J-L). Of 3/4 untreated matched controls, the samples recorded at two consecutive time points showed up to six quasispecies sequences at the beginning were survived by only one dominant quasispecies at later time point (Figure 1C, Figure 3M-O). Conversely, 1/4 untreated matched controls (i.e., patient 13) recorded at a single time point showed two quasispecies (Figure 1C). In 7/9 responder patients, declines in QS complexity were associated with declines in HBV DNA and/or detection of anti-HBs (Figures S7 and S8).

3.4 | HBV quasispecies diversity

Quasispecies diversity was assessed by comparing the mean Kimura 2-parameter distances between time points T1-T2 and T2-T3 throughout REP 2139-Ca therapy for responders, nonresponders and matched controls. The responders showed significantly higher genetic mean distance (*d*) as compared to nonresponders both at nucleotide level (0.0485 vs 0.0138 substitutions/site) and amino acid level (0.1392 vs 0.0416 substitutions/site; Table 2).

Furthermore, responders showed significantly higher dN and dS than nonresponders (dN 0.00613 vs 0.0031; dS 0.00125 vs 0.00033) and matched controls (dN 0.00613 vs 0.0023; dS 0.001255 vs 0.0156). The dN values were higher than dS values in responders than nonresponders, suggesting nonsynonymous substitutions were more predominant mutations among responders (Table 2).

The nucleotide substitutions in form of transitions and transversions were significantly higher among responders as compared to nonresponders and matched controls. In addition, transversions were more frequent than the transitions in responders as compared to the nonresponders (Table 2).

3.5 | Amino acid variation in HBsAg quasispecies

The total QS population for responders and nonresponders were divided into three groups: major population >20%; intermediate population 5% to <20%; and minor population 1% to <5%. The amino acid substitutions occurring within the HBsAg were segmented into three regions: pre-'a' determinant, 'a' determinant and post-'a' determinant (Table 3). In responders, we observed significantly high number of amino acid substitutions (n = 30) occurring outside the 'a' determinant in the major, intermediate and minor population groups (Table 3). Additionally, we found few amino acid substitutions within the 'a' determinant (T126S, T126I, G130R, D144E and G145R) among the intermediate and minor QS populations (Table 3). None of these amino acid mutations within the 'a' determinant were observed in major population of HBV quasispecies.

Several mutations were observed independently across all responder patients (Figure S2-S6). In responders, T126S/I substitutions were observed in patients 102-3 and 102-12, whereas mutations at position 109 (P109L and L109Q) were found in patients

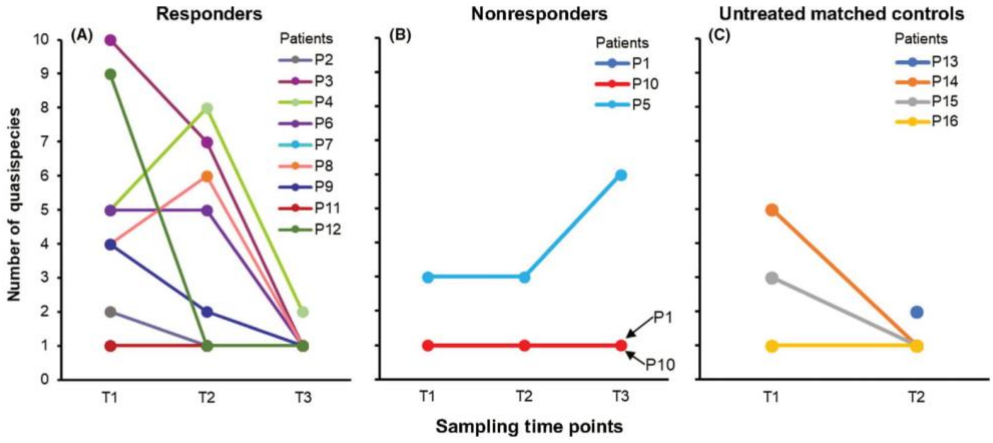

FIGURE 1 Kinetics of HBV quasispecies during REP 2139-Ca therapy. The number of quasispecies among (A) responders, (B) nonresponders and (C) matched controls are presented

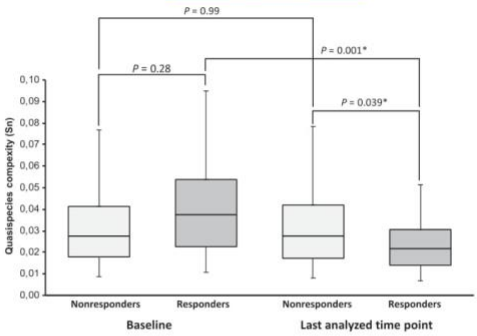

FIGURE 2 Position entropy plot showing quasispecies complexity in responders and nonresponders at baseline and end of REP 2139 therapy

102-3, 102-4 and 102-8. Similarly, amino acid substitution T115N was found in 102-4 and 102-8 patients. In nonresponders, only patient 102-5 reported six amino acid substitutions P88L, P120T, T126I, F170S, L173P and T189I; 5/6 detected in 1% to 5% of the total QS population (Table 3).

Several immune escape mutations within 'a' determinant described previously [35] were found only within the minor and intermediate quasispecies population of the responders include the following: genotype C (K122R, T126S, T126I, M103I, G145R and P120T) and genotype D (D144E).

4 | DISCUSSION

The treatment of chronic hepatitis B with REP 2139-Ca in combination with pegIFN in the REP 102 study showed promising results with reduced levels or complete loss of HBsAg and decline in HBV DNA. Seroconversion to anti-HBs was observed in 8/9 responders with reductions or loss of HBsAg levels indicating functional control of HBV infection.

We analysed evolutionary patterns of hepatitis B virus quasispecies at three different time points during REP 2139-Ca therapy. We observed mutations (i.e at the nucleotide and amino acid level) within the MHR of S region with frequencies that changed during the course of REP 2139-Ca therapy among responders as well as nonresponders, and controls matched to patient demographics of the REP 102 study.

Among the responders, 4/9 participants showed significantly higher genetic heterogeneity of quasispecies (QS) complexity, suggesting no virus adaption to the host was occurring in these patients. The change in viral QS complexity and diversity in some participants responding to REP 2139-Ca therapy may indicate selective targeting of viral QS by host immune response during profound declines in HBV DNA and HBsAg. In contrast, the pattern of QS complexity during REP 2139-Ca therapy among the nonresponders remained mostly unchanged. Consistent with this hypothesis, our analysis revealed that declines in QS complexity observed in responder patients were associated with declines in HBV DNA and detection of anti-HBs. Moreover, appearance of mutations during REP 2139 therapy is also likely driven by the removal of certain QS leading to the alteration of prevalence of other QS still present, and allowing their detection. Significant elevations in anti-HBs concentrations did not accompany QS decline in three responder patients (Figure S8A,E,F). In these cases, anti-HBs may still have been sequestered by reactive virions so free anti-HBs levels were not yet meaningful.

The emergence of QS during chronic HBV infection may also be a result of selective pressure of antiviral treatment. These QS may have altered epitopes within the MHR (particularly within the 'a' determinant) which are not recognized by virus-specific B cells or T cells present in the host. A wide range of amino acid substitutions occurring within the 'a' determinant of HBsAg have been described escaping viral neutralization by anti-HBs and interferes with HBsAg/anti-HBs binding in diagnostic assays.[7,36] However, our study revealed that most amino acid mutations observed in responders and nonresponders are found outside the 'a' determinant region (i.e., 76-123 aa and 148-200 aa) at varying time points after initiation of therapy. These mutations were not correlated with the changes in HBsAg serum levels, indicating that the HBsAg reductions observed cannot be attributed to changes in the 'a' determinant affecting the antigenic conformation as a result of structural changes that alter antibody binding and subsequent performance of diagnostic assays for HBsAg detection.

No specific mutation pattern was observed among the responders versus nonresponders that would suggest selective pressure potentially driven by a direct interaction of REP 2139 with the HBsAg and subsequent inhibition of HBsAg release from infected hepatocytes. These observations are consistent with the recent reports that REP 2139 does not directly interact with the envelope protein of HBV or HDV [37] or with HBsAg.[18] Moreover, our results are in agreement with our companion study analysing the entire pre-S/S ORF of HBV in the REP 102 study and revealed that no mutations were emerging during REP 2139-Ca which could effect the HBsAg production, secretion or detection.[19]

While there is evidence to suggest that mutations in the basic core promoter (BCP) may play a role in the weaker effect of interferon in HBeAg-negative patients,[38,39] the secretion of HBsAg is not affected by these BCP mutations. Moreover, no alteration over the entire HBsAg ORF from our previous study [19] or alteration in quasispecies in the MHR of the HBsAg examined in the current study were observed after the switch from REP 2139-Ca monotherapy therapy to combination therapy with REP 2139-Ca and either pegIFN or thymosin alpha-1. Although the analysis of HBsAg mutations with pegIFN monotherapy is outside the scope of this study, the data presented herein indicate that addition of pegIFN or thymosin alpha-1 does not alter QS species present in the HBsAg MHR or the mutation status throughout the entire HBsAg ORF as shown in our previous study.[19]

Additionally, in other studies QS dynamics under treatment with NUC's showed that high QS complexity in the precore region

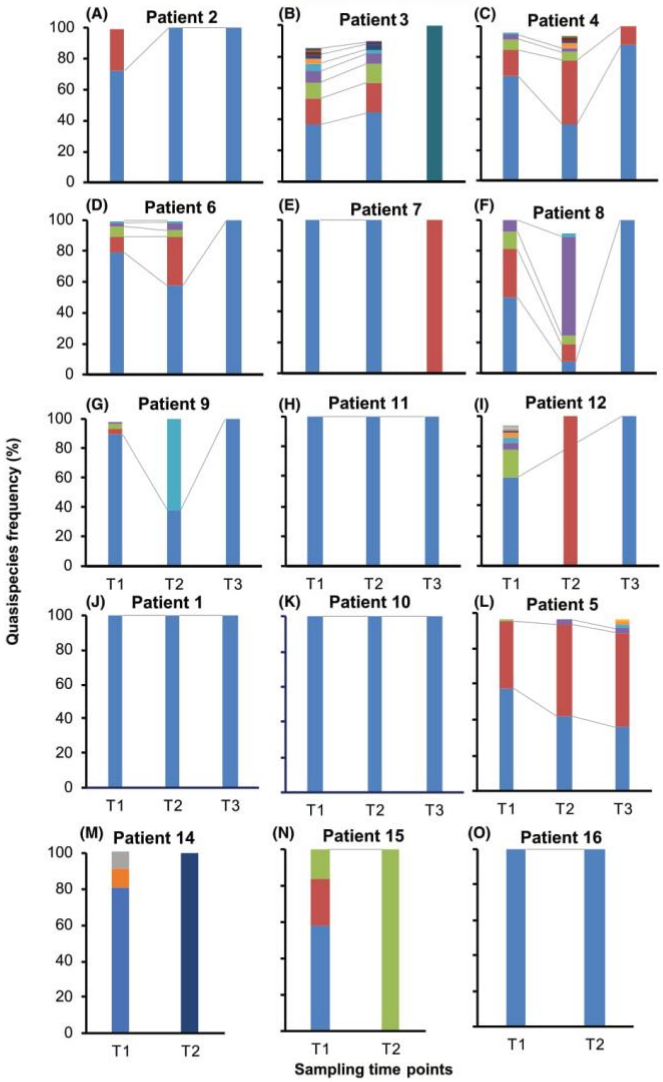

FIGURE 3 Kinetics of individual quasispecies frequency for responders (A-I), nonresponders (J, K, L) and matched control patients (M, N, O) over assessed time points

is associated with low viral replication and suggests active immune response in HBeAg-negative patients.[40] In a recent study,[41] the frequency of viral QS increased significantly in different genomic regions (Polymerase, S region) when compared before and during the NUC treatment. The variations in the polymerase region particularly within the RT1 motif region were greater during early period of NUC treatment than before treatment. It was suggested that the variations in the small S region which overlaps with RT region may inhibit the RT activity and synthesis and suppress the secretion of HBsAg.

In summary, our analysis indicates that treatment of chronic hepatitis B with REP 2139-Ca does exerts selective pressure on the existing populations of HBsAg QS within the MHR in patients responding to treatment versus the nonresponders. We have no indication that the antiviral effects of REP 2139 in chronically infected patients are achieved by directly targeting HBsAg clearance from the blood. However, we suggest that REP 2139 NAPs may somehow target one or more host cellular proteins involved in viral replication, assembly or release of subviral particles. Moreover, HBsAg reductions observed during REP

TABLE 3 Mutations observed in the major (>20%), intermediate (5%-20%) and minor (1%-5%) viral populations

Amino acid Position in HBsAg	Mutation	No of patients with mutations					
		Major population		Intermediate population		Minor population	
		R	NR	R	NR	R	NR
Pre-a-determinant (76-123)	F85C					1	
	L88P		1		1		1
	L98V	**1**		**1**		**1**	
	D99G					1	
	Q101K			1		1	
	Q101R			1		1	
	M103I	**1**		**1**		**1**	
	V106F					1	
	P108H					1	
	P109L	**1**		**1**		**1**	
	L109Q			2		1	
	G112R	**1**		**1**		**1**	
	T115N	**1**		**2**		**1**	
	G119R	**1**		**1**		**1**	
	P120T	1					1
	K122R	**1**		**1**		**1**	
a-determinant (124-147)	T126S			1		1	
	T126I			1		1	1
	D144E					1	
	G145R			1		1	
Post-a-determinant (148-200)	P153T	1					
	A159V	1					
	F161Y	**1**		**1**		**1**	
	V168A					1	
	F170S			1		1	1
	L173P						1
	G185E					1	
	P188L					1	
	P188H					1	
	T189I	1		1		1	1

Note: In bold are the mutations found in responders in all three of the viral population.
Abbreviations: R, responders (n = 9), NR, nonresponders (n = 3).

2139-Ca therapy are not a result of impaired detection of HBsAg by standard diagnostic assays. We suggest to evaluate the kinetics of quasispecies in a larger cohort of HBV patients receiving antiviral treatment to determine whether deep sequencing in the future may serve as a useful tool to predict responsiveness to treatment with NAP.

ACKNOWLEDGEMENTS

ZU was supported by grants from the Deutscher Akademischer Austauschdienst (DAAD; 57129429) and Replicor Inc. The authors thank Judith Seebach for her technical support.

CONFLICT OF INTEREST

A conflicting interest exists when professional judgement concerning a primary interest (such as patient's welfare or the validity of research) may be influenced by a secondary interest (such as financial gain or personal rivalry). It may arise for the authors when they have financial interest that may influence their interpretation of their results or those of others. Examples of potential conflicts of interest include employment, consultancies, stock ownership, honoraria, paid expert testimony, patent applications/registrations, and grants or other funding.

AUTHOR CONTRIBUTION

HK, HM and MD designed and performed the experiments; ZU and DF performed the bioinformatics and statistical analyses; M-AM, AV and MB enrolled patients and performed clinical evaluations; MB, MR and AV contributed with the study concept and interpreted the data; ZU, HM, HK, MR and AV wrote the manuscript.

ETHICAL APPROVAL

All procedures followed in the REP 102 studies were in accordance with the ethical standards of the responsible local institutional ethics committee on human experimentation in Dhaka, Bangladesh (Viral Hepatitis Foundation of Bangladesh) and with the Helsinki Declaration of 1975, as revised in 2008. REP 102 study protocols and amendments were approved by the Viral Hepatitis Foundation of Bangladesh.

RANDOMIZED CONTROLLED TRIAL

REP 102 (NCT02646189—https://clinicaltrials.gov/ct2/show/NCT02646189) studies were registered retroactively on www.clincaltrials.gov prior to publication as trial registration in a public database.

