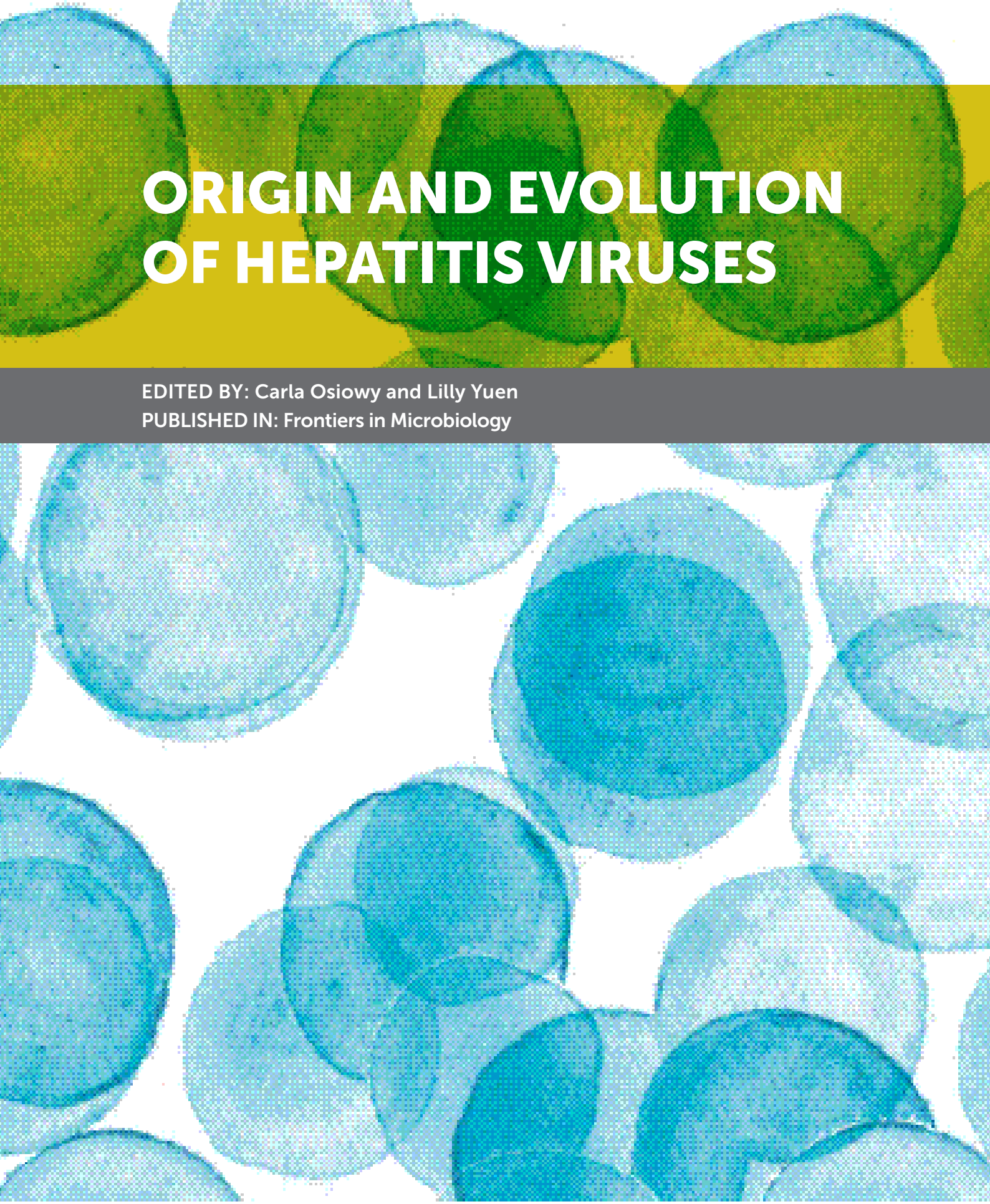

ORIGIN AND EVOLUTION OF HEPATITIS VIRUSES

EDITED BY: Carla Osiowy and Lilly Yuen
PUBLISHED IN: Frontiers in Microbiology

frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88971-488-9
DOI 10.3389/978-2-88971-488-9

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

ORIGIN AND EVOLUTION OF HEPATITIS VIRUSES

Topic Editors:
Carla Osiowy, Public Health Agency of Canada (PHAC), Canada
Lilly Yuen, Victorian Infectious Diseases Reference Laboratory, Australia

Citation: Osiowy, C., Yuen, L., eds. (2021). Origin and Evolution of Hepatitis Viruses. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88971-488-9

Table of Contents

frontiers in Microbiology

EDITORIAL
published: 25 August 2021
doi: 10.3389/fmicb.2021.740255

Editorial: Origin and Evolution of Hepatitis Viruses

Carla Osiowy[1]* *and Lilly Yuen*[2]

[1] National Microbiology Laboratory, Public Health Agency of Canada, Winnipeg, MB, Canada, [2] Victorian Infectious Diseases Reference Laboratory, Melbourne Health, Peter Doherty Institute, Melbourne, VIC, Australia

Keywords: phylogeny, geographic distribution, recombination, host shift, resistance, genotype, subgenotype

Editorial on the Research Topic

Origin and Evolution of Hepatitis Viruses

Viral infection with hepatitis A, B, C, D, or E viruses (HAV, HBV, HCV, HDV, HEV) results in the syndrome of hepatitis, characterized by inflammation of the liver. Each virus is classified within a different virus family, yet all have hepatocyte-specific tropism and similar clinical manifestations. Full permissive infection with human hepatitis viruses is limited to higher primates, except for HEV, yet hepatitis-like viruses are known to infect invertebrates and all manner of vertebrates, including animals of the Laurasiatheria clade, such as bats and other insectivorous small mammals, suggesting an ancient origin and complex evolutionary history for hepatitis viruses. The definitive origin of these viruses remains largely unknown.

This Special Research Topic includes papers on HBV, HCV, HDV, and HEV, and includes studies investigating the consequences of virus evolution, such as geographic distribution and clinical outcomes. Other papers investigate intra-patient evolution, including super-infection and recombination, but also evolution over extensive timescales, thus providing a glimpse into hepatitis virus origins.

Several papers within the Topic presented important perspectives on the origins of HBV and HDV. The paper by Locarnini et al. elucidates the origin of primate HBV in the context of host evolution and migration. The authors posit that the evolutionary impetus giving rise to contemporary human and Old World non-human primate (NHP) HBV involved early human migration out of Africa during the upper Paleolithic era and the Neolithic agricultural expansion. The authors conclude that HBV evolution has occurred over many thousands of years with lineages disappearing over time and extant genotypes arising from specific population movements, such as slave trading. Most importantly, their investigation suggests that co-evolution among human and NHP HBV is not supported. The study by Netter et al. also investigated possible co-evolution of HDV and HDV-like agents with their hosts. In this important study, the authors advance the idea that HDV, a satellite virus most similar to plant viroids, likely originated within a cellular transcriptome as a circular RNA with ribozyme activity. Although delta-like agents have been detected in a multitude of different animals, including birds, fish, and insects, the helper virus, if indeed one is required, is not known. The paper suggests that host shifting, not co-divergence, is the suggested mode of HDV evolution based on HDV/host phylogeny.

Viral genomic recombination, as a mechanism of evolution involving viral superinfection, was the focus of several papers within the Research Topic. Jose-Abrego et al. observed that a high percentage (56%) of HIV-HBV co-infected patients were infected with multiple heterologous HBV genotypes (gt), including gtH, providing opportunity for genomic recombination. The paper from Giersch et al. presents an elegant study of HDV superinfection amongst gts1 and 3 using a human chimeric liver mouse model, observing that recombination is not a significant evolutionary process for HDV. Productive infection with multiple HDV strains, regardless of genotype, was not observed

Edited and reviewed by:
Rosa Maria Pintó,
University of Barcelona, Spain

***Correspondence:**
Carla Osiowy
carla.osiowy@phac-aspc.gc.ca

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *12 July 2021*
Accepted: *06 August 2021*
Published: *25 August 2021*

Citation:
Osiowy C and Yuen L (2021) Editorial: Origin and Evolution of Hepatitis Viruses. Front. Microbiol. 12:740255. doi: 10.3389/fmicb.2021.740255

within a single cell, thus reducing the likelihood of recombination. Strain specific L-HDAg and/or the innate immune response induced during primary infection may play a role in suppressing superinfection. RNA recombination is thought to occur via the host polymerase switching viral genome templates; however, it is rarely observed in negative sense RNA viruses (Patiño-Galindo et al., 2021).

The distinct global distribution of hepatitis viruses provides evidence of ongoing evolution, with phylogenetic analyses delineating the relationship among viral genotypes. In the paper by de Bernardi Schneider et al. a novel quasi-subgenotype of HBV gtD was observed to be common among HBV-infected Inuit patients living in West Greenland communities. By use of Bayesian inference, the quasi-subgenotype was dated to approximately 629 CE, thus describing yet another unique HBV subtype identified within Indigenous populations (Bouckaert et al., 2017; Yuen et al., 2019), further confirming the sustained HBV/human host relationship. Designating new HBV subgenotypes requires that a specific genetic divergence criteria ($\leq$4%) be met. The paper by Nicot et al. describes the basis for a similar criteria to differentiate HEV gt3 subgenotypes. By analyzing full genome (FG) HEV sequences, the authors generated a robust cut-off (0.093 nucleotide substitutions/site) to define a gt3 subgenotype. Following on the topic of HEV genotypes, the study by O'Keefe et al. describes the molecular epidemiology of HEV in Australia. The incidence of HEV within resource-rich regions has historically been associated with travel to endemic regions; however, gts3 and 4, associated with zoonotic transmission primarily from the animal family Suidae (hogs, boar, pig), now circulate within Australia, and thus are considered autochthonous. Following phylogenetic analysis, O'Keefe et al. determined that several gt3 non-travel associated strains may represent new "Australian" subgenotypes, and were likely transmitted from consumption of pork products.