ORCID

Michael Roggendorf https://orcid.org/0000-0003-3293-6738

REFERENCES

1. Polaris Observatory Collaborators. Global prevalence, treatment, and prevention of hepatitis B virus infection in 2016: a modelling study. *Lancet Gastroenterol Hepatol*. 2018;3(6):383-403.
2. WHO. Hepatitis B fact sheet updated: World Health Organization. 2017.
3. Domingo E, Martin V, Perales C, Grande-Perez A, Garcia-Arriaza J, Arias A. Viruses as quasispecies: biological implications. *Curr Top Microbiol Immunol*. 2006;299:51-82.
4. Fishman SL, Branch AD. The quasispecies nature and biological implications of the hepatitis C virus. *Infect Genet Evol*. 2009;9(6):1158-1167.
5. Margeridon-Thermet S, Shulman N, Ahmed A, 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*. 2009;199(9):1275-1285.
6. Toh ST, Jin Y, Liu L, et al. Deep sequencing of the hepatitis B virus in hepatocellular carcinoma patients reveals enriched integration events, structural alterations and sequence variations. *Carcinogenesis*. 2013;34(4):787-798.
7. Carman WF. The clinical significance of surface antigen variants of hepatitis B virus. *J Viral Hepat*. 1997;4(Suppl 1):11-20.
8. Kondo Y, Ninomiya M, Kakazu E, Kimura O, Shimosegawa T. Hepatitis B surface antigen could contribute to the immunopathogenesis of hepatitis B virus infection. *ISRN Gastroenterol*. 2013;2013:935295.
9. Chotiyaputta W, Lok AS. Hepatitis B virus variants. *Nat Rev Gastroenterol Hepatol*. 2009;6(8):453-462.
10. Kay A, Zoulim F. Hepatitis B virus genetic variability and evolution. *Virus Res*. 2007;127(2):164-176.
11. Degertekin B, Han S-H, Keeffe EB, et al. Impact of virologic breakthrough and HBIG regimen on hepatitis B recurrence after liver transplantation. *Am J Transplant*. 2010;10(8):1823-1833.
12. Kalinina T, Iwanski A, Will H, Sterneck M. Deficiency in virion secretion and decreased stability of the hepatitis B virus immune escape mutant G145R. *Hepatology*. 2003;38(5):1274-1281.
13. Lok A, McMahon BJ, Brown RS, et al. Antiviral therapy for chronic hepatitis B viral infection in adults: a systematic review and meta-analysis. *Hepatology*. 2016;63(1):284-306.
14. Vaillant A. REP 2139: antiviral mechanisms and applications in achieving functional control of HBV and HDV infection. *ACS Infect Dis*. 2019;5(5):675-687.
15. Noordeen F, Scougall CA, Grosse A, et al. Therapeutic antiviral effect of the nucleic acid polymer REP 2055 against persistent duck hepatitis B virus infection. *PLoS ONE*. 2015;10(11):e0140909.
16. Quinet J, Jamard C, Burtin M, et al. Nucleic acid polymer REP 2139 and nucleos (T)ide analogues act synergistically against chronic hepadnaviral infection in vivo in Pekin ducks. *Hepatology*. 2018;67(6):2127-2140.
17. Al-Mahtab M, Bazinet M, Vaillant A. Safety and efficacy of nucleic acid polymers in monotherapy and combined with immunotherapy in treatment-naive Bangladeshi patients with HBeAg+ chronic hepatitis B infection. *PLoS ONE*. 2016;11(6):e0156667.
18. Bazinet M, Pântea V, Cebotarescu V, et al. Safety and efficacy of REP 2139 and pegylated interferon alfa-2a for treatment-naive patients with chronic hepatitis B virus and hepatitis D virus co-infection (REP 301 and REP 301-LTF): a non-randomised, open-label, phase 2 trial. *Lancet Gastroenterol Hepatol*. 2017;2(12):877-889.
19. Mijočević H, Karimzadeh H, Seebach J, et al. Variants of hepatitis B virus surface antigen observed during therapy with nucleic acid polymer REP 2139-Ca have no influence on treatment outcome and its detection by diagnostic assays. *J Viral Hepat*. 2019;26(4):485-495.
20. Brown J, Pirrung M, McCue LA. FQC Dashboard: integrates FastQC results into a web-based, interactive, and extensible FASTQ quality control tool. *Bioinformatics*. 2017.
21. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J*. 2011;17:10-12.
22. Pruitt KD, Tatusova T, Maglott DR. NCBIReference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res*. 2005;33(Database issue): D501-D504.
23. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. 2013.
24. Li H, Handsaker B, Wysoker A, et al. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25(16):2078-2079.
25. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26(6):841-842.
26. Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL. NCBI BLAST: a better web interface. *Nucleic Acids Res*. 2008;36(Web Server issue): W5-W9.
27. Cai Q, Zhu H, Zhang Y, Li X, Zhang Z. Hepatitis B virus genotype A: design of reference sequences for sub-genotypes. *Virus Genes*. 2016;52(3):325-333.
28. Zhu HL, Wang CT, Xia JB, Li X, Zhang ZH. Establishment of reference sequences of hepatitis B virus genotype C subgenotypes. *Genet Mol Res*. 2015;14(4):16521-16534.
29. Spitz N, Moreira AS, Mello F, Franz H, Gomes SA, Araujo NM. Complete genome sequence of a hepatitis B virus isolate of genotype D2, subtype adrq+, from Brazil. *Arch Virol*. 2017;162(6):1695-1700.
30. Topfer A, Zagordi O, Prabhakaran S, Roth V, Halperin E, Beerenwinkel N. Probabilistic inference of viral quasispecies subject to recombination. *J Comput Biol*. 2013;20(2):113-123.
31. Bolcic F, Sede M, Moretti F, et al. Analysis of the PKR-eIF2alpha phosphorylation homology domain (PePHD) of hepatitis C virus

genotype 1 in HIV-coinfected patients by ultra-deep pyrosequencing and its relationship to responses to pegylated interferon-ribavirin treatment. *Arch Virol*. 2012;157(4):703-711.
32. Eriksson N, Pachter L, Mitsuya Y, et al. Viral population estimation using pyrosequencing. *PLoS Comput Biol*. 2008;4(4):e1000074.
33. Nishijima N, Marusawa H, Ueda Y, et al. Dynamics of hepatitis B virus quasispecies in association with nucleos(t)ide analogue treatment determined by ultra-deep sequencing. *PLoS ONE*. 2012;7(4):e35052.
34. Felsenstein J. PHYLIP (Phylogeny Inference Package) version 3.6. *Cladistics*. 2005.
35. Coppola N, Onorato L, Minichini C, et al. Clinical significance of hepatitis B surface antigen mutants. *World J Hepatol*. 2015;7(27):2729-2739.
36. Sheldon J, Soriano V. Hepatitis B virus escape mutants induced by antiviral therapy. *J Antimicrob Chemother*. 2008;61(4):766-768.
37. Beilstein F, Blanchet M, Vaillant A, Sureau C. Nucleic acid polymers are active against hepatitis delta virus infection in vitro. *J Virol*. 2018, 92(4). e01416-e1417.
38. Erhardt A, Reineke U, Blondin D, et al. Mutations of the core promoter and response to interferon treatment in chronic replicative hepatitis B. *Hepatology*. 2000;31(3):716-725.
39. Wang Y, Wei L, Jiang D, et al. In vitro resistance to interferon-alpha of hepatitis B virus with basic core promoter double mutation. *Antiviral Res*. 2007;75(2):139-145.
40. Homs M, Caballero A, Gregori J, et al. Clinical application of estimating hepatitis B virus quasispecies complexity by massive sequencing: correlation between natural evolution and on-treatment evolution. *PLoS ONE*. 2014;9(11):e112306.
41. Liang Y, Yano Y, Putri W, et al. Early changes in quasispecies variant after antiviral therapy for chronic hepatitis B. *Mol Med Rep*. 2018;17(4):5528-5537.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Usman Z, Mijočević H, Karimzadeh H, et al. Kinetics of hepatitis B surface antigen quasispecies during REP 2139-Ca therapy in HBeAg-positive chronic HBV infection. *J Viral Hepat*. 2019;26:1454–1464. https://doi.org/10.1111/jvh.13180

B Research Article II

Received: 6 August 2018 | Accepted: 24 January 2019

DOI: 10.1111/jvh.13086

ORIGINAL ARTICLE

WILEY JVH JOURNAL OF VIRAL HEPATITIS

Genetic diversity of hepatitis D virus genotype-1 in Europe allows classification into subtypes

Hadi Karimzadeh[1,2] | Zainab Usman[3] | Dmitrij Frishman[3,4] | Michael Roggendorf[1]

[1]Institute of Virology, Technische Universität München, Munich, Germany

[2]Department of Medicine II, University Hospital, LMU, Munich, Germany

[3]Department of Bioinformatics, Wissenschaftszentrum Weihenstephan, Technische Universität München, Freising, Germany

[4]Laboratory of Bioinformatics, RASA Research Center, St. Petersburg State Polytechnical University, Saint Petersburg, Russia

Correspondence
Michael Roggendorf, Institute of Virology, Technische Universität München, Munich, Germany.
Email: michael.roggendorf@tum.de

Funding information
Medical Faculty of the University of Duisburg-Essen/Technical University of Munich; European Association for the Study of the Liver (EASL); German Academic Exchange Service (DAAD)

Summary

Hepatitis delta virus (HDV) is an RNA virus which leads to both acute and chronic forms of hepatitis. At present, HDV isolates have been classified into eight major genotypes distributed over different geographical regions. Recent increase in HDV sequences in Europe and worldwide has enabled us to revisit the taxonomic classification of HDV. A total of 116 large hepatitis delta antigen (L-HDAg) nucleotide sequences and 13 full-length HDV genome sequences belonging to genotype-1 from our European cohort, as well as 621 L-HDAg nucleotide sequences belonging to genotype-1 to genotype-8 retrieved from the GenBank NCBI were included in this study. All 116 isolates of our cohort and 341 of 621 isolates (60%) account for genotype-1, while the remaining 40% of isolates were unevenly distributed across genotype-2 to genotype-8. Phylogenetic analysis of 98 L-HDAg sequences selected after elimination of redundant sequences of all 737 isolates was performed to identify plausible subtypes within HDV genotype-1. Pairwise genetic distances for L-HDAg sequences were calculated to estimate the inter-genotype and inter-subtype differences. The HDV genotype-1 isolates phylogenetically formed five distinct clusters (genotype 1a-1e), each of them corresponding to a distinct geographic region. Two distinct subtypes for HDV genotype-2 and -4 (ie -2a and -2b; -4a and -4b, respectively) could be identified based on isolate sequences from GenBank. The previously defined genotype-1 to genotype-8 have an inter-genotypic difference of ≥10%, while the newly defined subtypes of genotype-1, -2 and -4 show an inter-subtype difference of ≥3% to <10% from the average diversity. In addition, we identified unique amino acid residues, known as specificity-determining positions, amongst the proposed subtypes.

KEYWORDS
amino acid signatures, hepatitis delta, large hepatitis delta antigen (L-HDAg), subtyping

Abbreviations: AAS, amino acid signatures; CD-HIT, Cluster Database at High Identity with Tolerance method; HDV, hepatitis delta virus; L-HDAg, large hepatitis delta antigen; MI, mutual information; SDPs, specificity-determining positions; S-HDAg, small hepatitis delta antigen.

1 | INTRODUCTION

Hepatitis delta virus (HDV) is a defective RNA virus (36 nm), which depends on hepatitis B virus envelope protein (HBsAg) for the assembly of new infectious virions. There are two types of HDV infection: (a) Simultaneous infection with both HDV and HBV, clinically indistinguishable from HBV mono-infection alone; the recovery rate in adults is about 95%.[1] (b) Superinfection with HDV in HBV-infected individuals results in overt acute hepatitis infection (about 90% of patients) and may rapidly lead to liver cirrhosis and hepatocellular carcinoma.[2]

The prevalence of HDV infection varies greatly across different geographical regions. An estimated 15 to 20 million individuals have been infected with HDV worldwide.[3-5] To date, HDV strains isolated worldwide have been classified into eight genotypes (HDV-1 to HDV-8). HDV-1 is the most prevalent and geographically widespread[6] genotype distributed across Europe, Middle East, East Asia, America and Africa, whereas all other genotypes are associated with distinct geographical and ethnic regions. HDV-2 and HDV-4 are found in East Asia[7-9]; HDV-3 isolates are exclusively found in South America (Amazon Basin of Brazil, Peru and Venezuela)[10-13]; and HDV-5 to HDV-8 prevail amongst the individuals of African origin.[14-18] HDV infections with different genotypes exhibit different clinical course and outcomes: HDV-1 strains show a broad spectrum of virulent and pathogenic phenotypes, HDV-2 and HDV-4 cause milder forms of liver disease, HDV-3 isolates are associated with outbreaks of fulminant hepatitis in South America. The pathogenic properties of HDV 5-8 isolates are not well understood.[19]

Preliminary sequence comparison of 116 large hepatitis delta antigen (L-HDAg) as well as 13 full-length HDV genotype-1 isolates from our European cohort in different medical centres indicated possible existence of specific geographic clusters within this genotype of HDV. In the present study, we therefore revisit the taxonomic classification of HDV into genotypes and investigate potential subtypes within genotype-1 by performing phylogenetic analysis of 116 L-HDAg sequences of our European cohort together with 621 published sequences from the GenBank database. Phylogenetic analysis of the L-HDAg was performed after reducing the redundant sequences. Moreover, with a set of 13 complete genome sequences of our cohort and 76 complete genome sequences from the GenBank, we intended to identify plausible subtypes. Our analysis confirmed the existence of 8 genotypes and proposed 5 subtypes for HDV genotype-1 and 2 subtypes for genotype-2 and genotype-4.

2 | MATERIALS AND METHODS

2.1 | Patient samples and public data

Serum samples of 116 chronically infected HDV subjects were retrieved from six different medical centres which include Germany (n = 72) from Institute of Virology, Technische Universität München, Munich, Germany; Institute of Virology, University of Duisburg-Essen, Germany and Hannover Medical School, Hannover, Germany; Italy (n = 18) from Department of Gastroenterology and Hepatology, Molinette Hospital, Turin, Italy; Spain (n = 12) from Liver Unit, Department of Internal Medicine, Hospital Universitari Vall d'Hebron and Universitat Autònoma de Barcelona, Spain; and Iran (n = 14) from Tehran Hepatitis Center, Tehran, Iran. From these study subjects, 116 L-HDAg sequences were amplified and sequenced. Additionally, 13 full-length HDV genomes (ranging from 1670 to 1685 nt) from these 116 isolates were also amplified and sequenced. These include seven isolates from Germany, eight isolates from Italy and two isolates from Spain. This well-characterized cohort of patients was established to study CD8 epitopes within the HDV protein (L-HDAg). HLA-typing of these patients was also performed.[20] The local ethics committees have approved this study according to the 1975 Declaration of Helsinki guidelines. All patients were interviewed and informed about the study before they agreed to participate by signing a written consent. All serum samples were transferred and preserved at -80 Celsius until use.

2.2 | PCR amplification

Total RNA was extracted from the patients' serum samples and was reverse transcribed into cDNA by reverse transcriptase from the Moloney murine leukaemia virus (Promega, Madison, WI) at 37°C for 1 hour using an HDV-specific primer, 771R (Figure 1). In order to obtain sequences of the L-HDAg encoding region, cDNA products were amplified by Pfu DNA polymerases (Promega) in a two-step nested PCR using HDV-specific primers 891-F, 339-R, 912-F and 1674-R under the following thermal profile: 94°C for 10 minutes and a 35 cycle of 94°C for 30 seconds, 54°C for 45 seconds, and 72°C for 90 seconds and then an elongation step at 72°C for 7 minutes PCR products were purified via QIAquick® Gel Extraction (QIAGEN) and were directly sequenced on an ABI 3730xl DNA Analyzer using internal primers, 912-F and 1674-R. All sequences were submitted to the GenBank, NCBI (Accession numbers: MF175257-MF175360[20] and MH447633-MH447644).

2.3 | Redundancy reduction of the HDV dataset

To perform evolutionary and phylogenetic analysis, a total of 116 L-HDAg isolates of our cohort and 621 L-HDAg nucleotide sequences belonging to HDV genotype-1 to genotype-8 retrieved from the GenBank® database were included in our study. This dataset of 737 (621 + 116) L-HDAg nucleotide sequences included several almost identical or highly similar sequences. The presence of multiple copies of highly similar sequences can produce misleading results due to overrepresentation of identical sequences. However, this problem can be solved by employing a sequence clustering method to establish a non-redundant dataset. To eliminate highly similar sequences from the initial dataset of 737 isolates, Cluster Database at High Identity with Tolerance method (CD-HIT) was employed.[21]

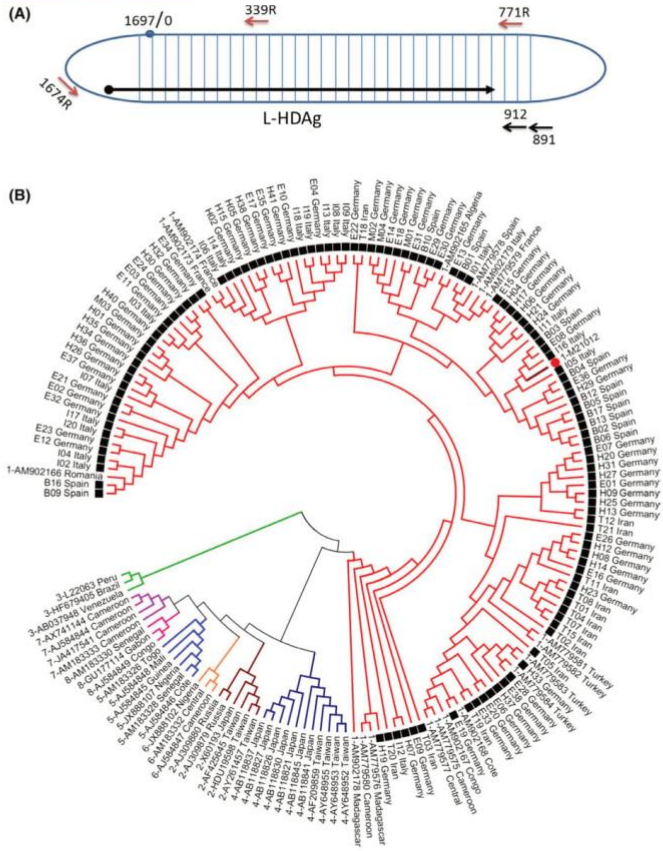

FIGURE 1 PCR amplification and phylogenetic tree obtained by distance analysis. A, Schematic representation of hepatitis delta virus (HDV) genome and the location of PCR primers. B, Maximum-likelihood phylogenetic tree based on the whole large delta antigen sequences of 116 strains from this study (marked by ■) and reference sequences from GenBank representing HDV genotype-1 to genotype-8

This method is based on a greedy incremental algorithm to cluster highly identical sequences and removes redundant sequences. First, sequences are sorted in a descending manner according to their length. The longest sequence is taken as the representative member of the first cluster. Then each remaining sequence is compared to the representatives of the existing clusters and if it is similar to one of them, it is inserted into the cluster. Otherwise, a new cluster is formed with that sequence as a representative. This method reports non-redundant sequence representatives for each cluster. For 116 isolates, sequence identity threshold was set to the default value of 90%. For the 621 GenBank isolates, we tested sequence identity thresholds in the range from 85% to 99% and selected 93% as this value resulted in the maximum non-redundant representation for all 8 genotypes along with fewer singletons (sequences which do not share sufficient identity with any other HDV sequences) (Figure 2).

2.4 | HDV sequence identification and phylogenetic analysis

Non-redundant L-HDAg nucleotide sequences were aligned using CLUSTAL Omega.[22] Moreover, the aligned sequences were analysed using other multiple sequence alignment tools such as MUSCLE and MAFFT.[23,24] To revisit the taxonomic classification of HDV genotypes, phylogenetic analysis of non-redundant L-HDAg sequences were performed using MEGA version 7.[25] Phylogenetic trees were reconstructed using the neighbour-joining[26] and the Maximum-likelihood[27] methods based on the Tamura-Nei nucleotide substitution model.[28] The authenticity of the phylogenetic tree analysis was assessed and assured by performing bootstrap (at 1000 pseudoreplicates), which generated the bootstrap probability for each interior branch in the tree.[29] Bootstrap values higher than 70% were

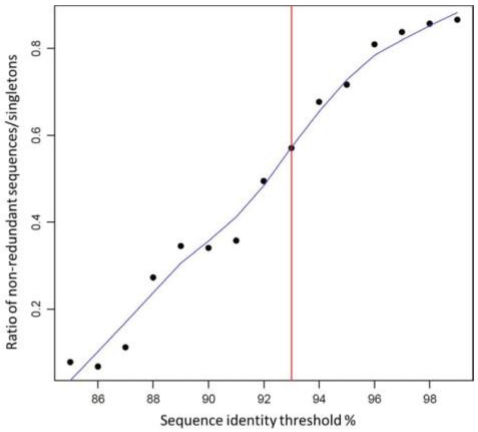

FIGURE 2 Distribution of the ratio of non-redundant sequences to singleton sequences over different identity thresholds. A range of sequence identity thresholds was tested to find a reasonable balance between the non-redundant sequences and singleton sequences for 621 Genbank sequences using Cluster Database at High Identity with Tolerance method. The selected identity threshold (93%, median) is represented by a red line

considered significant. Furthermore, we validated our phylogenetic analyses using a Bayesian approach; 5000 trees were initially built from $2x10^6$ bootstrap replicates using the MrBayes program version 3.0 ß4.[30] The posterior probabilities for a specific node to be observed were recorded. The consensus tree was obtained and visualized using Figtree v1.3.1.[31]

Pairwise genetic distances for the non-redundant nucleotide sequences were estimated using Kimura's 2 parameter method[32] which was used to calculate inter-genotypic differences for eight HDV genotypes (1 to 8). Similarly, genetic distances amongst each subtype were used to calculate inter-subtype differences. Differences in the genetic distances were expressed as mean ± standard deviation.

2.5 | Recombinant analysis

All 116 HDV isolate sequences were analysed for evidence of recombination using RDP4,[33] a software package for statistical identification and characterization of recombination events in nucleotide sequences. In short, RDP4 simultaneously utilizes a range of non-parametric recombination detection methods: RDP, GENECONV, BOOTSCAN, MAXICHI, CHIMAERA, SISCAN and 3SEQ. RDP4 treats every sequence within the analysed alignment as a potential recombinant, systematically screens sequence triplets/or quartets to identify viruses that contain a recombinant and identifies two sequences that could serve as parents while performing a statistical evaluation of recombination signals. The HDV sequence was considered a recombinant if the recombination signal was supported by at least three methods with a *P*-value of ≤0.05 after Bonferroni correction for multiple comparisons implemented in RDP4.

2.6 | Subtype-specific amino acid signatures

Subtype-specific amino acid signatures (AAS) were predicted for HDV genotypes-1, -2 and -4. Together with 116 L-HDAg isolates of our cohort, 295 isolates for HDV-1, 30 isolates for HDV-2 and 142 isolates for HDV-4 were retrieved from GenBank. Amino acid sequences of HDV-1, -2 and -4 were multiply aligned, and subtype-specific residues were predicted using the SDPpred method.[34,35] SDPpred uses mutual information (MI) to score the difference between the amino acid distributions at the given multiple sequence alignment (MSA) position *p* to distribution of proteins into specificity groups.

MI reflects statistical association between two random variables α and *i*, computed as:

$$MI_p = \sum_{i=1}^{N} \sum_{\alpha=1}^{20} f_p(\alpha,i) \log \frac{f_p(\alpha,i)}{f_p(\alpha) f(i)},$$

where $f_p(\alpha,i)$ is the frequency of the amino acid α in group *i* in position *p*, $f_p(\alpha)$ is the frequency of amino acid in position *p* in the whole alignment, $f(i)$ is the fraction of protein in group *i*.

After the calculation of MI values, *Z*-scores are calculated for each position in the MSA.

$$Z_i^I = \frac{I_{i-(I_i^{exp})}}{\sigma(I_i^{exp})}$$

The estimated *Z*-scores for each position in the MSA were evaluated by setting an appropriate *Z*-score threshold using the Bernoulli estimator. Subsequently, positions that are least probable to occur by chance and have scores higher than the *Z*-score threshold are referred specificity-determining position (SDPs).[34,35]

3 | RESULTS

3.1 | Generation of a non-redundant HDV dataset

The first set of phylogenetic analysis on the large delta antigen region indicated that all 116 isolates of our well-characterized cohort of European patients[20] cluster together with HDV genotype-1 (Figure 1). Then, these 116 sequences (covering L-HDAg) from our European cohort and 621 L-HDAg sequences from the GenBank were used to create a non-redundant sequence dataset. Based on the sequence clustering method CD-HIT, this initial dataset of 737 L-HDAg sequences was reduced to a set of 98 non-redundant L-HDAg nucleotide sequences using an identity threshold of 93%, median, which includes 12 European isolates from our cohort and 86 GenBank sequences (Figure 2). The selected identity threshold is a balance of non-redundant and singletons (ie unique sequences that are represented only once and do not overlap with other sequences). In total, 56 of 98 non-redundant sequences belonged to

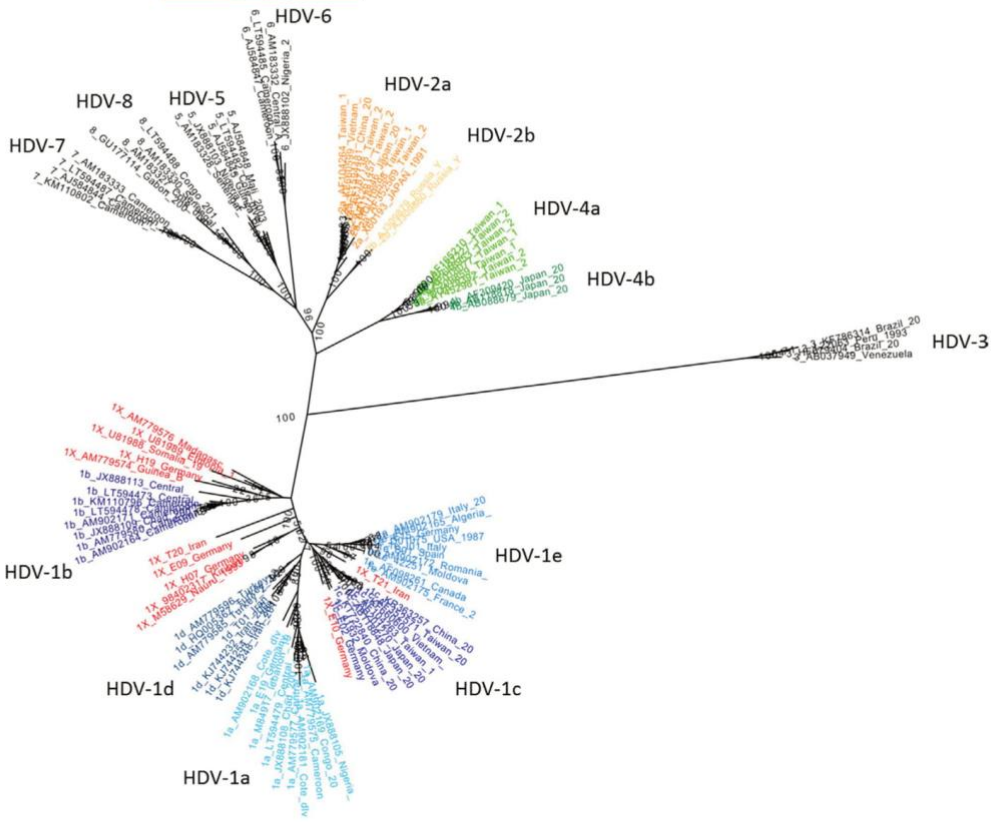

FIGURE 3 Phylogenetic tree of 98 non-redundant large hepatitis delta antigens nucleotide sequences using MrBayes, including 12 sequences from our European cohort and 86 GenBank sequences. Hepatitis delta virus (HDV)-1 subtypes are highlighted in shades of blue, whereas the sequences highlighted in red are border line cases. Taxas highlighted in orange and green represent HDV-2 and HDV-4 subtypes, respectively

HDV-1 (57%) while the remaining 42 sequences (43%) belonged to HDV genotype-2 to genotype-8. Specifically, the 56 HDV-1 sequences from different countries include America, Europe, Middle East, East Asia and Africa; 11 selected sequences of HDV-2 belong to countries such as Japan, Taiwan and Russia; 4 sequences of HDV-3 belong to South American countries such as Peru, Venezuela and Brazil; 10 sequences of HDV-4 correspond to Asian countries, *that is* Japan and Taiwan; and 5 nucleotide sequences belong to HDV-5, whereas HDV-6, -7 and -8 comprised of 4 nucleotide sequences of African origin (Figure 3).

3.2 | Genotype classification

To revisit the actual HDV classification scheme, a phylogenetic analysis was performed for the dataset of 98 L-HDAg non-redundant nucleotide sequences which were aligned using Clustal Omega. Moreover, applying other alignment tools including MUSCLE, MAFFT resulted in identical alignments and consistent phylogenetical topologies. The verified alignment was subjected for the phylogenetic analysis using neighbour-joining, maximum-likelihood and bayesian methods.

Phylogenetic analysis of these L-HDAg nucleotide sequences revealed the grouping of sequences into eight distinct genotypes. Pairwise genetic distances of these nucleotide sequences were used to estimate inter-genotype differences. The lowest inter-genotype divergence of ≥10% was reported between HDV-5 and HDV-2, whereas the highest inter-genotype difference was estimated for HDV-3 against all other HDV genotypes with an average of 28.44% (Figure 4). Additionally, inter-genotype nucleotide differences estimated for complete HDV genome and small hepatitis delta antigen (S-HDAg) nucleotide sequences also revealed an inter-genotype sequence divergence of ≥10% (data not shown).

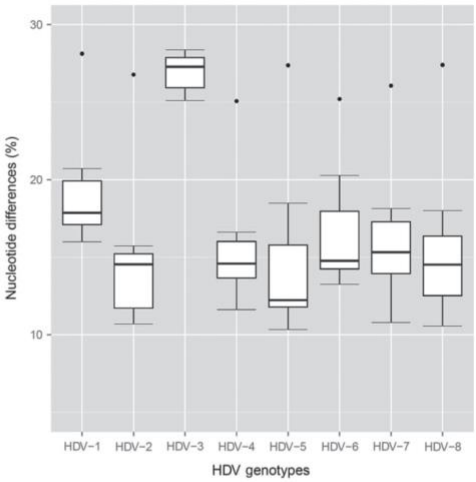

FIGURE 4 Box plot showing large hepatitis delta antigens nucleotide divergence (645 nt) amongst the hepatitis delta virus (HDV) eight genotypes. The whiskers in the box plots correspond to the maximum and the minimum value, whereas the dots correspond to the outliers. Amongst the genotypes, HDV-3 being the most diverse exhibits the highest nucleotide difference against other genotypes are shown as dots

3.3 | Subtypes of HDV genotype-1

Preliminary analysis of 116 L-HDAg isolates from our European cohort shed a light on possible existence of clusters within the HDV genotype-1 (Figure 1). Phylogenetic and sequence analysis of 98 L-HDAg non-redundant sequences revealed five subtypes within HDV-1 (Figure 3). These subtypes (supported by >70% bootstrap value) were classified as HDV-1a, -1b, -1c, -1d and -1e (Figure 3). Similarly, distinct clusters were also found for HDV genotype-2 and -4 (Figure 3). The bootstrap values confirmed these proposed subtypes within HDV-1, -2 and -4 (Figure S1). In addition, phylogenetic analysis of 89 full-length genomic sequences involving 13 isolates from our European cohort (MH457142-MH457154) revealed several clusters within HDV genotype-1 (Figure S2).

The average sequence identity between the 56 HDV-1 sequences was found to be 92.06%. Inter-subtype differences amongst HDV-1 nucleotide sequences were computed by calculating their genetic distances (Table 1). The inter-subtype difference between L-HDAg nucleotide sequences of the HDV-1a and HDV-1b,c,d,e was 5.34 ± 1.40%, that between HDV-1b and HDV-1a,c,d,e was 7.27 ± 0.33%, that between HDV-1c and HDV-1a,b,d,e was 4.47 ± 1.96%, that between HDV-1d and HDV-1a,b,c,e was 4.34 ± 1.82% and that between HDV-1e and HDV-1a,b,c,d was 5.0 ± 2.19% (Tables S1 and 1, left part). The lowest inter-subtype difference—3.2%—was between subtypes 1c and 1e, while the highest sequence diversity was observed between subtypes 1a and 1b (8.7%). The same calculations were performed for 53 full-length genomic and 56 S-HDAg sequences belonging to HDV-1 (Table 1, right part). No major sequence differences were observed amongst L-HDAg, S-HDAg and full-length genomic sequences.

3.4 | Unique isolates

Adhering to the above-mentioned classification, 44 of 56 non-redundant HDV-1 isolates were successfully clustered into five HDV-1 subtypes. However, 12 isolates could not be assigned to any of the five HDV-1 subtypes as a result of low bootstrap value (<70%) and low sequence identity (<90%) (Figure 3, Table S2). Six non-redundant isolates belonged to sequences from our European cohort study while the remaining six isolates originated from the GenBank database (Figure 3 highlighted in Red).