Differences among hepatitis virus genotypes regarding pathogenesis, clinical, and treatment outcomes, have been described for HBV (Pujol et al., 2020), HCV (Sarrazin, 2021), and HDV, resulting in part from viral evolution and genomic mutation. The paper by Micas et al. describes an example of this with HEV, such that infection with gt4 was associated with higher ALT and AST activity than gt3 infection. This difference may be related to genotype-specific host innate and inflammatory responses. Both Araujo et al. and Rahimi et al. analyzed clinically significant viral hepatitis mutations. By parsing out clinically relevant immune escape, antiviral resistance and hepatocellular carcinoma-related mutations in >6,000 FG HBV sequences, Araujo et al. observed associations between specific mutation profiles and certain genotypes or subgenotypes. The study by Rahimi et al. investigated direct acting antiviral (DAA) resistance-associated substitutions (RAS) within the NS5A and NS5B genomic regions of HCV in treatment-naïve Iranian patients. They found different NS5A RAS prevalence among gts1a and 3a, although no NS5B RAS were observed, which is consistent with the reduced fitness of NS5B RAS in the absence of DAA.

Although the Special Research Topic did not include any studies of HAV, there have been important advancements in the past several years toward understanding the origin and evolution of HAV. Targeted exploration among non-primate mammals, such as bats, rodents, marsupials and harbor seals, has identified hepatovirus genus or HAV-like sequences, suggesting the capacity for HAV to infect diverse species (Anthony et al., 2015; Drexler et al., 2015; de Oliveira Carneiro et al., 2018; He et al., 2021). As observed with other hepatitis viruses, the HAV evolutionary pathway appears complex, involving host shifts (de Oliveira Carneiro et al., 2018) and a possible ancestral origin as an insect-borne virus (Drexler et al., 2015). This complex evolution may have resulted in a unique optimized host codon usage by the virus, resulting in increased capsid translational and protein-folding fitness in relation to other picornaviruses (D'Andrea et al., 2019).

The origin and evolution of hepatitis viruses continues to fascinate as evidenced by a continual increase in publications on the topic. As emerging and zoonotic viruses pose a serious threat to human health, a greater understanding of hepatitis-like viruses and human hepatitis virus evolution will help us prepare for future challenges. It is hoped that the Special Research Topic has contributed to this.

AUTHOR CONTRIBUTIONS

CO wrote the editorial. LY reviewed and provided feedback.

REFERENCES

Anthony, S. J., St. Leger, J. A., Liang, E., Hicks, A. L., Sanchez-Leon, M. D., Jain, K., et al. (2015). Discovery of a novel hepatovirus (phopivirus of seals) related to human hepatitis A virus. *mBio* 6:1180. doi: 10.1128/mBio.01180-15

Bouckaert, R., Simons, B. C., Krarup, H., Friesen, T. M., and Osiowy, C. (2017). Tracing hepatitis B virus (HBV) genotype B5 (formerly B6) evolutionary history in the circumpolar Arctic through phylogeographic modelling. *PeerJ* 5:e3757. doi: 10.7717/peerj.3757

D'Andrea, L., Pérez-Rodríguez, F. J., de Castellarnau, M., Guix, S., Ribes, E., Quer, J., et al. (2019). The critical role of codon composition on the translation efficiency robustness of the hepatitis A virus capsid. *Genome Biol. Evol.* 11, 2439–2456. doi: 10.1093/gbe/evz146

de Oliveira Carneiro, I., Sander, A. L., Silva, N., Moreira-Soto, A., Normann, A., Flehmig, B., et al. (2018). A novel marsupial hepatitis A virus corroborates complex evolutionary patterns shaping the genus hepatovirus. *J. Virol.* 92, e00082–18. doi: 10.1128/JVI.00082-18

Drexler, J. F., Corman, V. M., Lukashev, A. N., van den Brand, J. M., Gmyl, A. P., Bruenink, S., et al. (2015). Evolutionary origins of hepatitis A virus in small mammals. *Proc. Natl. Acad. Sci. U.S.A.* 112, 15190–15195. doi: 10.1073/pnas.1516992112

He, W., Gao, Y., Wen, Y., Ke, X., Ou, Z., Li, Y., et al. (2021). Detection of virus-related sequences associated with potential etiologies of hepatitis in liver tissue samples from rats, mice, shrews, and bats. *Front. Microbiol.* 12:653873. doi: 10.3389/fmicb.2021.653873

Patiño-Galindo, J. Á., Filip, I., and Rabadan, R. (2021). Global patterns of recombination across human viruses. *Mol. Biol. Evol.* 38, 2520–2531. doi: 10.1093/molbev/msab046

Pujol, F., Jaspe, R. C., Loureiro, C. L., and Chemin, I. (2020). Hepatitis B virus American genotypes: pathogenic variants? *Clin. Res. Hepatol. Gastroenterol.* 44, 825–835. doi: 10.1016/j.clinre.2020.04.018

Sarrazin, C. (2021). Treatment failure with DAA therapy: importance of resistance. *J. Hepatol.* 74, 1472–1482. doi: 10.1016/j.jhep.2021.03.004

Yuen, L. K., Littlejohn, M., Duchêne, S., Edwards, R., Bukulatjpi, S., Binks, P., et al. (2019). Tracing ancient human migrations into Sahul using hepatitis B virus genomes. *Mol. Biol. Evol.* 36, 942–954. doi: 10.1093/molbev/msz021

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 14 December 2020
doi: 10.3389/fmicb.2020.616023

Comprehensive Analysis of Clinically Significant Hepatitis B Virus Mutations in Relation to Genotype, Subgenotype and Geographic Region

Natalia M. Araujo[1], Sheila A. Teles[2] and Natália Spitz[1]*

[1] Laboratory of Molecular Virology, Oswaldo Cruz Institute, FIOCRUZ, Rio de Janeiro, Brazil, [2] Faculty of Nursing, Federal University of Goias, Goiânia, Brazil

Hepatitis B virus (HBV) is a highly variable DNA virus due to its unique life cycle, which involves an error-prone reverse transcriptase. The high substitution rate drives the evolution of HBV by generating genetic variants upon which selection operates. HBV mutants with clinical implications have been documented worldwide, indicating the potential for spreading and developing their own epidemiology. However, the prevalence of such mutants among the different HBV genotypes and subgenotypes has not been systematically analyzed. In the current study, we performed large-scale analysis of 6,479 full-length HBV genome sequences from genotypes A-H, with the aim of gaining comprehensive insights into the relationships of relevant mutations associated with immune escape, antiviral resistance and hepatocellular carcinoma (HCC) development with HBV (sub)genotypes and geographic regions. Immune escape mutations were detected in 10.7% of the sequences, the most common being I/T126S (1.8%), G145R (1.2%), M133T (1.2%), and Q129R (1.0%). HBV genotype B showed the highest rate of escape mutations (14.7%) while genotype H had no mutations ($P < 0.001$). HCC-associated mutations were detected in 33.7% of the sequences, with significantly higher frequency of C1653T, T1753V and A1762T/G1764A in genotype G than C ($P < 0.001$). The overall frequencies of lamivudine-, telbivudine-, adefovir-, and entecavir-resistant mutants were 7.3, 7.2, 0.5, and 0.2%, respectively, while only 0.05% showed reduced susceptibility to tenofovir. In particular, the highest frequency of lamivudine-resistant mutations was observed in genotype G and the lowest frequency in genotype E (32.5 and 0.3%; $P < 0.001$). The prevalence of HBV mutants was also biased by geographic location, with North America identified as one of the regions with the highest rates of immune escape, antiviral resistance, and HCC-associated mutants. The collective findings were discussed in light of natural selection and the known characteristics of HBV (sub)genotypes. Our data provide relevant information on the prevalence of clinically relevant HBV mutations, which may contribute to further improvement of diagnostic procedures, immunization programs, therapeutic protocols, and disease prognosis.

Keywords: hepatitis B virus, genotypes, mutation, immune escape, antiviral resistance, hepatocellular carcinoma, HBV evolution, natural selection

OPEN ACCESS

Edited by:
Carla Osiowy, National Microbiology Laboratory, Public Health Agency of Canada (PHAC), Canada

Reviewed by:
Masaya Sugiyama, National Center for Global Health and Medicine, Japan
Keiji Ueda, Osaka University, Japan

****Correspondence:***
Natalia M. Araujo
nmaraujo@ioc.fiocruz.br

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *10 October 2020*
Accepted: *25 November 2020*
Published: *14 December 2020*

Citation:
Araujo NM, Teles SA and Spitz N (2020) Comprehensive Analysis of Clinically Significant Hepatitis B Virus Mutations in Relation to Genotype, Subgenotype and Geographic Region. Front. Microbiol. 11:616023. doi: 10.3389/fmicb.2020.616023

INTRODUCTION

Hepatitis B is one of the most prevalent viral infections in humans and a major global public health problem. Hepatitis B virus (HBV) has infected one-third of the global population, with 257 million chronic infections worldwide and more than 880,000 deaths per year, mostly from cirrhosis and hepatocellular carcinoma (HCC) (WHO, 2020).