We used these six isolates from our cohort (T20, T21, H19, E10, E09, H07), which did not correspond to any subtype, and performed sequence comparisons against our complete dataset. We observed that T20, H19 and E09 isolates showed ≤90% sequence identity against each of the five HDV-1 subtypes, whereas H07, E10 and T21 isolates showed sequence identity between 90% and 94%. These three isolates are considered as unique cases because they share the same sequence identity amongst the three subtypes and exhibit low bootstrap support (Table S2, Figure 3). Additionally, T21 isolate is novel in having a 648 nucleotide sequence. Extension of the sequence by three base pairs might result in a change of the HDV protein conformation and its virulence. Furthermore, the identity scores of these isolates were found to be considerably lower as compared to the average identity score for each of the five subtypes, *that is* 1a (94.78%), 1b (96.40%), 1c (97.41%), 1d (96.53%) and 1e (96.76%); therefore, these isolates were not assigned to any of the 5 subtypes of genotype-1.

Furthermore, we observed that the isolate H19 outgroups the subtype "1b" along with other highly diverged sequences from Ethiopia, Madagascar and Guinea Bissau, while the three isolates (T20, H07 and E09) clustered outside the HDV-1 alongside Kiribati and Nauru isolates (Figure 3). These highly diverged sequences were not assigned to any subtype due to a very low bootstrap support (<70%) and are considered to be borderline cases.

3.5 | Subtypes for HDV genotype-2 and -4

We employed the same approach using the 621 GenBank sequences and identified two distinct subtypes for HDV genotype-2 (2a-2b) and -4 (4a-4b), respectively (Figure 3). Inter-subtype difference between the nucleotide sequences of the genotype-4a and -4b was found to be 6.49 ± 0.81%. Whereas the nucleotide sequences of genotype-2a and -2b showed sequence divergence of 7.35 ± 0.49% (Table 2).

Similarly, the inter-subtype differences for HDV genotype-2 and -4 were also estimated for complete HDV genome and S-HDAg nucleotide sequences. The subtype differences estimated for S-HDAg within HDV genotype-2 and -4 were in accordance with that of L-HDAg nucleotide sequences.

TABLE 1 Comparison of nucleotide sequences amongst HDV-1 subtypes

HDV-1 Subtypes	HDV-1e	HDV-1d	HDV-1c	HDV-1b	HDV-1a	HDV-1a	HDV-1b	HDV-1c	HDV-1d	HDV-1e	HDV-1 Subtypes
HDV-1a	8.00	6.39	6.77	9.45	94.14 / 94.78	93.86 / 94.78	8.77	6.72	5.51	7.18	HDV-1a
HDV-1b	9.56	8.54	9.06	96.19 / 96.40	7.08	7.08	96.17 / 96.40	8.52	7.40	8.05	HDV-1b
HDV-1c	4.46	4.00	95.48 / 97.41	7.18	4.65	4.65	7.18	97.30 / 97.41	4.12	4.48	HDV-1c
HDV-1d	4.96	95.85 / 96.53	**3.00**	7.02	3.85	3.85	7.02	**3.00**	95.68 / 96.53	4.49	HDV-1d
HDV-1e	97.02 / 96.76	3.46	3.03	**7.75**	5.75	5.75	**7.75**	3.03	3.46	97.44 / 96.76	HDV-1e

Small Hepatitis Delta Antigen (S-HDAg) nucleotide sequence

Complete HDV genome nucleotide sequence

Large Hepatitis Delta Antigen (L-HDAg) nucleotide sequence

Complete genomes (ranging from 1670 to 1685 nt), large hepatitis delta antigens (L-HDAg; 645 nt) and partial delta antigens (S-HDAg; 585 nt) of hepatitis delta virus (HDV-1) subtypes are compared. The table is partitioned into three triangles. The small triangles on the right and left side contain (a) per cent identity scores for each HDV-1 subtype (grey boxes) and (b) inter-subtype sequence divergence scores (white boxes) for 44 complete genomes and partial delta antigens, respectively. The large triangle in the centre represents the per cent identity scores of HDV-1 subtypes (grey boxes) and inter-subtype difference scores based on 44 L-HDAg sequences. The nucleotide sequences include 10 HDV-1a, 8 HDV-1b, 8 HDV-1c, 7 HDV-1d and 9 HDV-1e. For L-HDAg sequences, the lowest inter-subtype divergence score observed between HDV-1c and -1d was 3.00%, whereas the highest inter-subtype difference between HDV-1b and -1e was 7.75%.

3.6 | Geographical distribution of genotype-1 subtypes

The five distinct subtypes within HDV genotype-1 were identified by phylogenetic analysis with ≥90% bootstrap, belonging to a different geographical region. (Figures 3 and S1). We identified two major distinct clusters for sequences belonging to Sub-Saharan Africa, designated as HDV-1a and HDV-1b with a significant bootstrap value of 100% and 100%, respectively. These two clusters consisted of 10 and 8 nucleotide sequences, respectively. The remaining three clusters designated as HDV-1c, HDV-1d and HDV-1e, represent a distinct geographical region. Subtype HDV-1c with significant bootstrap support of 90% consists of 9 L-HDAg sequences from Central Asian countries such as China, Vietnam, Japan and Taiwan. Subtype HDV-1d accounts for 7 L-HDAg sequences from Turkey and Iran with a strong bootstrap support of 100%. Lastly, subtype HDV-1e comprises of 10 nucleotide sequences from European countries such as Spain, Italy, Germany, Romania with highly significant (100%) bootstrap support (Figures 3 and 5).

TABLE 2 Inter-subtype differences for HDV-2 and HDV-4 nucleotide sequences

HDV subtypes	% of Nucleotide difference (mean ± SD)
HDV-2a vs -2b	7.35 ± 0.49
HDV-4a vs -4b	6.49 ± 0.81

HDV, hepatitis delta virus.

3.7 | Geographical distribution of subtypes of genotype-2 and -4

The sequences belonging to genotype-2 formed two distinct clusters with strong bootstrap support of 100% and were named HDV-2a and HDV-2b. The sequences from Japan and Taiwan were classified as HDV-2a and the sequences from Russia (Yakutia) were classified as HDV-2b (Figures 3 and 5). Similarly, genotype-4 diverged into two clusters with highly significant (100%) bootstrap support and was designated as HDV-4a and HDV-4b. The nucleotide sequences from Taiwan were designated as HDV-4a, while the sequences from Japan were designated as HDV-4b (Figures 3 and S1).

3.8 | AAS of HDV subtypes

In addition to phylogenetic and sequence analysis at the nucleotide level, comparative protein sequence analysis was performed for the proposed subtypes of genotype-1, -2 and -4 to identify subtype-specific residues which may create AAS within the amino acid sequences of HDV genotype-1, -2 and -4. We could find several sequence positions which showed specificity for different subtypes. SDPpred predicted six subtype-specific residue positions for HDV-1

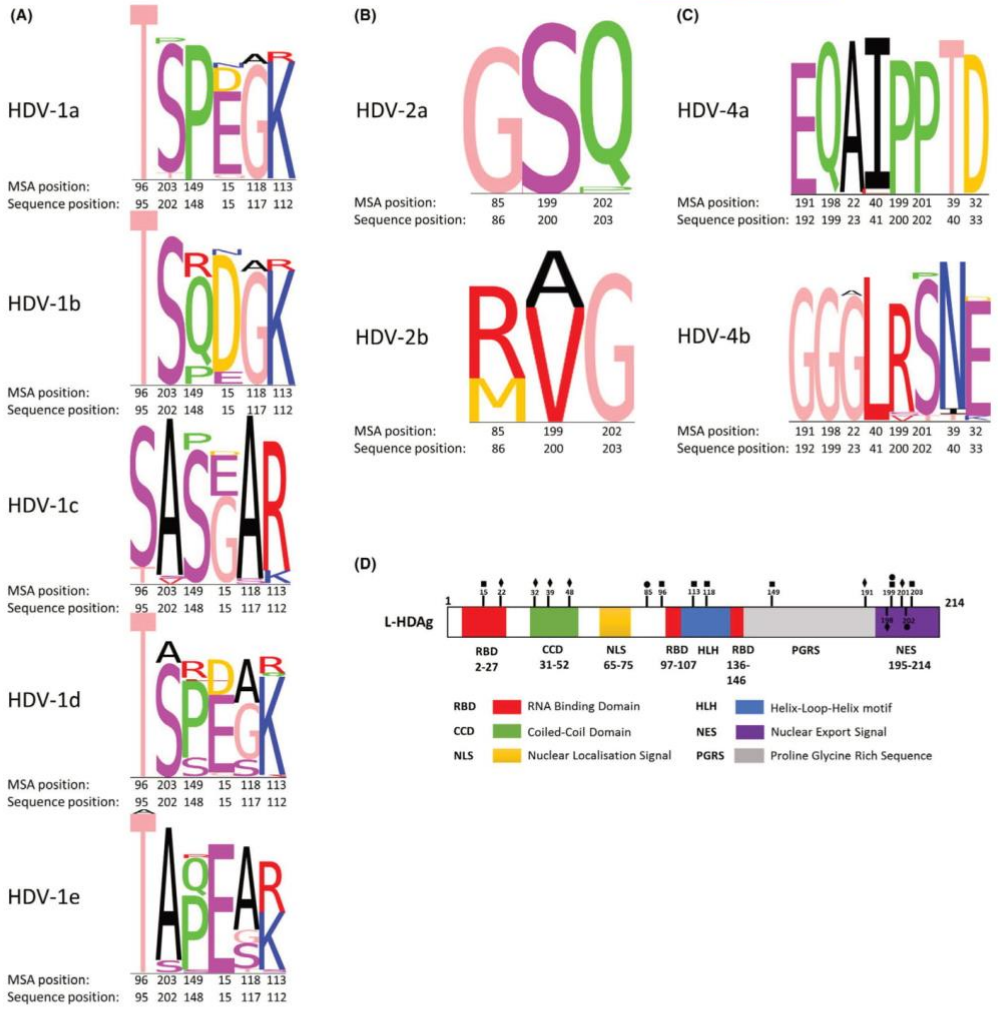

FIGURE 5 Subtype-specific amino acid residues predicted for hepatitis delta virus (HDV)-1, -2 and -4 using SDPpred. One letter codon of amino acids is demonstrated for each residue and the height of each letter reflects the frequency of that amino acid. A, Subtype-specific amino acid signatures (AAS) correlated with HDV-1 subtypes are located at residues 95, 202, 148, 15, 117 and 112. B, Subtype-specific AAS correlated with HDV-2 subtypes are located at residues 86, 200 and 203. C, Subtype-specific AAS correlated with HDV-4 subtypes are located at residues 192, 199, 23, 41, 200, 202, 40 and 33. D, Functionally important domains of large hepatitis delta antigens nucleotide. The predicted specificity-determining positions for HDV-1, HDV-2 and HDV-4 subtypes are represented with filled squares, circles and diamonds, respectively

(Figure 5), while three and eight subtype-specific AAS were predicted for HDV-2 and HDV-4, respectively (Figure 5A,B).

These amino acid preferences at a particular SDP can be used to predict subtypes. For instance, occurrence of an amino acid "A or Alanine" (99%-100%) at position 117 and 202 serves as an indicator for subtype "1c" (Figure 5A). Similarly, amino acid preferences (98%-100%) of "A or Alanine," and "E or Glutamic acid" at positions 202 and 15, respectively, indicate subtype "1e" (Figure 5A). Subtypes of genotype-2 and -4 also reveal different amino acid specificities (>90%) at different positions (Figure 5B,C). These specificities determining positions occur within the functional domains of L-HDAg, including RNA binding motifs (2-27), a coiled coil domain (31-52), a

nuclear localization signal (66-88), a helix loop helix motif (111-138), proline/glycine rich sequence (PGRS) and virus assembly signal or nuclear export signal (195-214) (Figure 5D).

3.9 | Analysis of potential recombination of HDV isolates

Recombination of HDV genomes has been found in a small number of patients with chronic HDV infection.[36,37] To exclude such events in our cohort which may influence attribution of single isolates to certain sub-genotypes, all 116 experimentally sequenced isolates of our cohort were analysed for evidence of recombination. Recombination analysis was performed using seven methods implemented in RDP4. The HDV sequence was considered a recombinant if the recombination signal was supported by at least three methods with *P*-value of≤0.05 after Bonferroni correction for multiple comparisons. However, none of the 116 sequences showed signs of recombination (data not shown).

4 | DISCUSSION

In this study, we revisited the current classification of HDV isolates by performing sequence comparisons and phylogenetic analysis of L-HDAg (n = 116) and 13 complete genome sequences of our cohort together with 621 publicly available sequences. To prevent a bias in classification due to a large number of redundant sequences in this data set, we employed the sequence clustering method CD-HIT for HDV sequence classification. As a result, we used 98 non-redundant sequences for HDV genotype classification and confirmed the known eight genotypes (1 to 8) of HDV.

Most recently, based on geographic distribution and genetic variability of HDV, different classification models have been proposed by Le Gal et al.[15] Delfino et al.[38] and Miao et al.[39] Our study is in agreement with studies by Le Gal and Miao, in confirming the previous definition of eight genotypes and identification of possible subtypes within these known genotypes. Delfino et al. also demonstrated the existence of specific amino acids signatures similar to our findings for different genotypes. Moreover, they agreed on genotype one and three, however summarized all other genotypes into one clade referred as genotype-2. Reducing the number of genotypes to just three seems not to be helpful for studies on epidemiology of HDV and clinical aspects of HDV infection.

These different concepts of genotyping proposed in these recent publications including our study indicate that there is an urgent need to involve the International Committee on Taxonomy of Viruses to evaluate the recent classifications and propose an updated system of nomenclature which is accepted by the scientific community. This would be also very important for clinical aspects to study differences in pathogenesis and response on antiviral therapy.

Additionally, we propose a sub-classification of HDV genotypes by defining five distinct subtypes within HDV genotype-1 (1a-1e) as well as two subtypes for HDV genotype-2(2a-2b) and genotype-4(4a-4b). By comparative sequence analysis, we identified inter-subtype differences within HDV genotypes-1, -2 and -4 and propose that isolates with genetic difference between 3% and 10% can be assigned to a potential subtype. The classification of HDV into the proposed subtypes at the nucleotide level was confirmed using amino acid sequence comparisons followed by the identification of AAS amongst different subtypes.[34,35] These amino acid preferences amongst different subtypes may suggest adaptation to the host and environment and are likely to be functionally relevant. Definition of subtypes and AAS may help to understand the key factors that contribute to replication competence, pathogenesis, response to therapy and prevalence in certain populations. For treatment studies in patients with chronic HDV infection, subtyping may be of importance to determine subtypes which may have different response rates upon treatment with certain drugs. This has been shown for other viruses like HCV. Direct acting antiviral treatment of HCV with protease inhibitors in chronic hepatitis C has shown that the rate of sustained direct response (SVR) was different for HCV subtype 1b vs 1a.[40]

In addition, subtyping of HDV genotypes may be important to recognize shift of the mode of infection, age of patients or certain populations. Such a shift has been shown for instance in patients infected with HCV and drug addiction. Therefore, subtyping of HDV may facilitate the tracking of replacement of certain subtypes through different modes of transmission in populations. With respect to subtypes of genotype 1 of HCV, there was a notable variation in the distribution of the prevalent subtypes 1a and 1b in different age groups. In individuals older than 50 years, subtype 1b was most frequent.[41] Thus, subtype 1b has been gradually substituted for subtype 1a over the last 20 years.

Despite the attribution of most of 56 non-redundant genotype-1 isolates to a defined subtype, we identified 12 of them highly divergent sequences (<70% bootstrap values) more distantly related with all other sequences within the genotype-1. Such isolates are referred as borderline cases exhibiting as low as 83% sequence identity and hence could not be assigned to any subtype of HDV genotype-1. The diversification trend of these 12 isolates might be explained by a high mutation rate experienced at early time point of infection or might suggest the emergence of new subtypes and more complex classification in the future. For instance, such borderline cases from the Micronesian island (Nauru and Kiribati) publically available, at GenBank, may appear as a new subtype. However, additional Micronesian isolates are needed to create such a potential new subtype.

In summary, based on our study we propose that the clusters exhibiting an inter-genotypic difference ≥10% are classified as genotypes, while the clusters showing inter-subtype difference ≥3% to <10% can be classified as subtypes. In addition, comparative sequence analysis at the protein level revealed subtype-specific amino acids for genotypes HDV-1, -2 and -4. In-depth understanding of evolutionary adaptation processes shaping population diversity and adaptation to different hosts may help in identifying optimal amino acid sequences or residues for immunotherapies and prophylactic

vaccines. Subtyping may have also benefit to determine shift in the epidemiology due to migration or use of i.v. drugs or to discriminate responders had non-responders to new antiviral drugs.

ACKNOWLEDGEMENTS

This work was financially supported by internal funds of the Medical Faculty of the University of Duisburg-Essen/Technical University of Munich and by Registry Grant from European Association for the Study of the Liver (EASL). HK and ZU were supported by a grant from the German Academic Exchange Service (DAAD). We thank all the people who volunteered in this research and the staff working in the diagnostic sections, institutes of virology, University Hospital Essen and TU Munich, Germany, for their close collaborations recruiting the samples and technical assistance.

CONFLICT OF INTEREST

The authors declared that they do not have anything to disclose regarding funding or conflict of interest with respect to this study.

ORCID

Michael Roggendorf https://orcid.org/0000-0003-3293-6738

REFERENCES

1. Noureddin M, Gish R. Hepatitis delta: epidemiology, diagnosis and management 36 years after discovery. *Curr Gastroenterol Rep*. 2014;16(1):365.
2. Smedile A, Farci P, Verme G, et al. Influence of delta infection on severity of hepatitis B. *Lancet*. 1982;2(8305):945-947.
3. Alfaiate D, Deny P, Durantel D. Hepatitis delta virus: from biological and medical aspects to current and investigational therapeutic options. *Antiviral Res*. 2015;122:112-129.
4. Rizzetto M, Hepatitis D. Virus: introduction and epidemiology. *Cold Spring Harb Perspect Med*. 2015;5(7):a021576.
5. Chen HY, Shen DT, Ji DZ, et al. Prevalence and burden of hepatitis D virus infection in the global population: a systematic review and meta-analysis. *Gut*. 2018.
6. Shakil AO, Hadziyannis S, Hoofnagle JH, Di Bisceglie AM, Gerin JL, Casey JL. Geographic distribution and genetic variability of hepatitis delta virus genotype I. *Virology*. 1997;234(1):160-167.
7. Ivaniushina V, Radjef N, Alexeeva M, et al. Hepatitis delta virus genotypes I and II cocirculate in an endemic area of Yakutia, Russia. *J Gen Virol*. 2001;82(Pt 11):2709-2718.
8. Lee CM, Changchien CS, Chung JC, Liaw YF. Characterization of a new genotype II hepatitis delta virus from Taiwan. *J Med Virol*. 1996;49(2):145-154.
9. Sakugawa H, Nakasone H, Nakayoshi T, et al. Hepatitis delta virus genotype IIb predominates in an endemic area, Okinawa, Japan. *J Med Virol*. 1999;58(4):366-372.
10. Casey JL. Hepatitis delta virus. Genetics and pathogenesis. *Clin Lab Med*. 1996;16(2):451-464.
11. Casey JL, Brown TL, Colan EJ, Wignall FS, Gerin JL. A genotype of hepatitis D virus that occurs in northern South America. *Proc Natl Acad Sci U S A*. 1993;90(19):9016-9020.
12. Gomes-Gouvea MS, Soares MC, Bensabath G, et al. Hepatitis B virus and hepatitis delta virus genotypes in outbreaks of fulminant hepatitis (Labrea black fever) in the western Brazilian Amazon region. *J Gen Virol*. 2009;90(Pt 11):2638-2643.
13. Quintero A, Uzcategui N, Loureiro CL, et al. Hepatitis delta virus genotypes I and III circulate associated with hepatitis B virus genotype F in Venezuela. *J Med Virol*. 2001;64(3):356-359.
14. Le Gal F, Brichler S, Drugan T, et al. Genetic diversity and worldwide distribution of the deltavirus genus: a study of 2,152 clinical strains. *Hepatology*. 2017;66(6):1826-1841.
15. Le Gal F, Gault E, Ripault MP, et al. Eighth major clade for hepatitis delta virus. *Emerg Infect Dis*. 2006;12(9):1447-1450.
16. Makuwa M, Caron M, Souquiere S, Malonga-Mouelet G, Mahe A, Kazanji M. Prevalence and genetic diversity of hepatitis B and delta viruses in pregnant women in Gabon: molecular evidence that hepatitis delta virus clade 8 originates from and is endemic in central Africa. *J Clin Microbiol*. 2008;46(2):754-756.
17. Makuwa M, Mintsa-Ndong A, Souquiere S, Nkoghe D, Leroy EM, Kazanji M. Prevalence and molecular diversity of hepatitis B virus and hepatitis delta virus in urban and rural populations in northern Gabon in central Africa. *J Clin Microbiol*. 2009;47(7):2265-2268.
18. Radjef N, Gordien E, Ivaniushina V, et al. Molecular phylogenetic analyses indicate a wide and ancient radiation of African hepatitis delta virus, suggesting a deltavirus genus of at least seven major clades. *J Virol*. 2004;78(5):2537-2544.
19. Barros LM, Gomes-Gouvea MS, Pinho JR, et al. Hepatitis Delta virus genotype 8 infection in Northeast Brazil: inheritance from African slaves? *Virus Res*. 2011;160(1–2):333-339.
20. Karimzadeh H, Kiraithe MM, Kosinska AD, et al. Amino acid substitutions within HLA-B*27-restricted T cell epitopes prevent recognition by hepatitis delta virus-specific CD8 + T cells. *J Virol*. 2018;92(13):e01891-17.
21. Li W, Godzik A. CD-HIT: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*. 2006;22(13):1658-1659.
22. Sievers F, Higgins DG. Clustal Omega, accurate alignment of very large numbers of sequences. *Methods Mol Biol*. 2014;1079:105-116.
23. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 2004;32(5):1792-1797.
24. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res*. 2002;30(14):3059-3066.
25. Kumar S, Nei M, Dudley J, Tamura K. MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief Bioinform*. 2008;9(4):299-306.
26. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol*. 1987;4(4):406-425.
27. Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol*. 2003;52(5):696-704.
28. Tamura K, Nei M, Kumar S. Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc Natl Acad Sci U S A*. 2004;101(30):11030-11035.
29. Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*. 1985;39(4):783-791.
30. Ronquist F, Teslenko M, van der Mark P, et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol*. 2012;61(3):539-542.
31. Rambaut A. *FigTree version 1.3.1*. Edinburgh, UK: University of Edinburgh; 2009.
32. Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol*. 1980;16(2):111-120.
33. Martin DP, Murrell B, Golden M, Khoosal A, Muhire B. RDP4: detection and analysis of recombination patterns in virus genomes. *Virus Evol*. 2015;1(1):vev003.

34. Kalinina OV, Novichkov PS, Mironov AA, Gelfand MS, Rakhmaninova AB. SDPpred: a tool for prediction of amino acid residues that determine differences in functional specificity of homologous proteins. *Nucleic Acids Res*. 2004;32:W424-W428.
35. Kalinina OV, Mironov AA, Gelfand MS, Rakhmaninova AB. Automated selection of positions determining functional specificity of proteins by comparative analysis of orthologous groups in protein families. *Protein Sci*. 2004;13(2):443-456.
36. Lin CC, Lee CC, Lin SH, et al. RNA recombination in hepatitis delta virus: identification of a novel naturally occurring recombinant. *J Microbiol Immunol Infect*. 2017;50(6):771-780.
37. Wang TC, Chao M. RNA recombination of hepatitis delta virus in natural mixed-genotype infection and transfected cultured cells. *J Virol*. 2005;79(4):2221-2229.
38. Delfino CM, Cerrudo CS, Biglione M, Oubina JR, Ghiringhelli PD, Mathet VL. A comprehensive bioinformatic analysis of hepatitis D virus full-length genomes. *J Viral Hepat*. 2018;25(7):860-869.
39. Miao Z, Zhang S, Ma Z, et al. Recombinant identification, molecular classification and proposed reference genomes for hepatitis delta virus. *J Viral Hepat*. 2018.
40. Hezode C, Fontaine H, Dorival C, et al. Effectiveness of telaprevir or boceprevir in treatment-experienced patients with HCV genotype 1 infection and cirrhosis. *Gastroenterology*. 2014;147(1):132-142.e134.
41. Ross RS, Viazov S, Renzing-Kohler K, Roggendorf M. Changes in the epidemiology of hepatitis C infection in Germany: shift in the predominance of hepatitis C subtypes. *J Med Virol*. 2000;60(2):122-125.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Karimzadeh H, Usman Z, Frishman D, Roggendorf M. Genetic diversity of hepatitis D virus genotype-1 in Europe allows classification into subtypes. *J Viral Hepat*. 2019;26:900–910. https://doi.org/10.1111/jvh.13086

C Research Article III

viruses

MDPI

Article

HDVdb: A Comprehensive Hepatitis D Virus Database

Zainab Usman [1,†], Stoyan Velkov [2,†], Ulrike Protzer [2], Michael Roggendorf [2], Dmitrij Frishman [1] and Hadi Karimzadeh [2,3,*]

1 Department of Bioinformatics, Wissenschaftszentrum Weihenstephan, Technische Universität München, 85354 Freising, Germany; zainabasrar@gmail.com (Z.U.); d.frishman@wzw.tum.de (D.F.)
2 Institute of Virology, Technische Universität München, 81675 Munich, Germany; stoyan.velkov@tum.de (S.V.); protzer@tum.de (U.P.); michael.roggendorf@tum.de (M.R.)
3 Division of Clinical Pharmacology, University Hospital, LMU Munich, 80337 Munich, Germany
* Correspondence: Hadi.Karimzadeh@med.uni-muenchen.de
† These authors contributed equally to this work.

Received: 26 March 2020; Accepted: 11 May 2020; Published: 14 May 2020

Abstract: Hepatitis D virus (HDV) causes the most severe form of viral hepatitis, which may rapidly progress to liver cirrhosis and hepatocellular carcinoma (HCC). It has been estimated that 15–20 million people worldwide are suffering from the chronic HDV infection. Currently, no effective therapies are available to treat acute or chronic HDV infection. The remarkable sequence variability of the HDV genome, particularly within the hypervariable region has resulted in the provisional classification of eight major genotypes and various subtypes. We have developed a specialized database, HDVdb, which contains a collection of partial and complete HDV genomic sequences obtained from the GenBank and from our own patient cohort. HDVdb enables the researchers to investigate the genetic variability of all available HDV sequences, correlation of genotypes to epidemiology and pathogenesis. Additionally, it will contribute in understanding the drug resistant mutations and develop effective vaccines against HDV infection. The database can be accessed through a web interface that allows for static and dynamic queries and offers integrated generic and specialized sequence analysis tools, such as annotation, genotyping, primer prediction, and phylogenetic analyses.

Keywords: hepatitis delta virus; database; genotyping; Webserver

1. Introduction

Hepatitis D virus (HDV) infection remains the most difficult-to-treat form of viral hepatitis, affecting 15–20 million patients worldwide with chronic hepatitis, liver cirrhosis and hepatocellular carcinoma (HCC) [1]. The HDV infection in humans occurs so far only together with hepatitis B virus (HBV) because HDV needs the envelope proteins from HBV to complete its life cycle. Therefore, two main forms of HDV infection have been described: (1) coinfection; with a high rate of viral clearance in adults similar to HBV mono-infection [2], or (2) super-infection in the presence of a pre-existing HBV infection. The latter results in a persistent chronic HDV infection in 70–90% of the cases and is associated with an early risk to develop cirrhosis and HCC [3]. The current anti-HDV therapy is mainly based on administration of interferon with a very low response rate in patients [4,5] and high chance of relapse upon discontinuation [6]. Nevertheless, efforts have been made recently to develop new anti-HDV drugs to treat chronic HDV infection, with promising results in the clinical trials [7–11].

HDV is a small, spherical virus of 35–37 nm in diameter, with an envelope containing the hepatitis B surface antigen (HBsAg), which surrounds the genomic RNA-nucleoprotein complex [12]. The genome is a negative sense single-stranded RNA (1.67 kb), whose complementary strand (antigenomic RNA)

contains one single functional open reading frame (ORF) encoding two isoforms of the hepatitis delta antigen (HDAg), the small (*S*-HDAg, 195 aa) and the large (*L*-HDAg, 214 aa) [13,14]. The sequence encoding these isoform proteins resides in the antigenomic RNA, which, as a result of the cellular editing activity of ADAR-1, modifies the amber stop codon (UAG) to (UGG) of *S*-HDAg, resulting in the extension of the amino acid sequences by 19–20 aa at the C terminus [15].

HDV RNA sequences identified so far have been classified into eight known genotypes (HDV-1–8) based both on the nucleotide and amino acid sequences of the coding region of HDAg [16]. These genotypes are distributed across different geographical regions. In our recent studies, we identified and introduced different subtypes for the genotype 1, genotype 2 and genotype 4 [17]. Subgrouping of the so far identified HDV genotypes into distinct clusters or subtypes has been also suggested by others [18,19]. These data provide a clearer picture of the geographical a global distribution of HDV isolates. However, it is not known how these subtypes correlate with the clinical manifestation and response to therapy.

HDV-1 is the most geographically widespread genotype distributed across major regions such as Europe, Middle East, East Asia, America and Africa; whereas all other genotypes (HDV-2–8) are associated with distinct geographical and ethnic regions. HDV-2 and HDV-4 are found in North Asia and East Asia, respectively [20–23]; HDV-3 is exclusively found in the north part of South America (Brazil, Peru, Colombia, Argentina, Ecuador and Venezuela) [24–28] and HDV-5 to HDV-8 were previously described to be found "only" in Africa [29,30], however, a recent study reported HDV-8 isolates from Northeast Brazil, which presumably crossed the ocean through slave trades in the 16–18th centuries [31]. In humans, HDV infections with different genotypes exhibit different clinical courses and outcomes. For instance, HDV-1 strains show a broad spectrum of virulent and pathogenic phenotypes [32], HDV-2 (and HDV-4) cause milder forms of liver disease [33], whereas HDV-3 isolates are associated with outbreaks of fulminant hepatitis in South America [24]. The pathogenic properties of HDV 5–8 isolates are not well characterized [34].

For decades, HDV has been thought to have evolved in humans in combination with HBV as a helper virus providing a viral envelope to form infectious particles. Recently, however, HDV like sequences have been identified in a variety of animals and insects [35–37]. Sequence analysis in ducks revealed an approximately 1700-nt circular RNA genome with self-complementary, unbranched rod-like structures, and coiled-coil domains [36]. The predicted HDV-like protein discovered in ducks shares 32% amino acid similarity with the small delta antigen (*S*-HDAg) of the human HDV (hHDV).

This discovery of an HDV-like agent in ducks was followed by the identification of a deltavirus in snakes (*Boa constrictor*), designated as snake HDV (sHDV) [37]. Sequence comparison of the snake delta antigen (sHDAg) showed that its aa sequence is 55% identical to its human counterparts. Anti-sera raised against a recombinant sHDAg was used in immunohistology studies. A broad viral target was demonstrated in different snake cells, including neurons, epithelial cells and leukocytes. The duck and snake viruses constitute divergent phylogenetic lineages as compared to the human HDV (hHDV), which so far seem quite distant related to the known human isolates.

Using additional meta-transcriptomic data, highly divergent HDV-like viruses were also found to be present in fish, amphibians and invertebrates. These newly identified viruses share human HDV-like genomic features such as a small genome size of 1.7 kb in length [35].

The identification of a much broader range of hosts as initially anticipated and the fact that the HDV RNA genome can efficiently replicate in different tissues and species, raise the possibility that HDV is able to be transmitted independently of HBV. Perez-Vargas J. et al. [38] have shown, that envelope glycoproteins (GPs) of unrelated viruses can act as helper viruses for HDV including vesiculovirus, flavivirus and hepacivirus. These GPs can package HDV RNPs, allowing efficient egress of HDV particles in the extracellular milieu of coinfected cells and subsequent entry into the cells expressing the corresponding receptors. In vivo studies in humanized mice indicate that HDV RNPs packaged into an HCV envelope can propagate HDV infection in the liver of coinfected mice [38].

In recent years, the amount of HDV genomic data has increased exponentially. Intensive sequencing efforts have resulted in approximately 2621 nucleotide HDV sequences (partial and full length) deposited into the DDBJ, EMBL and GenBank databases. The GenBank is part of the International Nucleotide Sequence Database Collaboration (INSDC), which comprises of the DNA DataBank of Japan (DDBJ), the European Nucleotide Archive (ENA) and GenBank at NCBI. Those three organizations are synchronized and exchange data on a daily basis. Therefore, the sequences dataset can be retrieved using any of the platforms, i.e., the GenBank database. In order to exploit this large and growing collection of sequences efficiently and to facilitate sequence analysis we sought to develop a specialized database. Databases established for other types of viruses, in particular for HIV [39], HBV [40] and HCV [41] have proved to be very helpful for epidemiological and clinical studies, more importantly in characterizing resistance to direct anti-viral drugs. Here, we present the hepatitis delta virus database (HDVdb; http://hdvdb.bio.wzw.tum.de/). This comprehensive database collates HDV sequences and is mainly oriented towards the sequence analysis of HDV isolates, including the complete viral genomic sequences, large and small HDV antigen sequences (*L*-HDAg and *S*-HDAg, respectively). HDVdb provides a platform for genotyping and phylogenetic analyses including prediction of HDV genotypes for user-supplied HDV sequence entries. Moreover, the database will help in identifying the emerging variants related to immune escape from the B and T cell response as described recently [42,43] and in detecting therapy resistant variants across different HDV genotypes, which can be correlated with clinical studies.