Hepatitis B virus is the prototype member of the family *Hepadnaviridae* containing a partially double-stranded relaxed circular DNA genome (∼3.2 kb) with a compact coding organization containing four partly or completely overlapping open reading frames (Seeger et al., 2013). Based on >7.5% genomic sequence divergence, HBV has been phylogenetically classified into nine genotypes (A–I) and one putative genotype (J). The significant diversity within specific HBV genotypes has led to further classification into numerous subgenotypes (Kramvis, 2014). HBV genotypes and subgenotypes have distinct geographic distributions. The unique replication cycle of HBV includes the activity of an error-prone reverse transcriptase (RT) that generates numerous viral variants (quasispecies). Constant HBV evolution leads to continued selection of mutations through pressure exerted by endogenous (host immune system) and exogenous (antiviral therapy and vaccination) factors, which fuels a molecular arms race between virus and host, resulting in emergence of mutations involved in several clinical effects (Araujo et al., 2011; Rajoriya et al., 2017; Revill et al., 2020).

The HBV surface antigen (HBsAg) contains a highly conserved antibody-neutralizing epitope cluster termed "a" determinant, which spans amino acids 124–147. Mutations occurring within or around this region lead to conformational changes, which can affect binding of neutralizing antibodies produced during natural infection or following active or passive immunization (Zuckerman and Zuckerman, 2003). These immune escape mutations account for several possible consequences, such as false-negative results by commercial HBsAg assays (occult hepatitis B) and evasion of anti-HBV immunoglobulin therapy and vaccine-induced immunity. G145R was the first escape mutation described and the most frequently detected HBV variant with proven vaccine escape properties in humans (Carman et al., 1990). Over the years, a number of other mutations associated with immune escape have been documented worldwide (Cooreman et al., 2001; Lazarevic et al., 2019).

Nucleos(t)ide analogs (NAs) are potent inhibitors of HBV RT activity. To date, six NAs, specifically, lamivudine (LAM), telbivudine (LdT), adefovir dipivoxil (ADV), entecavir (ETV), tenofovir disoproxil fumarate (TDF) and tenofovir alafenamide (TAF), have been approved for treatment of chronic HBV infections (Sarin et al., 2016; Easl, 2017; Terrault et al., 2018). However, drug-resistant mutations often arise during long-term use of therapies with a low barrier to resistance, such as LAM, leading to treatment failure and progression to liver disease (Zoulim, 2011). Additionally, owing to overlapping reading frames in the HBV genome, mutations in the polymerase gene resulting from antiviral selection pressure may affect neutralization epitopes within HBsAg (Pollicino et al., 2009).

The influence of HBV diversity on severity of liver disease has been investigated mainly for genotypes A-D (McMahon, 2009; Kao et al., 2010; Levrero and Zucman-Rossi, 2016). Genotype C is considered more oncogenic than genotypes A, B, and D (Chan et al., 2004; Wong et al., 2013). Likewise, mutations in the basal core promoter (BCP) resulting in decreased expression of HBeAg but enhanced viral replication and deletions in the Pre-S region are associated with increased risk of HCC (Liu et al., 2009).

The widespread distribution of HBV mutations with clinical implications poses a considerable challenge for design of diagnostic assays and current treatment strategies and is a potential threat to the long-term success of mass vaccination programs (Lazarevic, 2014; Yan et al., 2017). However, little is known on the frequencies of these mutations in different HBV genotypes. Knowledge of HBV subgenotypes is even more limited. Additionally, the unique distribution patterns of HBV (sub)genotypes in different geographic areas has hampered simultaneous comparisons. In the current study, we performed large-scale analysis of 6,479 full-length HBV genome sequences from genotypes A-H retrieved from public database with the aim of gaining comprehensive insights into the relationships of the most relevant escape-, resistance- and HCC-associated mutations with HBV (sub)genotypes and geographic regions. Knowledge of the prevalence of clinically relevant mutations among the different HBV genotypes and subgenotypes and their worldwide distribution should facilitate improvement of diagnostic procedures, immunization programs, therapeutic protocols, and disease prognosis.

MATERIALS AND METHODS

HBV Genome Sequences

All complete genome sequences (n = 6,479) of HBV genotypes A to H from the Hepatitis B Virus Database (HBVdb) available at https://hbvdb.lyon.inserm.fr/HBVdb/HBVdbIndex (Hayer et al., 2013) were obtained. Precomputed datasets for each genotype containing sequences aligned in the Clustal W format were downloaded on the following dates: genotype A (n = 871), 18th September 2019; B (n = 1,755), 13th November 2019; C (n = 2,187), 14th January 2020; D (n = 1,059), 31st January 2020; E (n = 292), 17th February 2020; F (n = 249), G (n = 40), and H (n = 26), 18th February 2020. Information on the country of sequence origin was extracted from existing GenBank annotations. Sequences were further grouped into 21 geographic subregions according to the UN Statistics Division[1].

HBV Subgenotyping

Subgenotyping of all sequences was assessed via Maximum Likelihood (ML) phylogenetic analysis. ML phylogenetic tree was inferred with the IQ-TREE v.1.6.12 program (Nguyen et al., 2015) under the GTR + I + G nucleotide substitution model selected with ModelFinder (Kalyaanamoorthy et al., 2017). A heuristic tree search was performed with the aid of the nearest neighbor interchange (NNI) algorithm and the reliability of

[1]https://unstats.un.org/unsd/methodology/m49/#geo-regions

phylogeny estimated with the approximate likelihood-ratio test (Anisimova and Gascuel, 2006) based on a Shimodaira–Hasegawa-like procedure (SH-aLRT). We employed the newer proposals for subgenotype classification reviewed elsewhere (Kramvis, 2014). New classifications of subgenotypes proposed subsequently were not included in the analyses. GenBank accession numbers of the reference sequences are as follows: AB241115, AY233278 (subgenotype A1); AB116079, AY738141 (A2); AB194951, AB194952, AM180623, AY934764, FJ692609, FJ692613 (QS-A3); GQ331047, GQ331048 (A4); AB014366, D00329 (B1); AP011084, AY596111 (B2); AB219427, AP011085, AP011086, AP011091, AP011093, AP011094, EF473977, M54923 (QS-B3); AB073835, AB115551 (B4); AB287316, DQ463787 (B5); AB031265, AB112066 (C1); AB033553, AB368297, AF533983, HQ622095 (QS-C2); KU695745, KU695746, X75656, X75665 (C3); AB048704, AB048705, KF873543, KF873545 (C4); AB241109, AP011099, EU410081, KM999992 (C5); KM999993, AP011103, AP011102 (C6); EU670263 (C7); AP011104, AP011105, AP011106, AP011107 (C8); AP011108 (C9); AB540583 (C10); AB554019, AB554020 (C11); AB554018, AB554025 (C12); AB644280, AB644281 (C13); GQ377555, HM011493 (C14); AB644286 (C15); AB644287 (C16); FJ904424, X80926 (D1); AB109475, Z35716 (D2); AB493845, AB493846, AY233291, X65257 (D3); AB033559, AB048702 (D4); AB033558, DQ315779 (D5); AM494716, FJ904430 (D6); AB091256, HM363610, X75664 (E); AY090459, DQ823095, HM585194 (F1); AY090455, DQ899143, X69798 (F2); AB036905, AB036915, MH051986 (F3); AB166850, DQ823090, KJ843175 (F4); AB056513, AB064310, EF634480 (G); AB516395, AY090454, AY090457 (H); FJ023664, FJ023660 (I); AB486012 (J).

Mutational Analysis

Sequence alignments were analyzed using MEGA version 10.1.8 (Kumar et al., 2018). An Excel database containing information on each nucleotide/amino acid in the respective positions of each sequence was constructed (**Supplementary Table S1**). A total of 21 clinically relevant positions according to previous reports were examined, including 10 amino acid positions within HBsAg (120, 126, 129, 130, 133, 141, 142, 143, 144, and 145) related to immune escape (Cooreman et al., 2001; Ma and Wang, 2012; Qin and Liao, 2018; Lazarevic et al., 2019), four nucleotide positions within the core promoter gene (1653 in enhancer II and 1753, 1762, and 1764 in BCP) and Pre-S deletions related to HCC development (Chou et al., 2008; Liu et al., 2009; Kao et al., 2010; Tseng et al., 2015), and seven amino acid positions within the RT domain of HBV polymerase (180, 181, 184, 202, 204, 236, and 250) related to antiviral resistance (Zoulim and Locarnini, 2009). According to the most recent clinical practice guidelines of the European Association for the Study of the Liver (Easl, 2017), the following amino acid substitution profiles for HBV-resistant mutants were used: rtM204V/I (LAM resistance), rtM204I or L180M + rtM204V (LdT resistance), rtA181T/V or rtN236T (ADV resistance), rtL180M + rtM204I/V ± rtT184S/G ± rtS202I/G ± rtM250V (ETV resistance), and rtA181T/V + rtN236T (TDF/TAF-reduced susceptibility).

Statistical Analysis

Statistical analysis was performed using SPSS statistical software package version 23.0 (IBM SPSS Inc., Chicago, IL, United States). Frequencies were compared using the chi-squared test or Fisher's exact test. P-values < 0.05 were considered statistically significant.