2. Materials and Methods

The HDVdb building process began with the manual retrieval of all the HDV entries using the keyword: "hepatitis delta virus" from GenBank hosted at NCBI [44]. All results corresponding to taxon "Hepatitis delta virus" (taxid: 12475) were considered. Currently, a total of 2621 hepatitis delta virus nucleotide sequences are deposited into GenBank. These entries contain full sequence records of both HDV "complete genomic" sequences and subgenomic fragments (*S*-HDAg; 1–195 aa and *L*-HDAg; 1–214 aa) as well as partial cds sequences. GenBank entries containing complete HDV protein sequences were also incorporated. Majority of the sequences were retained to provide the maximum data information to our visitors, however, sequences shorter than 90 bases were not included into the dataset. The sequence dataset was than parsed by creating an automatic pipeline using Java programming language to extract essential information for each accession number such as strain name, genotype, country and date. In addition, 152 sequences lacking the genotype information were assigned to a genotype by performing similarity search using BLAST [45].

The HDVdb web interface is hosted using Apache HTTP server and runs on PHP Laravel framework. The HDVdb is updated on an annual basis. The software for the automatic annotation, as well as for the querying and the managing the database is implemented in Java and Bash programming languages. It makes use of all defined genotypes and their subgenotypes. The workflow of database construction is schematically demonstrated in Figure 1.

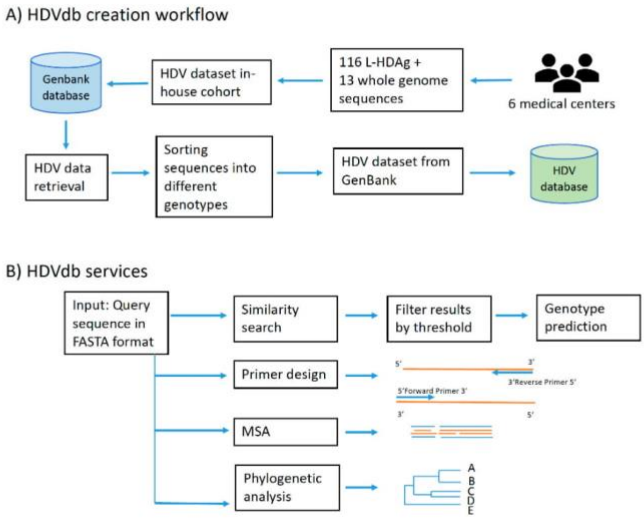

Figure 1. HDVdb construction and analysis workflow. (**A**) Building blocks of HDVdb based on publicly available and in-house isolates. (**B**) List of services available at the HDVdb. In Primer design, the orange lines represent template and blue lines represent the primers. In MSA (Multiple sequence alignment): blue and orange lines represent different aligned sequences, schematically.

3. Results and Discussion

The HDVdb is accessible online through the website: http://hdvdb.bio.wzw.tum.de/. HDVdb contains entries for human hepatitis delta virus sequences, with 512 complete genome sequences, as well as 1066 *L*-HDAg and *S*-HDAg and 1281 partial cds nucleotide sequences as well as protein sequences for *L*-HDAg and *S*-HDAg. These sequences can be directly downloaded from the database for any further analysis. Links to protein sequences for both *L*-HDAg and *S*-HDAg sequences are directly provided at the home page. Additionally, we included 13 complete genome (Accession MH457142-MH457154) and 116 *L*-HDAg sequences (Accession MF175257-MF175360, MH447633-MH447644) retrieved from six different medical centers of our European study cohort [17]. In this study, sequence conservation at each position across the entire length of the 322 multiply aligned genome sequences (i.e., genotype-1) was visualized. The multiple sequence alignments were performed using MUSCLE v3.8.5551 [46] whereas the evaluation was performed using customized Ruby scripts (Figure 2). We concluded that despite low conservation rate throughout the HDV genome, there were no significant differences on genotyping results using the whole genome or the *L*-HDAg encoding region.

The HDVdb is divided into a static and a dynamic part as demonstrated in Figure 3. The static part allows the user to access the general information about HDV. The homepage provides a data summary of updated number of *S*-HDAg, *L*-HDAg and complete genomes of all the eight known genotypes on the database. The user can retrieve pre-compiled protein and nucleotide datasets for complete genome, *L*-HDAg, and *S*-HDAg separately for each genotype, alternatively the user can also download a single FASTA file containing these datasets for all genotypes. In addition, the database also provides a tutorial to help the users with necessary technical information required to access tools available on the database.

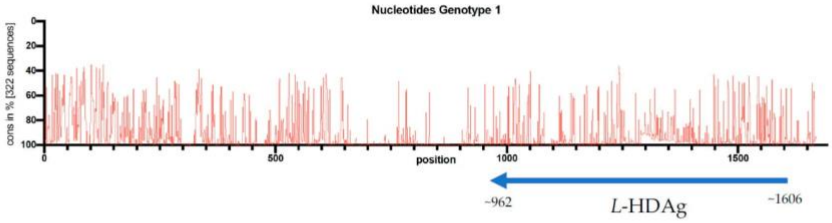

Figure 2. Conservation rate of 322 full genome nucleotide sequences of the HDV genotype 1. The single open reading frame of HDAg is located on antigenomic strand between position 962 and 1606 and indicated with an arrow under the genome.

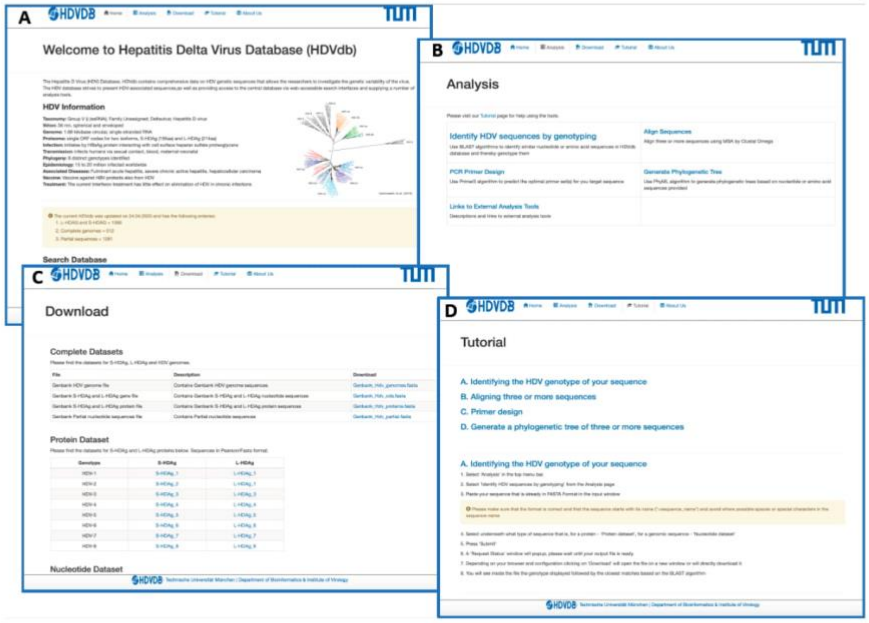

Figure 3. Web interface of the HDVdb. (**A**) The homepage, with the menu bar and all the menus repeated as buttons for ease-of-use. The page summarizes the characteristics and statistics of HDV. The nucleotide as well as protein datasets can be directly accessed through the homepage. The sequence files can be viewed for each genotype both in protein and nucleotide. (**B**) List of tools available on the website to analyze HDV related sequences. (**C**) List of sequence datasets available in FASTA format and updated on a regular basis. (**D**) A snapshot of the tutorial page with step-by-step instructions on how to use different tools available on this database.

The dynamic part allows the analysis of user-provided queries. The homepage presents an interactive search box that allows the visitors of our database with options to search sequences based on accession number, genotype, country and date for protein, complete genome, coding sequences for *L*-HDAg and *S*-HDAg, as well as partial sequences. The nucleotide and protein sequences queries can be genotyped using "Identify HDV sequences by genotyping" option. The webservice uses BLAST [45], which performs local alignments and scores the most relevant sequences to access the query genotype. A minimum of 75% identity score against the database is required to classify the query sequence to one

of the HDV genotypes. This threshold prevents the false positives to be classified and is based on our previous research [17].

Furthermore, we integrated computational tools for multiple sequence analysis (Clustal Omega, version 1.2.3 [47]), primer design (Primer3, version 2.3.7 [48]) and phylogenetic analyses (Phylip (PhyML), version 3.696 [49]), Figure 3. The user can also graphically visualize the phylogenetic trees on completion generated by FigTree, version 1.4.4. The request and response from these services was handled using PHP Laravel framework and bash scripting.

4. Conclusions

Hepatitis D has received a lot of attention in recent years, resulting in a flood of new findings and information, including next generation sequencing data. However, a platform capable of collecting and analyzing this growing body of data has so far been missing. Here we introduced HDVdb as a comprehensive database of human HDV sequences with a potential of expansion to the recently identified isolates from animals and insects. HDVdb allows the user to download structured data of all known HDV sequences. It also permits the user to use this data and perform comparative sequence analysis using multiple bioinformatics services available directly on HDVdb website.

Author Contributions: H.K. and M.R. conceived the idea of study. Z.U. extracted the data, created the webservices, and developed the database. Z.U. and S.V. implemented software services. Z.U., S.V., U.P., M.R., D.F. and H.K. wrote the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was financially supported by internal funds of the Medical Faculty of the Technical University of Munich and by Registry Grant from European Association for the Study of the Liver (EASL). Z.U. was supported by a grant (57129429) from the German Academic Exchange Service (DAAD).

Acknowledgments: We thank all the staff working at the diagnostic sections of the Institute of Virology of the Technical University of Munich for their technical assistance. We would also like to thank the network and system administrators both at TUM Department of Bioinformatics who helped make the service available online and accessible for all.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Chen, H.Y.; Shen, D.T.; Ji, D.Z.; Han, P.C.; Zhang, W.M.; Ma, J.F.; Chen, W.S.; Goyal, H.; Pan, S.; Xu, H.G. Prevalence and burden of hepatitis D virus infection in the global population: A systematic review and meta-analysis. *Gut* **2018**. [CrossRef] [PubMed]
2. Noureddin, M.; Gish, R. Hepatitis delta: Epidemiology, diagnosis and management 36 years after discovery. *Curr. Gastroenterol. Rep.* **2014**, *16*, 365. [CrossRef] [PubMed]
3. Smedile, A.; Farci, P.; Verme, G.; Caredda, F.; Cargnel, A.; Caporaso, N.; Dentico, P.; Trepo, C.; Opolon, P.; Gimson, A.; et al. Influence of delta infection on severity of hepatitis B. *Lancet* **1982**, *2*, 945–947. [CrossRef]
4. Wedemeyer, H.; Yurdaydin, C.; Dalekos, G.N.; Erhardt, A.; Çakaloğlu, Y.; Değertekin, H.; Gürel, S.; Zeuzem, S.; Zachou, K.; Bozkaya, H.; et al. Peginterferon plus adefovir versus either drug alone for hepatitis delta. *N. Engl. J. Med.* **2011**, *364*, 322–331. [CrossRef]
5. Wedemeyer, H.; Yurdaydin, C.; Hardtke, S.; Caruntu, F.A.; Curescu, M.G.; Yalcin, K.; Akarca, U.S.; Gurel, S.; Zeuzem, S.; Erhardt, A.; et al. Peginterferon alfa-2a plus tenofovir disoproxil fumarate for hepatitis D (HIDIT-II): A randomised, placebo controlled, phase 2 trial. *Lancet Infect. Dis.* **2019**, *19*, 275–286. [CrossRef]
6. Heidrich, B.; Yurdaydin, C.; Kabacam, G.; Ratsch, B.A.; Zachou, K.; Bremer, B.; Dalekos, G.N.; Erhardt, A.; Tabak, F.; Yalcin, K.; et al. Late HDV RNA relapse after peginterferon alpha-based therapy of chronic hepatitis delta. *Hepatology* **2014**, *60*, 87–97. [CrossRef]
7. Bogomolov, P.; Alexandrov, A.; Voronkova, N.; Macievich, M.; Kokina, K.; Petrachenkova, M.; Lehr, T.; Lempp, F.A.; Wedemeyer, H.; Haag, M.; et al. Treatment of chronic hepatitis D with the entry inhibitor myrcludex B: First results of a phase Ib/IIa study. *J. Hepatol.* **2016**, *65*, 490–498. [CrossRef]

8. Beilstein, F.; Blanchet, M.; Vaillant, A.; Sureau, C. Nucleic acid polymers are active against hepatitis delta virus infection in vitro. *J. Virol.* **2018**, *92*. [CrossRef]
9. Zhang, Z.; Filzmayer, C.; Ni, Y.; Sultmann, H.; Mutz, P.; Hiet, M.S.; Vondran, F.W.R.; Bartenschlager, R.; Urban, S. Hepatitis D virus replication is sensed by MDA5 and induces IFN-beta/lambda responses in hepatocytes. *J. Hepatol.* **2018**, *69*, 25–35. [CrossRef]
10. Bazinet, M.; Pantea, V.; Cebotarescu, V.; Cojuhari, L.; Jimbei, P.; Albrecht, J.; Schmid, P.; Le Gal, F.; Gordien, E.; Krawczyk, A.; et al. Safety and efficacy of REP 2139 and pegylated interferon alfa-2a for treatment-naive patients with chronic hepatitis B virus and hepatitis D virus co-infection (REP 301 and REP 301-LTF): A non-randomised, open-label, phase 2 trial. *Lancet Gastroenterol. Hepatol.* **2017**, *2*, 877–889. [CrossRef]
11. Lempp, F.A.; Urban, S. Hepatitis delta virus: Replication strategy and upcoming therapeutic options for a neglected human pathogen. *Viruses* **2017**, *9*, 172. [CrossRef] [PubMed]
12. Lai, M.M. The molecular biology of hepatitis delta virus. *Annu. Rev. Biochem.* **1995**, *64*, 259–286. [CrossRef] [PubMed]
13. Huang, C.R.; Lo, S.J. Evolution and diversity of the human hepatitis d virus genome. *Adv. Bioinform.* **2010**. [CrossRef] [PubMed]
14. Weiner, A.J.; Choo, Q.L.; Wang, K.S.; Govindarajan, S.; Redeker, A.G.; Gerin, J.L.; Houghton, M. A single antigenomic open reading frame of the hepatitis delta virus encodes the epitope(s) of both hepatitis delta antigen polypeptides p24 delta and p27 delta. *J. Virol.* **1988**, *62*, 594–599. [CrossRef]
15. Casey, J.L. RNA editing in hepatitis delta virus. *Curr. Top. Microbiol. Immunol.* **2006**, *307*, 67–89.
16. Le Gal, F.; Gault, E.; Ripault, M.P.; Serpaggi, J.; Trinchet, J.C.; Gordien, E.; Deny, P. Eighth major clade for hepatitis delta virus. *Emerg. Infect. Dis.* **2006**, *12*, 1447–1450. [CrossRef]
17. Karimzadeh, H.; Usman, Z.; Frishman, D.; Roggendorf, M. Genetic diversity of hepatitis D virus genotype-1 in Europe allows classification into subtypes. *J. Viral Hepat.* **2019**. [CrossRef]
18. Le Gal, F.; Brichler, S.; Drugan, T.; Alloui, C.; Roulot, D.; Pawlotsky, J.M.; Deny, P.; Gordien, E. Genetic diversity and worldwide distribution of the deltavirus genus: A study of 2,152 clinical strains. *Hepatology* **2017**, *66*, 1826–1841. [CrossRef]
19. Paul Dény, C.D.; Gatherer, D.; Kay, A.C.; Le Gal, F.; Drugan, T. Hepatitis delta virus clades and genotypes: A practical approach. In *Hepatitis D. Virology, Management and Methodology*; Mario Rizzetto, A.S., Ed.; Il Pensiero Scientifico: Rome, Italy, 2019.
20. Ivaniushina, V.; Radjef, N.; Alexeeva, M.; Gault, E.; Semenov, S.; Salhi, M.; Kiselev, O.; Deny, P. Hepatitis delta virus genotypes I and II cocirculate in an endemic area of Yakutia, Russia. *J. Gen. Virol.* **2001**, *82*, 2709–2718. [CrossRef]
21. Lee, C.M.; Changchien, C.S.; Chung, J.C.; Liaw, Y.F. Characterization of a new genotype II hepatitis delta virus from Taiwan. *J. Med. Virol.* **1996**, *49*, 145–154. [CrossRef]
22. Sakugawa, H.; Nakasone, H.; Nakayoshi, T.; Kawakami, Y.; Miyazato, S.; Kinjo, F.; Saito, A.; Ma, S.P.; Hotta, H.; Kinoshita, M. Hepatitis delta virus genotype IIb predominates in an endemic area, Okinawa, Japan. *J. Med. Virol.* **1999**, *58*, 366–372. [CrossRef]
23. Marino, Q.; Cisse, M.; Gerber, A.; Dolo, O.; Sayon, S.; Ba, A.; Brichler, S.; Tata Traore, F.; Gordien, E.; Togo, J.; et al. Low hepatitis D seroprevalence in blood donors of Bamako, Mali. *Infect. Dis. (Lond.)* **2019**, *51*, 622–624. [CrossRef] [PubMed]
24. Casey, J.L.; Brown, T.L.; Colan, E.J.; Wignall, F.S.; Gerin, J.L. A genotype of hepatitis D virus that occurs in northern South America. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 9016–9020. [CrossRef]
25. Gomes-Gouvea, M.S.; Soares, M.C.; Bensabath, G.; de Carvalho-Mello, I.M.; Brito, E.M.; Souza, O.S.; Queiroz, A.T.; Carrilho, F.J.; Pinho, J.R. Hepatitis B virus and hepatitis delta virus genotypes in outbreaks of fulminant hepatitis (Labrea black fever) in the western Brazilian Amazon region. *J. Gen. Virol.* **2009**, *90*, 2638–2643. [CrossRef]
26. Quintero, A.; Uzcategui, N.; Loureiro, C.L.; Villegas, L.; Illarramendi, X.; Guevara, M.E.; Ludert, J.E.; Blitz, L.; Liprandi, F.; Pujol, F.H. Hepatitis delta virus genotypes I and III circulate associated with hepatitis B virus genotype F in Venezuela. *J. Med. Virol.* **2001**, *64*, 356–359. [CrossRef]
27. Scarponi, C.F.O.; Silva, R.D.N.D.; Souza Filho, J.A.; Guerra, M.R.L.; Pedrosa, M.A.F.; Mol, M.P.G. Hepatitis delta prevalence in South America: A systematic review and meta-analysis. *Rev. Soc. Bras. Med. Trop.* **2019**, *52*, e20180289. [CrossRef]