RESULTS

HBV Subgenotypes and Geographic Distribution

For comprehensive evaluation, all HBV genome sequences from genotypes A, B, C, D, and F (n = 6,121) were classified into subgenotypes via phylogenetic analysis of each genotype dataset (**Supplementary Figure S1**). The phylogenetic tree of the complete data used in the study is shown in **Figure 1**. Subgenotyping results were further compared with information on subgenotypes of the sequences in GenBank, where available. Sequences with conflicting results (n = 29) or undetermined subgenotypes (n = 45) were not included in subsequent analyses. The following subgenotypes were identified for each genotype: A1 (n = 265, 30.4%), A2 (n = 553, 63.5%), QS-A3 (n = 48, 5.5%), A4 (n = 3, 0.3%), and not determined (n = 2, 0.2%) (genotype A, n = 871); B1 (n = 57, 3.2%), B2 (n = 1310, 74.6%), QS-B3 (n = 207, 11.8%), B4 (n = 128, 7.3%), B5 (n = 52, 3.0%), and not determined (n = 1, 0.1%) (genotype B, n = 1755); C1 (n = 451, 20.6%), QS-C2 (n = 1518, 69.4%), C3 (n = 28, 1.3%), C4 (n = 13, 0.6%), C5 (n = 19, 0.9%), C6–C15 (n = 103, 4.7%), not determined (n = 36, 1.6%) and conflicting results (n = 19, 0.9%) (genotype C, n = 2187); D1 (n = 516, 48.7%), D2 (n = 291, 27.5%), D3 (n = 189, 17.8%), D4 (n = 21, 2.0%), D5 (n = 21, 2.0%), D6 (n = 7, 0.7%), not determined (n = 6, 0.6%), and conflicting results (n = 8, 0.8%) (genotype D, n = 1059); F1 (n = 123, 49.4%), F2 (n = 29, 11.6%), F3 (n = 39, 15.7%), F4 (n = 56, 22.5%), and conflicting results (n = 2, 0.8%) (genotype F, n = 249). In addition, sequences with information on the country of origin available in GenBank were grouped into geographic subregions. For simplification purposes, sequences classified in subgenotypes C6–C15 were grouped as C6–C15. The global geographical distribution of HBV (sub)genotypes from 6,137 full-length genome sequences present in the HBVdb database is shown in **Figure 2**. The sequence list and corresponding background information is available as **Supplementary Table S1**.

Immune escape Mutations

Ten positions within HBsAg (P120, I/T126, Q129, G130, M133, K141, P142, S/T143, D144, and G145) incorporating 25 immune escape mutations were analyzed. The amino acid frequencies at each position in different HBV genotypes are shown in **Table 1**. Overall, immune escape mutations were detected in 691 of 6,479 sequences (10.7%). HBV genotype B showed the highest frequency of mutation (14.7%), followed by genotypes C (11.2%), G (10%), D (9.3%), A (7.5%), E (5.1%), and F (2.0%) ($P < 0.001$). No escape mutations were detected in genotype H sequences. The most common substitutions were I/T126S (1.8%),

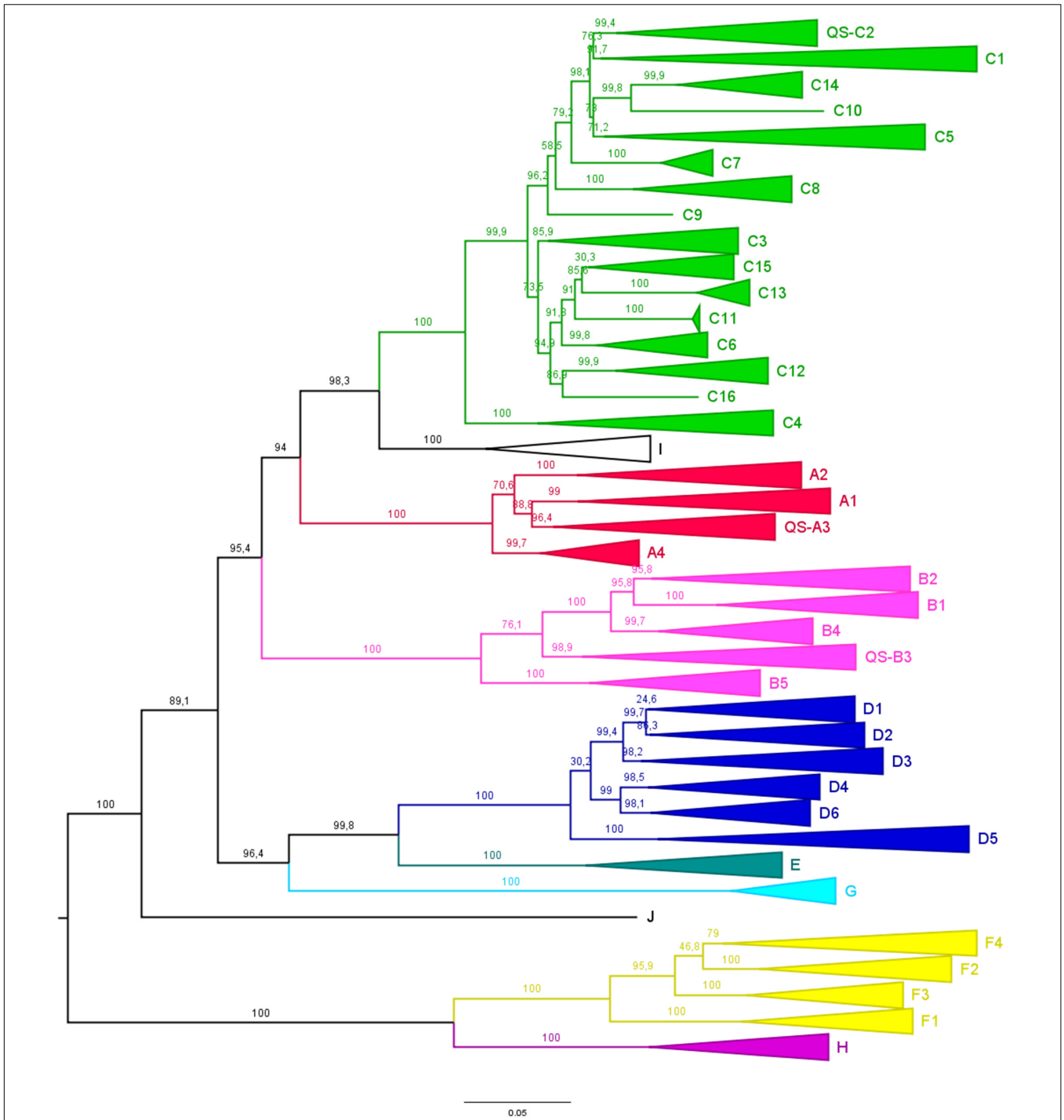

FIGURE 1 | Phylogenetic analysis of HBV complete genome sequences representative of genotypes A–J. Maximum likelihood phylogenetic tree of 6,538 complete genome sequences (6,434 from HBVdb database plus 104 reference sequences). Sequences whose subgenotype could not be determined (n = 45) were excluded from the analysis. The branches are colored according to the genotypes. (Sub)genotypes are indicated next to their respective clusters. The numbers in branches indicate the statistical support (aLRT value).

G145R (1.2%), M133T (1.2%), Q129R (1.0%), I/T126A (0.8%), and P120T (0.8%). These major mutations were more frequent in genotypes A (P120T, 3.2%), B (I/T126A, 2.8%; Q129R, 2.2%), C (I/T126S, 3.9%; M133T, 1.9%), and G (G145R, 5.0%). Notably, G145R was only detected in genotypes G (5.0%), D (2.5%), C (2.1%), and B (0.2%) ($P < 0.001$) (**Table 1**).

The HBV subgenotypes A2, B1, C6–C15, D4 and F2 displayed the highest frequencies of immune escape mutants within their

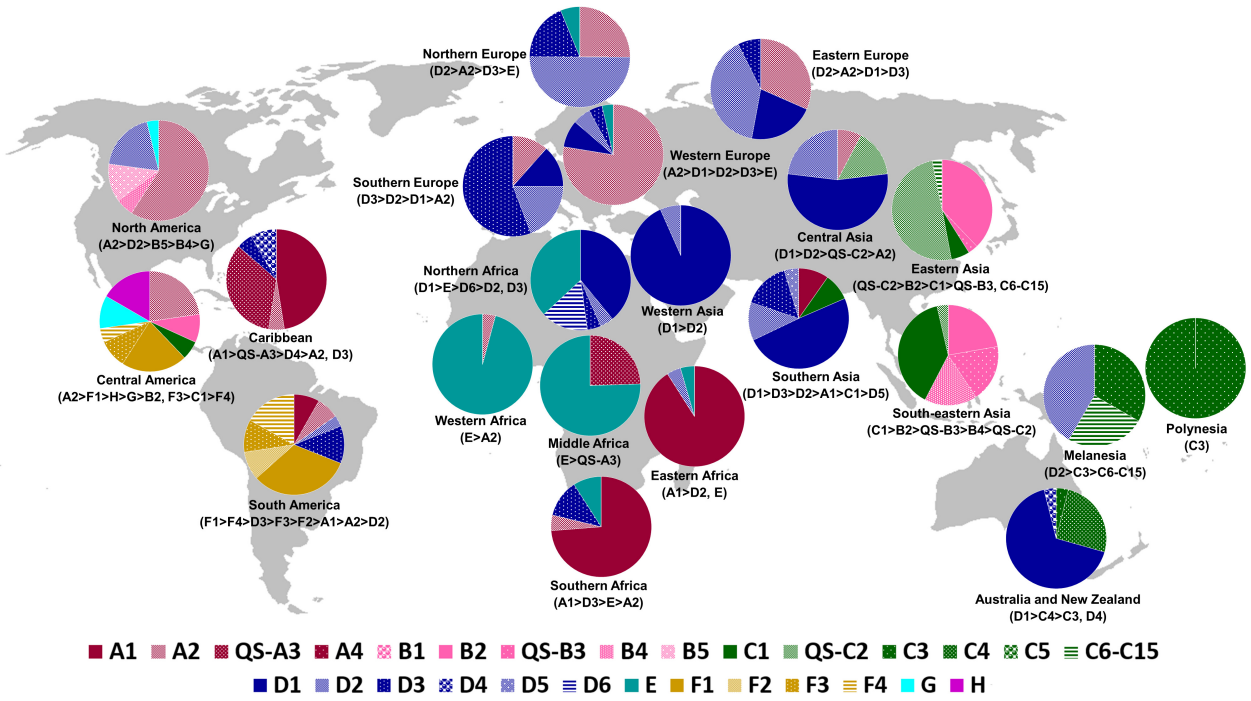

FIGURE 2 | Geographic distribution of HBV genotypes and subgenotypes based on 6,137 complete genome sequences present in the HBVdb database. (Sub)genotypes with frequencies < 3% were not included in the figure. The map was reconstructed using Wikimedia Commons (https://commons.wikimedia.org/wiki/File:BlankMap-World-noborders.png), this figure is similar but not identical to the original image and is therefore for illustrative purposes only. QS, quasi-subgenotype.

respective genotypes (A2: 9.6%; B1: 15.8%; C6-C15: 36.9%; D4: 19%; F2: 6.9%), with significant differences within genotypes A and C ($P = 0.018$ and $P < 0.001$, respectively). On the other hand, subgenotypes A4, C4, C5, and D6 had no mutations in the listed HBsAg positions. In particular, significant differences in the frequency of G145R were observed within genotypes C and D, with the highest rates recorded for subgenotypes C6–C15 and D2 (34 and 8.2%, $P < 0.001$ for both; **Tables 2–6**).

HCC-Associated Mutations

The frequencies of C1653T, T1753V and A1762T/G1764A (double mutant) and Pre-S deletions in different HBV genotypes are shown in **Table 1**. HCC-associated mutations were detected in 2,183 out of 6,479 sequences (33.7%). HBV genotype G showed the highest frequency of mutations (97.5%), followed by genotype C (49.7%), D (30.9%), F (30.5%), A (29.9%), E (24.3%), B (18.1%), and H (15.4%) ($P < 0.001$). C1653T, T1753V and A1762T/G1764A were highly frequent in genotype G (95, 95, and 97.5%, respectively), followed by genotype C (12, 18.1, and 46.1%, respectively), while Pre-S deletions prevailed in genotype C (3.3%) (**Table 1**).

All HBV subgenotypes displayed HCC-associated gene variations. QS-A3, B5, C6–C15, D1 and F2 displayed the highest mutation frequencies within their respective genotypes (QS-A3: 35.4%; B5: 30.8%; C6–C15: 57.3%; D1: 36.4%; F2: 51.7%). Significant differences in A1762T/G1764A rates were found within genotypes B, C, D, and F, with subgenotypes B2, C6-C15, D1 and F1 showing the highest rates (17.7, 56.3, 28.3, and 29.3%, respectively). Similarly, Pre-S deletion frequencies were markedly different within genotypes A and B, with the highest rates detected for subgenotypes A1 and QS-B3 (3.0 and 2.4%, respectively, $P < 0.05$ for both) (**Tables 2–6**).

Antiviral Resistance Mutations

Seven amino acid positions within the RT domain of HBV polymerase involving 11 mutations associated with resistance to NAs were analyzed. The frequencies of mutations and resistant mutants for each NA in different HBV genotypes are shown in **Table 1**. Overall, rtL180M, rtA181T/V, rtT184G/S, rtS202G/I, rtM204I, rtM204V, rtN236T, and rtM250V were more frequently detected in genotypes A (13.9%), C (0.8%), D (0.1%), H (3.8%), G (25%), A (13.1%), B (0.7%), and C (0.1%), respectively. Significant differences were evident among HBV genotypes in terms of resistance to NAs. LAM- and LdT-resistant mutants were more frequent in genotype G (32.5% for both), followed by genotypes A (14.1 and 13.8%), C (8.6 and 8.4%), H (7.7% for both), D (7.4% for both), B (3.5% for both), F (3.2% for both) and E (0.3% for both) ($P < 0.001$). ADV-resistant mutants were only found in genotypes C (0.9%), B (0.7%), A (0.1%), and D (0.1%) ($P = 0.016$), whereas ETV-resistant mutants were specifically detected in genotypes H (3.8%), C (0.4%), and D (0.2%)

($P < 0.001$). Only genotypes B and C contained mutations with reduced susceptibility to TDF and TAF (0.1% for both genotypes) (**Table 1**).

The HBV subgenotypes A2, B4, QS-C2, D2 and F2 displayed the highest frequencies of LAM- and LdT-resistant mutants within their respective genotypes (A2: 21.5 and 21.3%; B4: 18.8% for both; QS-C2: 11.3% and 11.1%; D2: 16.2% for both; F2: 6.9% for both). Subgenotypes A2, B2, QS-C2 and D1 showed the highest frequencies of ADV-resistant mutants within their respective genotypes (A2: 0.2%; B2: 1.0%; QS-C2: 1.3%; D1: 0.2%). Subgenotypes C6–C15 and D2 showed the highest frequencies of ETV-resistant mutants within their respective genotypes (C6–C12: 1.0%; D2: 0.3%). Only B2 and QS-C2 had mutants with reduced susceptibility to TDF and TAF (0.1% for both) (**Tables 2–6**).

Geographic Distribution of HBV Mutants

The frequencies of HBV mutants according to geographic region are shown in **Figure 3** and **Table 7**. Central Asia was the only region where no mutants were detected. The highest rates of HCC-associated mutants were observed in Southern Asia (41%), followed by North America (37%) and Eastern Asia (36.5%) and the lowest rates in Middle Africa (14.6%), Western Europe (17.5%), and Australia and New Zealand (19.6%) ($P < 0.001$). Immune escape mutants were detected more often in North America (16.3%), Eastern Asia (12.3%) and South-eastern Asia (12%) and less frequently in Middle Africa (0), Eastern Africa (1.4%) and the Caribbean (1.6%) ($P < 0.001$). Regarding antiviral therapy resistance, HBV sequences from Australia and New Zealand, Caribbean, Melanesia, Northern Africa, Northern Europe, Polynesia, Southern Africa, and Western Africa contained no NA-resistant mutations. Higher frequencies of LAM- and LdT-resistant mutants were found in North America (40.9% for both), Southern Europe (24% for both) and Western Europe (9.1 and 8.4%), and lower frequencies in Eastern Europe (0.7% for both), Middle Africa (1.0% for both) and Eastern Africa (1.4% for both) ($P < 0.001$). ADV-resistant mutants were specifically identified in Eastern Asia (1.1%), North America (0.2%) and Southern Asia (0.2%) and ETV-resistant mutants in Southern Asia (0.4%) and Eastern Asia (0.1%) ($P > 0.05$ for all). Mutants with reduced TDF and TAF susceptibility were solely detected in Eastern Asia (0.1%) (**Figure 3** and **Table 7**).

The comparison of mutation profiles of same (sub)genotype among geographic regions is shown in **Supplementary Table S2**. A1 isolates from Southern Asia displayed significantly higher rates of HCC-associated mutants (61.4%) than A1 isolates from Caribbean (32.1%), Eastern Africa (25%), South America (24%), and Southern Africa (21.9%) ($P < 0.05$). Similar results were observed between C isolates from Asia (Southern Asia, 57.1%; Eastern Asia, 50.5%; South-eastern Asia, 43.4%) and Oceania (Australia and New Zealand, 20%; Melanesia, 25%; Polynesia, 27.3%) ($P < 0.05$), E isolates from Western Africa and Middle Africa (33.1% and 8.5%, respectively, $P < 0.05$), and F isolates from North America and South America (87.5% and 30%, respectively, $P < 0.05$). A2 isolates from North America and Western Europe showed significant differences in the rates of HCC-associated mutants (40.3 and 9.3%, respectively, $P < 0.05$) and immune escape mutants (16.5 and 3.7%, respectively, $P < 0.05$). Moreover, isolates from North America showed the highest frequencies of LAM-resistant mutants within (sub)genotypes A2, B, D, and G ($P < 0.05$) (**Supplementary Table S2**).

DISCUSSION

Evolution is an important aspect of epidemiology of viral diseases. HBV evolution occurs through both mutation and recombination processes (Kay and Zoulim, 2007; Araujo, 2015). Since HBV replication involves an error-prone reverse transcription step, nucleotide substitution rates are higher than those for other DNA viruses. The estimated HBV mutation rates are reported to range from 10^{-4} to 10^{-6} substitutions/site/year (Zehender et al., 2014; Paraskevis et al., 2015; Spitz et al., 2019). This high substitution rate allows HBV to evolve rapidly and adapt to changes in the host environment, leading to the emergence of mutations with important implications for prevention, diagnosis, treatment, and prognosis of infection. HBV mutants have been documented worldwide (Zhang et al., 2016), indicating the potential to spread in a range of human populations and develop their own epidemiology.