28. di Filippo Villa, D.; Cortes-Mancera, F.; Payares, E.; Montes, N.; de la Hoz, F.; Arbelaez, M.P.; Correa, G.; Navas, M.C. Hepatitis D virus and hepatitis B virus infection in Amerindian communities of the Amazonas state, Colombia. *Virol. J.* **2015**, *12*, 172. [CrossRef]
29. Makuwa, M.; Caron, M.; Souquiere, S.; Malonga-Mouelet, G.; Mahe, A.; Kazanji, M. Prevalence and genetic diversity of hepatitis B and delta viruses in pregnant women in Gabon: Molecular evidence that hepatitis delta virus clade 8 originates from and is endemic in central Africa. *J. Clin. Microbiol.* **2008**, *46*, 754–756. [CrossRef]
30. Radjef, N.; Gordien, E.; Ivaniushina, V.; Gault, E.; Anais, P.; Drugan, T.; Trinchet, J.C.; Roulot, D.; Tamby, M.; Milinkovitch, M.C.; et al. Molecular phylogenetic analyses indicate a wide and ancient radiation of African hepatitis delta virus, suggesting a deltavirus genus of at least seven major clades. *J. Virol.* **2004**, *78*, 2537–2544. [CrossRef]
31. Santos, M.D.; Gomes-Gouvêa, M.S.; Nunes, J.D.; Barros, L.M.; Carrilho, F.J.; Ferreira, A.e.S.; Pinho, J.R. The hepatitis delta genotype 8 in Northeast Brazil: The North Atlantic slave trade as the potential route for infection. *Virus Res.* **2016**, *224*, 6–11. [CrossRef]
32. Niro, G.A.; Smedile, A.; Andriulli, A.; Rizzetto, M.; Gerin, J.L.; Casey, J.L. The predominance of hepatitis delta virus genotype I among chronically infected Italian patients. *Hepatology* **1997**, *25*, 728–734. [CrossRef] [PubMed]
33. Su, C.W.; Huang, Y.H.; Huo, T.I.; Shih, H.H.; Sheen, I.J.; Chen, S.W.; Lee, P.C.; Lee, S.D.; Wu, J.C. Genotypes and viremia of hepatitis B and D viruses are associated with outcomes of chronic hepatitis D patients. *Gastroenterology* **2006**, *130*, 1625–1635. [CrossRef] [PubMed]
34. Barros, L.M.; Gomes-Gouvea, M.S.; Pinho, J.R.; Alvarado-Mora, M.V.; Dos Santos, A.; Mendes-Correa, M.C.; Caldas, A.J.; Sousa, M.T.; Santos, M.D.; Ferreira, A.S. Hepatitis Delta virus genotype 8 infection in Northeast Brazil: Inheritance from African slaves? *Virus Res.* **2011**, *160*, 333–339. [CrossRef]
35. Chang, W.S.; Pettersson, J.H.; Le Lay, C.; Shi, M.; Lo, N.; Wille, M.; Eden, J.S.; Holmes, E.C. Novel hepatitis D-like agents in vertebrates and invertebrates. *Virus Evol.* **2019**, *5*, vez021. [CrossRef]
36. Wille, M.; Netter, H.J.; Littlejohn, M.; Yuen, L.; Shi, M.; Eden, J.S.; Klaassen, M.; Holmes, E.C.; Hurt, A.C. A divergent hepatitis D-like agent in birds. *Viruses* **2018**, *10*, 720. [CrossRef] [PubMed]
37. Hetzel, U.; Szirovicza, L.; Smura, T.; Prahauser, B.; Vapalahti, O.; Kipar, A.; Hepojoki, J. Identification of a novel deltavirus in boa constrictors. *mBio* **2019**, *10*. [CrossRef]
38. Perez-Vargas, J.; Amirache, F.; Boson, B.; Mialon, C.; Freitas, N.; Sureau, C.; Fusil, F.; Cosset, F.L. Enveloped viruses distinct from HBV induce dissemination of hepatitis D virus in vivo. *Nat. Commun.* **2019**, *10*, 2098. [CrossRef]
39. Kuiken, C.; Korber, B.; Shafer, R.W. HIV sequence databases. *Aids Rev.* **2003**, *5*, 52–61.
40. Hayer, J.; Jadeau, F.; Deleage, G.; Kay, A.; Zoulim, F.; Combet, C. HBVdb: A knowledge database for Hepatitis B Virus. *Nucleic Acids Res.* **2013**, *41*, D566–D570. [CrossRef]
41. Kuiken, C.; Hraber, P.; Thurmond, J.; Yusim, K. The hepatitis C sequence database in Los Alamos. *Nucleic Acids Res.* **2008**, *36*, D512–D516. [CrossRef] [PubMed]
42. Kefalakes, H.; Koh, C.; Sidney, J.; Amanakis, G.; Sette, A.; Heller, T.; Rehermann, B. Hepatitis D Virus-specific CD8(+) T cells have a memory-like phenotype associated with viral immune escape in patients with chronic hepatitis D virus infection. *Gastroenterology* **2019**, *156*, 1805–1819. [CrossRef] [PubMed]
43. Karimzadeh, H.; Kiraithe, M.M.; Oberhardt, V.; Salimi Alizei, E.; Bockmann, J.; Schulze Zur Wiesch, J.; Budeus, B.; Hoffmann, D.; Wedemeyer, H.; Cornberg, M.; et al. Mutations in hepatitis D virus allow it to escape detection by CD8(+) T cells and evolve at the population level. *Gastroenterology* **2019**, *156*, 1820–1833. [CrossRef] [PubMed]
44. Benson, D.A.; Clark, K.; Karsch-Mizrachi, I.; Lipman, D.J.; Ostell, J.; Sayers, E.W. GenBank. *Nucleic Acids Res.* **2015**, *43*, D30–D35. [CrossRef]
45. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [CrossRef]
46. Edgar, R.C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797. [CrossRef] [PubMed]
47. Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T.J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Soding, J.; et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **2011**, *7*, 539. [CrossRef] [PubMed]

48. Untergasser, A.; Cutcutache, I.; Koressaar, T.; Ye, J.; Faircloth, B.C.; Remm, M.; Rozen, S.G. Primer3—New capabilities and interfaces. *Nucleic Acids Res.* **2012**, *40*, e115. [CrossRef] [PubMed]
49. Guindon, S.; Dufayard, J.F.; Lefort, V.; Anisimova, M.; Hordijk, W.; Gascuel, O. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst. Biol.* **2010**, *59*, 307–321. [CrossRef]

D Eidesstattliche Erklärung

TUM School of Life Sciences Technische Universität München

Name der promotionsführenden Einrichtung

Evolution of hepatotropic viruses during chronic infection in humans

Titel der wissenschaftlichen Abhandlung

Zainab Usman

Vorname und Name

Vollständiger Abdruck der von der promotionsführenden Einrichtung
der TUM School of Life Sciences
der Technischen Universität München zur Erlangung des akademischen Grades
eines Doktors **der Naturwissenschaften (Dr. rer. nat.)**
genehmigten Dissertation.

Vorsitzende/-r: Prof. Dr. Johan Philipp Benz

Prüfende/-r der Dissertation:

1. Prof. Dr. Dmitrij Frishmann

2. Prof. Dr. Ulrike Protzer

3.

Die Dissertation wurde am _____ bei der TUM School of Life Sciences Technischen Universität München
eingereicht und durch die promotionsführende Einrichtung
_____ angenommen.

Anhang I

Eidesstattliche Erklärung

Ich erkläre an Eides statt, dass ich die bei der promotionsführenden Einrichtung

Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt

der TUM zur Promotionsprüfung vorgelegte Arbeit mit dem Titel:

Evolution of hepatrophic viruses during chronic infection in humans

in Department of Bioinformatics

Fakultät, Institut, Lehrstuhl, Klinik, Krankenhaus, Abteilung

unter der Anleitung und Betreuung durch: Prof. Dr. Dmitrij Frishman ohne sonstige Hilfe erstellt und bei der Abfassung nur die gemäß § 6 Ab. 6 und 7 Satz 2 angebotenen Hilfsmittel benutzt habe.

☒ Ich habe keine Organisation eingeschaltet, die gegen Entgelt Betreuerinnen und Betreuer für die Anfertigung von Dissertationen sucht, oder die mir obliegenden Pflichten hinsichtlich der Prüfungsleistungen für mich ganz oder teilweise erledigt.

☒ Ich habe die Dissertation in dieser oder ähnlicher Form in keinem anderen Prüfungsverfahren als Prüfungsleistung vorgelegt.

☐ Die vollständige Dissertation wurde in _____ veröffentlicht. Die promotionsführende Einrichtung

_____

hat der Veröffentlichung zugestimmt.

☒ Ich habe den angestrebten Doktorgrad noch nicht erworben und bin nicht in einem früheren Promotionsverfahren für den angestrebten Doktorgrad endgültig gescheitert.

☐ Ich habe bereits am _____ bei der Fakultät für _____

_____

der Hochschule _____

unter Vorlage einer Dissertation mit dem Thema _____

_____

die Zulassung zur Promotion beantragt mit dem Ergebnis: _____

_____

Die öffentlich zugängliche Promotionsordnung der TUM ist mir bekannt, insbesondere habe ich die Bedeutung von § 28 (Nichtigkeit der Promotion) und § 29 (Entzug des Doktorgrades) zur Kenntnis genommen. Ich bin mir der Konsequenzen einer falschen Eidesstattlichen Erklärung bewusst.

Mit der Aufnahme meiner personenbezogenen Daten in die Alumni-Datei bei der TUM bin ich

☒ einverstanden, ☐ nicht einverstanden.

Freising, 20.05.2020, Unterschrift

E Publishers Consent for Article I

JOHN WILEY AND SONS LICENSE
TERMS AND CONDITIONS

May 12, 2020

This Agreement between Wissenschaftszentrum Weihenstephan, ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

License Number	4814820349629
License date	Apr 23, 2020
Licensed Content Publisher	John Wiley and Sons
Licensed Content Publication	Journal of Viral Hepatitis
Licensed Content Title	Kinetics of hepatitis B surface antigen quasispecies during REP 2139-Ca therapy in HBeAg-positive chronic HBV infection
Licensed Content Author	Michael Roggendorf, Andrew Vaillant, Dmitrij Frishman, et al
Licensed Content Date	Aug 13, 2019
Licensed Content Volume	26
Licensed Content Issue	12
Licensed Content Pages	11

Type of use	Dissertation/Thesis
Requestor type	Author of this Wiley article
Format	Print and electronic
Portion	Full article
Will you be translating?	No
Title	Phd student
Institution name	Technical University Munich
Expected presentation date	Nov 2020
Requestor Location	Department of Bioinformatics Wissenschaftszentrum Weihenstephan, Maximus-von-Imhof-Forum 3, Technical University Munich Freising, Bayern 85354 Germany Attn: Zainab Usman
Publisher Tax ID	EU826007151
Total	0.00 USD

Terms and Conditions

TERMS AND CONDITIONS

This copyrighted material is owned by or exclusively licensed to John Wiley & Sons, Inc. or one of its group companies (each a"Wiley Company") or handled on behalf of a society with which a Wiley Company has exclusive publishing rights in relation to a particular work (collectively "WILEY"). By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the billing and payment terms and conditions established by the Copyright Clearance Center Inc., ("CCC's Billing and Payment terms and conditions"), at the time that

you opened your RightsLink account (these are available at any time at http://myaccount.copyright.com).

Terms and Conditions

- The materials you have requested permission to reproduce or reuse (the "Wiley Materials") are protected by copyright.
- You are hereby granted a personal, non-exclusive, non-sub licensable (on a stand-alone basis), non-transferable, worldwide, limited license to reproduce the Wiley Materials for the purpose specified in the licensing process. This license, **and any CONTENT (PDF or image file) purchased as part of your order,** is for a one-time use only and limited to any maximum distribution number specified in the license. The first instance of republication or reuse granted by this license must be completed within two years of the date of the grant of this license (although copies prepared before the end date may be distributed thereafter). The Wiley Materials shall not be used in any other manner or for any other purpose, beyond what is granted in the license. Permission is granted subject to an appropriate acknowledgement given to the author, title of the material/book/journal and the publisher. You shall also duplicate the copyright notice that appears in the Wiley publication in your use of the Wiley Material. Permission is also granted on the understanding that nowhere in the text is a previously published source acknowledged for all or part of this Wiley Material. Any third party content is expressly excluded from this permission.
- With respect to the Wiley Materials, all rights are reserved. Except as expressly granted by the terms of the license, no part of the Wiley Materials may be copied, modified, adapted (except for minor reformatting required by the new Publication), translated, reproduced, transferred or distributed, in any form or by any means, and no derivative works may be made based on the Wiley Materials without the prior permission of the respective copyright owner.**For STM Signatory Publishers clearing permission under the terms of the STM Permissions Guidelines only, the terms of the license are extended to include subsequent editions and for editions in other languages, provided such editions are for the work as a whole in situ and does not involve the separate exploitation of the permitted figures or extracts,** You may not alter, remove or suppress in any manner any copyright, trademark or other notices displayed by the Wiley Materials. You may not license, rent, sell, loan, lease, pledge, offer as security, transfer or assign the Wiley Materials on a stand-alone basis, or any of the rights granted to you hereunder to any other person.
- The Wiley Materials and all of the intellectual property rights therein shall at all times remain the exclusive property of John Wiley & Sons Inc, the Wiley Companies, or their respective licensors, and your interest therein is only that of having possession of and the right to reproduce the Wiley Materials pursuant to Section 2 herein during the continuance of this Agreement. You agree that you own no right, title or interest in or to the Wiley Materials or any of the intellectual property rights therein. You shall have no rights hereunder other than the license as provided for above in Section 2. No right, license or interest to any trademark, trade name, service mark or other branding ("Marks") of WILEY or its licensors is granted hereunder, and you agree that you shall not assert any such right, license or interest with respect thereto
- NEITHER WILEY NOR ITS LICENSORS MAKES ANY WARRANTY OR REPRESENTATION OF ANY KIND TO YOU OR ANY THIRD PARTY, EXPRESS, IMPLIED OR STATUTORY, WITH RESPECT TO THE MATERIALS

OR THE ACCURACY OF ANY INFORMATION CONTAINED IN THE MATERIALS, INCLUDING, WITHOUT LIMITATION, ANY IMPLIED WARRANTY OF MERCHANTABILITY, ACCURACY, SATISFACTORY QUALITY, FITNESS FOR A PARTICULAR PURPOSE, USABILITY, INTEGRATION OR NON-INFRINGEMENT AND ALL SUCH WARRANTIES ARE HEREBY EXCLUDED BY WILEY AND ITS LICENSORS AND WAIVED BY YOU.