HBV genotypes and subgenotypes have been associated with differences in disease progression, response to antiviral therapy, and clinical outcomes (Kramvis and Kew, 2005; McMahon, 2009; Kim et al., 2011; Liu and Kao, 2013). One potential reason is that HBV mutations may be more common among some (sub)genotypes than among others. However, due to the unique distribution patterns of HBV genotypes in Asian and Western countries, epidemiological data on HBV mutants are based on comparisons between genotypes B and C or A and D, with little information on the other genotypes. In this study, mutations related to immune escape, antiviral resistance, and HCC development were analyzed in 6,479 HBV genome sequences classified into genotypes A–H. To our knowledge, the present study has conducted the most comprehensive analysis of the association of clinically relevant mutations of HBV with genotypes, subgenotypes, and geographic regions.

Overall, immune escape mutations were detected in 10.7% HBV genomes. Similar rates have been determined in cohorts from Argentina (7.5–10.7%) (Di Lello et al., 2019), China (9.01%) (Yan et al., 2017), Spain (6.6–12.5%) (Avellon and Echevarria, 2006), and Turkey (8.3%) (Sayan et al., 2010). Only 4 of the 25 escape mutations had total frequencies of ≥1% (I/T126S, 1.8%; G145R, 1.2%; M133T, 1.2%; and Q129R, 1.0%), consistent with a previous large-scale analysis showing frequencies of no less than 1% for most escape mutations (Ma and Wang, 2012). Particularly, the most common escape mutation identified, I/T126S, was previously associated with vaccine failure in Chinese adults undergoing the HBV vaccination program (He et al., 2001) and

TABLE 1 | Clinically relevant HBV mutations by genotype.

Sites	Genotypes, n (%)									Clinical issue	*P*
	Total (*n* = 6479)	A (*n* = 871)	B (*n* = 1755)	C (*n* = 2187)	D (*n* = 1059)	E (*n* = 292)	F (*n* = 249)	G (*n* = 40)	H (*n* = 26)		
P120										Immune escape	
L	2 (0.0)	0	0	0	1 (0.1)	1 (0.3)	0	0	0		0.096
Q	3 (0.0)	0	1 (0.1)	0	1 (0.1)	0	1 (0.4)	0	0		0.256
S	43 (0.7)	1 (0.1)	25 (1.4)	2 (0.1)	14 (1.3)	1 (0.3)	0	0	0		<0.001
T	50 (0.8)	28 (3.2)	5 (0.3)	8 (0.4)	9 (0.8)	0	0	0	0		<0.001
Total	98 (1.5)	29 (3.3)	31 (1.8)	10 (0.5)	25 (2.8)	2 (0.7)	1 (0.4)	0	0		<0.000
I/T126										Immune escape	
A	53 (0.8)	1 (0.1)	50 (2.8)	0	0	0	2 (0.8)	0	0		<0.001
N	24 (0.4)	0	1 (0.1)	21 (1.0)	2 (0.2)	0	0	0	0		<0.001
S	116 (1.8)	14 (1.6)	14 (0.8)	86 (3.9)	2 (0.2)	0	0	0	0		<0.001
Total	193 (3.0)	15 (1.7)	65 (3.7)	107 (4.9)	4 (0.4)	0	2 (0.8)	0	0		<0.001
Q129										Immune escape	
H	46 (0.7)	0	39 (2.2)	2 (0.1)	4 (0.4)	1 (0.3)	0	0	0		<0.001
N	4 (0.1)	0	0	4 (0.2)	0	0	0	0	0		0.346
R	63 (1.0)	4 (0.5)	39 (2.2)	6 (0.3)	14 (1.3)	0	0	0	0		<0.001
Total	113 (1.7)	4 (0.5)	78 (4.4)	12 (0.5)	18 (1.7)	1 (0.3)	0	0	0		<0.001
G130										Immune escape	
N	36 (0.6)	4 (0.5)	4 (0.2)	25 (1.1)	3 (0.3)	0	0	0	0		0.003
R	11 (0.2)	0	4 (0.2)	5 (0.2)	0	1 (0.3)	1 (0.4)	0	0		0.600
Total	47 (0.7)	4 (0.5)	8 (0.5)	30 (1.4)	3 (0.3)	1 (0.3)	1 (0.4)	0	0		0.006
M133										Immune escape	
I	16 (0.2)	3 (0.3)	4 (0.2)	3 (0.1)	3 (0.3)	3 (1.0)	0	0	0		0.219
L	47 (0.7)	0	43 (2.5)	2 (0.1)	1 (0.1)	1 (0.3)	0	0	0		<0.001
T	79 (1.2)	8 (0.9)	26 (1.5)	42 (1.9)	1 (0.1)	2 (0.7)	0	0	0		<0.001
Total	142 (2.2)	11 (1.3)	73 (4.2)	47 (2.1)	5 (0.5)	6 (2.1)	0	0	0		<0.001
K141										Immune escape	
E	3 (0.0)	1 (0.1)	1 (0.1)	1 (0.0)	0	0	0	0	0		0.975
I	2 (0.0)	0	0	2 (0.1)	0	0	0	0	0		0.788
Total	5 (0.1)	1 (0.1)	1 (0.1)	3 (0.1)	0	0	0	0	0		0.922
P142										Immune escape	
L	4 (0.1)	0	0	1 (0.0)	2 (0.2)	0	0	1 (2.5)	0		<0.001
S	2 (0.0)	1 (0.1)	1 (0.1)	0	0	0	0	0	0		0.828
Total	6 (0.1)	1 (0.1)	1 (0.1)	1 (0.0)	2 (0.2)	0	0	1 (2.5)	0		<0.001
S/T143L	22 (0.3)	0	0	1 (0.0)	20 (1.9)	1 (0.3)	0	0	0	Immune escape	<0.001
D144										Immune escape	
A	4 (0.1)	2 (0.2)	0	1 (0.0)	0	1 (0.3)	0	0	0		0.204
E	13 (0.2)	2 (0.2)	1 (0.1)	2 (0.1)	4 (0.4)	2 (0.7)	1 (0.4)	1 (2.5)	0		0.007
G	3 (0.0)	1 (0.1)	0	1 (0.0)	1 (0.1)	0	0	0	0		0.927
Total	20 (0.3)	5 (0.6)	1 (0.1)	4 (0.2)	5 (0.5)	3 (1.0)	1 (0.4)	1 (2.5)	0		0.008
G145										Immune escape	
A	34 (0.5)	1 (0.1)	4 (0.2)	26 (1.2)	2 (0.2)	1 (0.3)	0	0	0		<0.001
R	78 (1.2)	0	4 (0.2)	46 (2.1)	26 (2.5)	0	0	2 (5.0)	0		<0.001
Total	112 (1.7)	1 (0.1)	8 (0.5)	72 (3.3)	28 (2.6)	1 (0.3)	0	2 (5.0)	0		<0.001
Escape mutants[a]	691 (10.7)	65 (7.5)	258 (14.7)	245 (11.2)	99 (9.3)	15 (5.1)	5 (2.0)	4 (10)	0	Immune escape	<0.001
C1653T	474 (7.3)	28 (3.2)	28 (1.6)	262 (12)	93 (8.8)	16 (5.5)	7 (2.8)	38 (95)	2 (7.7)	HCC	<0.001
T1753V	820 (12.7)	60 (6.9)	78 (4.4)	395 (18.1)	176 (16.6)	40 (13.7)	32 (12.9)	38 (95)	1 (3.8)	HCC	<0.001
A1762T/G1764A	1871 (28.9)	234 (26.9)	272 (15.5)	1009 (46.1)	228 (21.5)	32 (11)	56 (22.5)	39 (97.5)	1 (3.8)	HCC	<0.001
Pre-S deletions	98 (1.5)	10 (1.1)	10 (0.6)	73 (3.3)	0	2 (0.7)	3 (1.2)	0	0	HCC	<0.001
HCC-associated mutants[b]	2183 (33.7)	260 (29.9)	318 (18.1)	1088 (49.7)	327 (30.9)	71 (24.3)	76 (30.5)	39 (97.5)	4 (15.4)	HCC	<0.001
rtL180M	309 (4.8)	121 (13.9)	24 (1.4)	90 (4.1)	60 (5.7)	1 (0.3)	8 (3.2)	3 (7.5)	2 (7.7)	Antiviral resistance	<0.001
rtA181										Antiviral resistance	
T	3 (0.0)	0	0	3 (0.1)	0	0	0	0	0		0.553

(Continued)

TABLE 1 | Continued

Sites	Genotypes, n (%)									Clinical issue	*P*
	Total (*n* = 6479)	A (*n* = 871)	B (*n* = 1755)	C (*n* = 2187)	D (*n* = 1059)	E (*n* = 292)	F (*n* = 249)	G (*n* = 40)	H (*n* = 26)		
V	18 (0.3)	0	2 (0.1)	15 (0.7)	1 (0.1)	0	0	0	0		0.005
Total	21 (0.3)	0	2 (0.1)	18 (0.8)	1 (0.1)	0	0	0	0		0.001
rtT184										Antiviral resistance	
G	0	0	0	0	0	0	0	0	0		–
S	2 (0.0)	0	0	1 (0.0)	1 (0.1)	0	0	0	0		0.924
Total	2 (0.0)	0	0	1 (0.0)	1 (0.1)	0	0	0	0		0.924
rtS202										Antiviral resistance	
G	11 (0.2)	0	3 (0.2)	6 (0.3)	1 (0.1)	0	0	0	1 (3.8)		0.001
I	1 (0.0)	0	0	0	1 (0.1)	0	0	0	0		0.645
Total	12 (0.2)	0	3 (0.2)	6 (0.3)	2 (0.2)	0	0	0	1 (3.8)		0.002
rtM204										Antiviral resistance	
I	237 (3.7)	9 (1.0)	41 (2.3)	125 (5.7)	51 (4.8)	0	1 (0.4)	10 (25)	0		<0.001
V	236 (3.6)	114 (13.1)	20 (1.1)	62 (2.8)	27 (2.5)	1 (0.3)	7 (2.8)	3 (7.5)	2 (7.7)		<0.001
Total	472 (7.3)	123 (14.1)	61 (3.5)	187 (8.6)	78 (7.4)	1 (0.3)	8 (3.2)	13 (32.5)	2 (7.7)		<0.001
rtN236T	17 (0.3)	1 (0.1)	12 (0.7)	4 (0.2)	0	0	0	0	0	Antiviral resistance	0.014
rtM250V	3 (0.0)	0	0	3 (0.1)	0	0	0	0	0	Antiviral resistance	0.053
LAM-resistant mutants[c]	473 (7.3)	123 (14.1)	61 (3.5)	187 (8.6)	78 (7.4)	1 (0.3)	8 (3.2)	13 (32.5)	2 (7.7)	Antiviral resistance	<0.001
LdT-resistant mutants[d]	466 (7.2)	120 (13.8)	61 (3.5)	183 (8.4)	78 (7.4)	1 (0.3)	8 (3.2)	13 (32.5)	2 (7.7)	Antiviral resistance	<0.001
ADV-resistant mutants[e]	35 (0.5)	1 (0.1)	13 (0.7)	20 (0.9)	1(0.1)	0	0	0	0	Antiviral resistance	0.016
ETV-resistant mutants[f]	11 (0.2)	0	0	8 (0.4)	2 (0.2)	0	0	0	1 (3.8)	Antiviral resistance	<0.001
TDF/TAF-reduced susceptibility mutants[g]	3 (0.0)	0	1 (0.1)	2 (0.1)	0	0	0	0		Antiviral resistance	0.949

The P-value was listed as " – " when the specific mutation was found in no genotype (statistical analysis not conducted).
LAM, lamivudine; LdT, telbivudine; ADV, adefovir; ETV, entecavir; TDF, tenofovir disoproxil fumarate; TAF, tenofovir alafenamide; rt, reverse transcriptase.
[a]*Sequences with at least one immune escape mutation.*
[b]*Sequences with at least one HCC-associated mutation.*
[c]*Amino acid substitution profile for LAM resistance: rtM204V/I.*
[d]*Amino acid substitution profile for LdT resistance: rtM204I or L180M + M204V.*
[e]*Amino acid substitution profile for ADV resistance: rtA181T/V or rtN236T.*
[f]*Amino acid substitution profile for ETV resistance: rtL180M + rtM204I/V ± rtT184S/G ± rtS202I/G ± rtM250V.*
[g]*Amino acid substitution profile for TDF/TAF-reduced susceptibility: rtA181T/V + rtN236T.*

most prevalent among patients under 15 years of age after mass vaccination in China (Yan et al., 2017).

Notably, significant differences in the frequencies of escape mutations among HBV genotypes were observed in this study, with genotypes A–D and G showing higher rates than E, F, and H. The lack of escape mutations observed in genotype H should be interpreted with caution due to the low number of genotype H sequences available for analysis. However, similar findings were reported by Ma and Wang (2012), who showed that HBV genotypes A-D harbor more escape mutations than the other genotypes. Likewise, Di Lello et al. (2019) observed a significantly higher rate of escape mutations in genotype D (33%) than genotype F (2.3%). The results collectively suggest low circulation of immune escape mutants among the Amerindian (and more divergent) HBV genotypes F and H, and highlight the necessity to monitor genotypes A (mainly A2), B, C (mainly C6-C15), D, and G.

Moreover, since the HBV vaccine used worldwide is produced with HBsAg subgenotype A2, stronger selection pressure of the humoral immune response against the "a" determinant of HBV-A2 isolates may occur. Consequently, higher rates of immune escape mutants are expected in subgenotype A2 than other (sub)genotypes. In fact, significant differences in the frequencies of immune escape mutants were observed within genotype A, with A2 showing the highest rates. However, other (sub)genotypes included higher rates of escape mutants than A2, suggesting that this potential selective pressure on subgenotype A2 is not yet a concern for virus vaccine design.

Hepatocellular carcinoma development due to HBV infection is a multifactorial process associated with viral, host and environmental factors. Patients with chronic HBV infection have a 20-fold higher risk of HCC than non-infected individuals (Parkin, 2006). The ability to identify patients at risk of HCC is crucial, since progression to liver cancer in the

TABLE 2 | Clinically relevant mutations in HBV/A subgenotypes.

	Subgenotypes, n (%)					
Sites	**A1 (*n* = 265)**	**A2 (*n* = 553)**	**QS-A3 (*n* = 48)**	**A4 (*n* = 3)**	**Clinical issue**	***P***
P120					Immune escape	
L	0	0	0	0		–
Q	0	0	0	0		–
S	0	1 (0.2)	0	0		0.903
T	2 (0.8)	26 (4.7)	0	0		0.013
Total	2 (0.8)	27 (4.9)	0	0		0.010
I/T126					Immune escape	
A	1 (0.4)	0	0	0		0.516
N	0	0	0	0		–
S	1 (0.4)	13 (2.4)	0	0		0.152
Total	2 (0.8)	13 (2.4)	0	0		0.303
Q129					Immune escape	
H	0	0	0	0		–
N	0	0	0	0		–
R	2 (0.8)	2 (0.4)	0	0		0.836
Total	2 (0.8)	2 (0.4)	0	0		0.836
G130					Immune escape	
N	3 (1.1)	1 (0.2)	0	0		0.285
R	0	0	0	0		–
Total	3 (1.1)	1 (0.2)	0	0		0.285
M133					Immune escape	
I	2 (0.8)	0	1 (2.1)	0		0.059
L	0	0	0	0		–
T	1 (0.4)	7 (1.3)	0	0		0.561
Total	3 (1.1)	7 (1.3)	1 (2.1)	0		0.954
K141					Immune escape	
E	0	1 (0.2)	0	0		0.903
I	0	0	0	0		–
Total	0	1 (0.2)	0	0		0.903
P142					Immune escape	
L	0	0	0	0		–
S	0	1 (0.2)	0	0		0.903
Total	0	1 (0.2)	0	0		0.903
S/T143L	0	0	0	0	Immune escape	–
D144					Immune escape	
A	1 (0.4)	1 (0.2)	0	0		0.935
E	0	2 (0.4)	0	0		0.766
G	1 (0.4)	0	0	0		0.516
Total	2 (0.8)	3 (0.5)	0	0		0.926
G145					Immune escape	
A	0	1 (0.2)	0	0		0.903
R	0	0	0	0		–
Total	0	1 (0.2)	0	0		0.903
Escape mutants[a]	11 (4.2)	53 (9.6)	1 (2.1)	0	Immune escape	0.018
C1653T	14 (5.3)	8 (1.4)	6 (12.5)	0	HCC	<0.001
T1753V	24 (9.1)	28 (5.1)	7 (14.6)	1 (33.3)	HCC	<0.006
A1762T/G1764A	70 (26.4)	148 (26.8)	15 (31.3)	1 (33.3)	HCC	0.905
Pre-S deletions	8 (3.0)	2 (0.4)	0	0	HCC	0.008
HCC-associated mutants[b]	85 (32.1)	157 (28.4)	17 (35.4)	1 (33.3)	HCC	0.591
rtL180M	2 (0.8)	119 (21.5)	0	0	Antiviral resistance	<0.001

(Continued)

TABLE 2 | Continued

Sites	Subgenotypes, n (%)				Clinical issue	*P*
	A1 (*n* = 265)	A2 (*n* = 553)	QS-A3 (*n* = 48)	A4 (*n* = 3)		
rtA181					Antiviral resistance	
T	0	0	0	0		–
V	0	0	0	0		–
rtT184					Antiviral resistance	
G	0	0	0	0		–
S	0	0	0	0		–
rtS202					Antiviral resistance	
G	0	0	0	0		–
I	0	0	0	0		–
rtM204					Antiviral resistance	
I	0	9 (1.6)	0	0		0.158
V	4 (1.5)	110 (19.9)	0	0		<0.001
Total	4 (1.5)	119 (21.5)	0	0		<0.001
rtN236T	0	1 (0.2)	0	0	Antiviral resistance	0.903
rtM250V	0	0	0	0	Antiviral resistance	–
LAM-resistant mutants[c]	4 (1.5)	119 (21.5)	0	0	Antiviral resistance	<0.001
LdT-resistant mutants[d]	2 (0.8)	118 (21.3)	0	0	Antiviral resistance	<0.001
ADV-resistant mutants[e]	0	1 (0.2)	0	0	Antiviral resistance	0.903
ETV-resistant mutants[f]	0	0	0	0	Antiviral resistance	–
TDF/TAF-reduced susceptibility mutants[g]	0	0	0	0	Antiviral resistance	–

The P-value was listed as "– " when the specific mutation was found in no genotype (statistical analysis not conducted).
LAM, lamivudine; LdT, telbivudine; ADV, adefovir; ETV, entecavir; TDF, tenofovir disoproxil fumarate; TAF, tenofovir alafenamide; QS, quasi-subgenotype.
[a]*Sequences with at least one immune escape mutation.*
[b]*Sequences with at least one HCC-associated mutation.*
[c]*Amino acid substitution profile for LAM resistance: rtM204V/I.*
[d]*Amino acid substitution profile for LdT resistance: rtM204I or L180M + M204V.*
[e]*Amino acid substitution profile for ADV resistance: rtA181T/V or rtN236T.*
[f]*Amino acid substitution profile for ETV resistance: rtL180M + rtM204I/V ± rtT184S/G ± rtS202I/G ± rtM250V.*
[g]*Amino acid substitution profile for TDF/TAF-reduced susceptibility: rtA181T/V + rtN236T.*

absence of cirrhosis has been documented in chronic HBV infection cases (McMahon, 2009; Chayanupatkul et al., 2017; Rajoriya et al., 2017).