- WILEY shall have the right to terminate this Agreement immediately upon breach of this Agreement by you.
- You shall indemnify, defend and hold harmless WILEY, its Licensors and their respective directors, officers, agents and employees, from and against any actual or threatened claims, demands, causes of action or proceedings arising from any breach of this Agreement by you.
- IN NO EVENT SHALL WILEY OR ITS LICENSORS BE LIABLE TO YOU OR ANY OTHER PARTY OR ANY OTHER PERSON OR ENTITY FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL, INDIRECT, EXEMPLARY OR PUNITIVE DAMAGES, HOWEVER CAUSED, ARISING OUT OF OR IN CONNECTION WITH THE DOWNLOADING, PROVISIONING, VIEWING OR USE OF THE MATERIALS REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE (INCLUDING, WITHOUT LIMITATION, DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE, BUSINESS OPPORTUNITY OR CLAIMS OF THIRD PARTIES), AND WHETHER OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS LIMITATION SHALL APPLY NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.
- Should any provision of this Agreement be held by a court of competent jurisdiction to be illegal, invalid, or unenforceable, that provision shall be deemed amended to achieve as nearly as possible the same economic effect as the original provision, and the legality, validity and enforceability of the remaining provisions of this Agreement shall not be affected or impaired thereby.
- The failure of either party to enforce any term or condition of this Agreement shall not constitute a waiver of either party's right to enforce each and every term and condition of this Agreement. No breach under this agreement shall be deemed waived or excused by either party unless such waiver or consent is in writing signed by the party granting such waiver or consent. The waiver by or consent of a party to a breach of any provision of this Agreement shall not operate or be construed as a waiver of or consent to any other or subsequent breach by such other party.
- This Agreement may not be assigned (including by operation of law or otherwise) by you without WILEY's prior written consent.
- Any fee required for this permission shall be non-refundable after thirty (30) days from receipt by the CCC.
- These terms and conditions together with CCC's Billing and Payment terms and conditions (which are incorporated herein) form the entire agreement between you and WILEY concerning this licensing transaction and (in the absence of fraud) supersedes

all prior agreements and representations of the parties, oral or written. This Agreement may not be amended except in writing signed by both parties. This Agreement shall be binding upon and inure to the benefit of the parties' successors, legal representatives, and authorized assigns.

- In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall prevail.
- WILEY expressly reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.
- This Agreement will be void if the Type of Use, Format, Circulation, or Requestor Type was misrepresented during the licensing process.
- This Agreement shall be governed by and construed in accordance with the laws of the State of New York, USA, without regards to such state's conflict of law rules. Any legal action, suit or proceeding arising out of or relating to these Terms and Conditions or the breach thereof shall be instituted in a court of competent jurisdiction in New York County in the State of New York in the United States of America and each party hereby consents and submits to the personal jurisdiction of such court, waives any objection to venue in such court and consents to service of process by registered or certified mail, return receipt requested, at the last known address of such party.

WILEY OPEN ACCESS TERMS AND CONDITIONS

Wiley Publishes Open Access Articles in fully Open Access Journals and in Subscription journals offering Online Open. Although most of the fully Open Access journals publish open access articles under the terms of the Creative Commons Attribution (CC BY) License only, the subscription journals and a few of the Open Access Journals offer a choice of Creative Commons Licenses. The license type is clearly identified on the article.

The Creative Commons Attribution License

The Creative Commons Attribution License (CC-BY) allows users to copy, distribute and transmit an article, adapt the article and make commercial use of the article. The CC-BY license permits commercial and non-

Creative Commons Attribution Non-Commercial License

The Creative Commons Attribution Non-Commercial (CC-BY-NC)License permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.(see below)

Creative Commons Attribution-Non-Commercial-NoDerivs License

The Creative Commons Attribution Non-Commercial-NoDerivs License (CC-BY-NC-ND) permits use, distribution and reproduction in any medium, provided the original work is properly cited, is not used for commercial purposes and no modifications or adaptations are made. (see below)

Use by commercial "for-profit" organizations

Use of Wiley Open Access articles for commercial, promotional, or marketing purposes requires further explicit permission from Wiley and will be subject to a fee.

Further details can be found on Wiley Online Library
http://olabout.wiley.com/WileyCDA/Section/id-410895.html

Other Terms and Conditions:

v1.10 Last updated September 2015

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

F Publishers Consent for Article II

JOHN WILEY AND SONS LICENSE
TERMS AND CONDITIONS

May 12, 2020

This Agreement between Zainab Usman ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

License Number	4814820195299
License date	Apr 23, 2020
Licensed Content Publisher	John Wiley and Sons
Licensed Content Publication	Journal of Viral Hepatitis
Licensed Content Title	Genetic diversity of hepatitis D virus genotype-1 in Europe allows classification into subtypes
Licensed Content Author	Michael Roggendorf, Dmitrij Frishman, Zainab Usman, et al
Licensed Content Date	Mar 15, 2019
Licensed Content Volume	26
Licensed Content Issue	7
Licensed Content Pages	11

Type of use	Dissertation/Thesis
Requestor type	Author of this Wiley article
Format	Print and electronic
Portion	Full article
Will you be translating?	No
Title	Phd student
Institution name	Technical University Munich
Expected presentation date	Nov 2020
Requestor Location	Department of Bioinformatics Wissenschaftszentrum Weihenstephan, Maximus-von-Imhof-Forum 3, Technical University Muenchen Freising, Bayern 85354 Germany Attn: Zainab Usman
Publisher Tax ID	EU826007151
Billing Type	Invoice
Billing Address	Wissenschaftszentrum Weihenstephan, Wissenschaftszentrum Weihenstephan, TU Muenchen,Maximus-von-Imhof-Forum 3, Freising, Germany 85354 Attn: Wissenschaftszentrum Weihenstephan,
Total	0.00 EUR

Terms and Conditions

TERMS AND CONDITIONS

This copyrighted material is owned by or exclusively licensed to John Wiley & Sons, Inc. or one of its group companies (each a"Wiley Company") or handled on behalf of a society with which a Wiley Company has exclusive publishing rights in relation to a particular work (collectively "WILEY"). By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the billing and payment terms and conditions established by the Copyright Clearance Center Inc., ("CCC's Billing and Payment terms and conditions"), at the time that you opened your RightsLink account (these are available at any time at http://myaccount.copyright.com).

Terms and Conditions

- The materials you have requested permission to reproduce or reuse (the "Wiley Materials") are protected by copyright.

- You are hereby granted a personal, non-exclusive, non-sub licensable (on a stand-alone basis), non-transferable, worldwide, limited license to reproduce the Wiley Materials for the purpose specified in the licensing process. This license, **and any CONTENT (PDF or image file) purchased as part of your order,** is for a one-time use only and limited to any maximum distribution number specified in the license. The first instance of republication or reuse granted by this license must be completed within two years of the date of the grant of this license (although copies prepared before the end date may be distributed thereafter). The Wiley Materials shall not be used in any other manner or for any other purpose, beyond what is granted in the license. Permission is granted subject to an appropriate acknowledgement given to the author, title of the material/book/journal and the publisher. You shall also duplicate the copyright notice that appears in the Wiley publication in your use of the Wiley Material. Permission is also granted on the understanding that nowhere in the text is a previously published source acknowledged for all or part of this Wiley Material. Any third party content is expressly excluded from this permission.

- With respect to the Wiley Materials, all rights are reserved. Except as expressly granted by the terms of the license, no part of the Wiley Materials may be copied, modified, adapted (except for minor reformatting required by the new Publication), translated, reproduced, transferred or distributed, in any form or by any means, and no derivative works may be made based on the Wiley Materials without the prior permission of the respective copyright owner.**For STM Signatory Publishers clearing permission under the terms of the STM Permissions Guidelines only, the terms of the license are extended to include subsequent editions and for editions in other languages, provided such editions are for the work as a whole in situ and does not involve the separate exploitation of the permitted figures or extracts,** You may not alter, remove or suppress in any manner any copyright, trademark or other notices displayed by the Wiley Materials. You may not license, rent, sell, loan, lease, pledge, offer as security, transfer or assign the Wiley Materials on a stand-alone basis, or any of the rights granted to you hereunder to any other person.

- The Wiley Materials and all of the intellectual property rights therein shall at all times remain the exclusive property of John Wiley & Sons Inc, the Wiley Companies, or their respective licensors, and your interest therein is only that of having possession of and the right to reproduce the Wiley Materials pursuant to Section 2 herein during the continuance of this Agreement. You agree that you own no right, title or interest in or

to the Wiley Materials or any of the intellectual property rights therein. You shall have no rights hereunder other than the license as provided for above in Section 2. No right, license or interest to any trademark, trade name, service mark or other branding ("Marks") of WILEY or its licensors is granted hereunder, and you agree that you shall not assert any such right, license or interest with respect thereto

- NEITHER WILEY NOR ITS LICENSORS MAKES ANY WARRANTY OR REPRESENTATION OF ANY KIND TO YOU OR ANY THIRD PARTY, EXPRESS, IMPLIED OR STATUTORY, WITH RESPECT TO THE MATERIALS OR THE ACCURACY OF ANY INFORMATION CONTAINED IN THE MATERIALS, INCLUDING, WITHOUT LIMITATION, ANY IMPLIED WARRANTY OF MERCHANTABILITY, ACCURACY, SATISFACTORY QUALITY, FITNESS FOR A PARTICULAR PURPOSE, USABILITY, INTEGRATION OR NON-INFRINGEMENT AND ALL SUCH WARRANTIES ARE HEREBY EXCLUDED BY WILEY AND ITS LICENSORS AND WAIVED BY YOU.

- WILEY shall have the right to terminate this Agreement immediately upon breach of this Agreement by you.

- You shall indemnify, defend and hold harmless WILEY, its Licensors and their respective directors, officers, agents and employees, from and against any actual or threatened claims, demands, causes of action or proceedings arising from any breach of this Agreement by you.

- IN NO EVENT SHALL WILEY OR ITS LICENSORS BE LIABLE TO YOU OR ANY OTHER PARTY OR ANY OTHER PERSON OR ENTITY FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL, INDIRECT, EXEMPLARY OR PUNITIVE DAMAGES, HOWEVER CAUSED, ARISING OUT OF OR IN CONNECTION WITH THE DOWNLOADING, PROVISIONING, VIEWING OR USE OF THE MATERIALS REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE (INCLUDING, WITHOUT LIMITATION, DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE, BUSINESS OPPORTUNITY OR CLAIMS OF THIRD PARTIES), AND WHETHER OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS LIMITATION SHALL APPLY NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.

- Should any provision of this Agreement be held by a court of competent jurisdiction to be illegal, invalid, or unenforceable, that provision shall be deemed amended to achieve as nearly as possible the same economic effect as the original provision, and the legality, validity and enforceability of the remaining provisions of this Agreement shall not be affected or impaired thereby.

- The failure of either party to enforce any term or condition of this Agreement shall not constitute a waiver of either party's right to enforce each and every term and condition of this Agreement. No breach under this agreement shall be deemed waived or excused by either party unless such waiver or consent is in writing signed by the party granting such waiver or consent. The waiver by or consent of a party to a breach of any provision of this Agreement shall not operate or be construed as a waiver of or consent to any other or subsequent breach by such other party.

- This Agreement may not be assigned (including by operation of law or otherwise) by you without WILEY's prior written consent.
- Any fee required for this permission shall be non-refundable after thirty (30) days from receipt by the CCC.
- These terms and conditions together with CCC's Billing and Payment terms and conditions (which are incorporated herein) form the entire agreement between you and WILEY concerning this licensing transaction and (in the absence of fraud) supersedes all prior agreements and representations of the parties, oral or written. This Agreement may not be amended except in writing signed by both parties. This Agreement shall be binding upon and inure to the benefit of the parties' successors, legal representatives, and authorized assigns.
- In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall prevail.
- WILEY expressly reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.
- This Agreement will be void if the Type of Use, Format, Circulation, or Requestor Type was misrepresented during the licensing process.
- This Agreement shall be governed by and construed in accordance with the laws of the State of New York, USA, without regards to such state's conflict of law rules. Any legal action, suit or proceeding arising out of or relating to these Terms and Conditions or the breach thereof shall be instituted in a court of competent jurisdiction in New York County in the State of New York in the United States of America and each party hereby consents and submits to the personal jurisdiction of such court, waives any objection to venue in such court and consents to service of process by registered or certified mail, return receipt requested, at the last known address of such party.

WILEY OPEN ACCESS TERMS AND CONDITIONS

Wiley Publishes Open Access Articles in fully Open Access Journals and in Subscription journals offering Online Open. Although most of the fully Open Access journals publish open access articles under the terms of the Creative Commons Attribution (CC BY) License only, the subscription journals and a few of the Open Access Journals offer a choice of Creative Commons Licenses. The license type is clearly identified on the article.

The Creative Commons Attribution License

The Creative Commons Attribution License (CC-BY) allows users to copy, distribute and transmit an article, adapt the article and make commercial use of the article. The CC-BY license permits commercial and non-

Creative Commons Attribution Non-Commercial License

The Creative Commons Attribution Non-Commercial (CC-BY-NC)License permits use, distribution and reproduction in any medium, provided the original work is properly cited

and is not used for commercial purposes.(see below)

Creative Commons Attribution-Non-Commercial-NoDerivs License

The Creative Commons Attribution Non-Commercial-NoDerivs License (CC-BY-NC-ND) permits use, distribution and reproduction in any medium, provided the original work is properly cited, is not used for commercial purposes and no modifications or adaptations are made. (see below)

Use by commercial "for-profit" organizations

Use of Wiley Open Access articles for commercial, promotional, or marketing purposes requires further explicit permission from Wiley and will be subject to a fee.

Further details can be found on Wiley Online Library
http://olabout.wiley.com/WileyCDA/Section/id-410895.html

Other Terms and Conditions:

v1.10 Last updated September 2015

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

G Publishers Consent for Article III

viruses

an Open Access Journal by MDPI

Indexed in: PubMed | CITESCORE 4.03 SCOPUS | IMPACT FACTOR 3.811

CERTIFICATE OF PUBLICATION

Certificate of publication for the article titled:

HDVdb: A Comprehensive Hepatitis D Virus Database

Authored by:

Zainab Usman; Stoyan Velkov; Ulrike Protzer; Michael Roggendorf; Dmitrij Frishman; Hadi Karimzadeh

Published in:

Viruses **2020**, Volume 12, Issue 5, 538

MDPI Academic Open Access Publishing since 1996

Basel, May 2020

H First Author Consent for Article II

HelmholtzZentrum münchen
Deutsches Forschungszentrum für Gesundheit und Umwelt

Institut für Virologie-Technische Universität München/Helmholtz-Zentrum München GmbH
Schneckenburgerstraße 8 . 81675 München . Germany

Technische Universität München

Institut für Virologie
Prof. Dr. med. Michael Roggendorf

Schneckenburgerstraße 8,
81675 München Germany

Tel: +49.89.4140.7360
Fax: +49.89.4140.7444

michael.roggendorf@tum.de

20th April 2020

To Whom It May Concern

Subject: No objection letter for use of publication

Dear Madam/Sir:

This is to certify that Ms. Zainab Usman, co-authored the research article "***Genetic diversity of hepatitis D virus genotype-1 in Europe allows classification into subtypes***" (PMID: 30801877) published in a peer-reviewed scientific journal "***Journal of Viral Hepatitis***" while she was working on her Doctoral studies under supervision of Prof. Dmitrij Frishman at Department of Bioinformatics, Technical University Munich.

It is noteworthy to mention contribution of all authors. HK and MR conceived the original idea of the study. HK collected the samples and executed wet lab experiments. ZU created the computational workflow and performed bioinformatics analysis. DF and MR supervised the project. ZU, HK, DF and MR interpreted the results and wrote the manuscript.

In view of her compelling contributions to the research article, we do hereby confirm that we have no objection to the use of aforementioned research and article material (text or figures) in her PhD dissertation.

Sincerely,

Prof. Dr. med. Michael Roggendorf, 19.04.2020
Signature & Date Corresponding author

Dr. Hadi Karimzadeh, 20.04.2020
Signature & Date First author