A previous meta-analysis showed that Pre-S deletions, C1653T in enhancer II, and T1753V and A1762T/G1764A in BCP are associated with increased risk of HCC compared with HBV without mutations (Liu et al., 2009). Significant differences in these mutations were observed among HBV genotypes ($P < 0.001$). Genotype G showed by far the highest frequency (≥95%) of C1653T, T1753V, and A1762T/G1764A. Genotype G is associated with more severe liver disease in HIV-HBV co-infected patients (Lacombe et al., 2006; Dao et al., 2011; Malagnino et al., 2019) but its role in HCC development remains to be established. Genotype C displayed the highest frequency of Pre-S deletions (3.3%) and the second highest rates of C1653T, T1753V and A1762T/G1764A variations. These findings are consistent with several previous reports demonstrating the association of genotype C with increased risk of HCC relative to other major HBV genotypes (Chan et al., 2004; Liu et al., 2009; Wong et al., 2013).

Interestingly, a correlation between genotype F infection and development of HCC among native Alaskan people was demonstrated in an earlier study (Livingston et al., 2007). Furthermore, Kowalec et al. (2013) showed that development of HCC in the native Alaskan population is significantly associated with specific mutations within the core promoter region of genotype F isolates (Kowalec et al., 2013). In the current study, all genotype F sequences from Alaska were classified as subgenotype F1. On the other hand, subgenotype F2 has been associated with development of HCC in Venezuelan patients (Puche et al., 2016; Pujol et al., 2020). In fact, both subgenotypes showed similar frequencies of A1762T/G1764A (F1, 29.3%; F2, 27.6%), which were higher than those observed for subgenotypes F3 and F4 (15.4% and 10.7%, $P = 0.027$). Moreover, significantly higher frequency of T1753V was observed in subgenotype F2 (44.8%, $P < 0.001$), whereas F1 was the only subgenotype that contained Pre-S deletions. The collective results suggest that HBV subgenotypes F1 and F2 are more prone to harboring HCC-associated mutations than subgenotypes F3 and F4, resulting in stronger oncogenic potential.

Likewise, the frequency of Pre-S deletions in subgenotype A1 was significantly higher than that in other A subgenotypes (3.0%, $P = 0.008$) and slightly lower than that in genotype C (3.3%). Notably, 62.5% (5 of 8) of A1 sequences with Pre-S deletions were obtained from African patients (**Supplementary Table S1**).

TABLE 3 | Clinically relevant mutations in HBV/B subgenotypes.

	Subgenotypes, n (%)						
Sites	**B1 (*n* = 57)**	**B2 (*n* = 1310)**	**QS-B3 (*n* = 207)**	**B4 (*n* = 128)**	**B5 (*n* = 52)**	**Clinical issue**	***P***
P120						Immune escape	
L	0	0	0	0	0		–
Q	0	1 (0.1)	0	0	0		0.987
S	1 (1.8)	20 (1.5)	2 (1.0)	2 (1.6)	0		0.875
T	0	4 (0.3)	0	1 (0.8)	0		0.730
Total	1 (1.8)	25 (1.9)	2 (1.0)	3 (2.3)	0		0.718
I/T126						Immune escape	
A	4 (7.0)	46 (3.5)	0	0	0		0.002
N	0	0	0	1 (0.8)	0		0.013
S	1 (1.8)	12 (0.9)	0	1 (0.8)	0		0.562
Total	5 (8.8)	58 (4.4)	0	2 (1.6)	0		0.01
Q129						Immune escape	
H	0	35 (2.7)	4 (1.9)	0	0		0.154
N	0	0	0	0	0		–
R	1 (1.8)	35 (2.7)	0	3 (2.3)	0		0.127
Total	1 (1.8)	35 (2.7)	4 (1.9)	3 (2.3)	0		0.036
G130						Immune escape	
N	1 (1.8)	0	0	2 (1.6)	1 (1.9)		<0.001
R	0	4 (0.3)	0	0	0		0.851
Total	1 (1.8)	4 (0.3)	0	2 (1.6)	1 (1.9)		0.047
M133						Immune escape	
I	0	2 (0.2)	1 (0.5)	1 (0.8)	0		0.576
L	2 (3.5)	22 (1.7)	16 (7.7)	3 (2.3)	0		<0.001
T	1 (1.8)	17 (1.3)	4 (1.9)	2 (1.6)	2 (3.8)		0.624
Total	3 (5.3)	41 (3.1)	21 (10.1)	6 (4.7)	2 (3.8)		<0.001
K141						Immune escape	
E	0	1 (0.1)	0	0	0		0.987
I	0	0	0	0	0		–
Total	0	1 (0.1)	0	0	0		0.987
P142						Immune escape	
L	0	0	0	0	0		–
S	0	1 (0.1)	0	0	0		0.987
Total	0	1 (0.1)	0	0	0		0.987
S/T143L	0	0	0	0	0	Immune escape	–
D144						Immune escape	
A	0	0	0	0	0		–
E	0	0	1 (0.5)	0	0		0.113
G	0	0	0	0	0		–
Total	0	0	1 (0.5)	0	0		0.113
G145						Immune escape	
A	0	3 (0.2)	1 (0.5)	0	0		0.889
R	0	3 (0.2)	1 (0.5)	0	0		0.889
Total	0	6 (0.5)	2 (1.0)	0	0		0.686
Escape mutants[a]	9 (15.8)	204 (15.6)	29 (14)	14 (10.9)	2 (3.8)	Immune escape	0.123
C1653T	2 (3.5)	14 (1.1)	1 (0.5)	1 (0.8)	10 (19.2)	HCC	<0.001
T1753V	10 (17.5)	45 (3.4)	10 (4.8)	2 (1.6)	11 (21.2)	HCC	<0.001
A1762T/G1764A	8 (14)	232 (17.7)	26 (12.6)	6 (4.7)	0	HCC	<0.001
Pre-S deletions	0	5 (0.4)	5 (2.4)	0	0	HCC	0.006
HCC-associated mutants[b]	11 (19.3)	250 (19.1)	34 (16.4)	7 (5.5)	16 (30.8)	HCC	<0.001
rtL180M	0	22 (1.7)	0	2 (1.6)	0	Antiviral resistance	0.252

(Continued)

TABLE 3 | Continued

Sites	Subgenotypes, n (%)					Clinical issue	*P*
	B1 (*n* = 57)	B2 (*n* = 1310)	QS-B3 (*n* = 207)	B4 (*n* = 128)	B5 (*n* = 52)		
rtA181						Antiviral resistance	
T	0	0	0	0	0		–
V	0	2 (0.2)	0	0	0		0.954
Total	0	2 (0.2)	0	0	0		0.954
rtT184						Antiviral resistance	
G	0	0	0	0	0		–
S	0	0	0	0	0		–
rtS202						Antiviral resistance	
G	0	2 (0.2)	1 (0.5)	0	0		0.807
I	0	0	0	0	0		–
Total	0	2 (0.2)	1 (0.5)	0	0		0.807
rtM204						Antiviral resistance	
I	0	17 (1.3)	0	24 (18.8)	0		<0.001
V	0	20 (1.5)	0	0	0		0.144
Total	0	37 (2.8)	0	24 (18.8)	0		<0.001
rtN236T	0	12 (0.9)	0	0	0	Antiviral resistance	0.393
rtM250V	0	0	0	0	0	Antiviral resistance	–
LAM-resistant mutants[c]	0	37 (2.8)	0	24 (18.8)	0	Antiviral resistance	<0.001
LdT-resistant mutants[d]	0	37 (2.8)	0	24 (18.8)	0	Antiviral resistance	<0.001
ADV-resistant mutants[e]	0	13 (1.0)	0	0	0	Antiviral resistance	0.350
ETV-resistant mutants[f]	0	0	0	0	0	Antiviral resistance	–
TDF/TAF-reduced susceptibility mutants[g]	0	1 (0.1)	0	0	0	Antiviral resistance	0.987

The P-value was listed as "– " when the specific mutation was found in no genotype (statistical analysis not conducted).
LAM, lamivudine; LdT, telbivudine; ADV, adefovir; ETV, entecavir; TDF, tenofovir disoproxil fumarate; TAF, tenofovir alafenamide; QS, quasi-subgenotype.
[a]*Sequences with at least one immune escape mutation.*
[b]*Sequences with at least one HCC-associated mutation.*
[c]*Amino acid substitution profile for LAM resistance: rtM204V/I.*
[d]*Amino acid substitution profile for LdT resistance: rtM204I or L180M + M204V.*
[e]*Amino acid substitution profile for ADV resistance: rtA181T/V or rtN236T.*
[f]*Amino acid substitution profile for ETV resistance: rtL180M + rtM204I/V ± rtT184S/G ± rtS202I/G ± rtM250V.*
[g]*Amino acid substitution profile for TDF/TAF-reduced susceptibility: rtA181T/V + rtN236T.*

The high frequency of Pre-S deletions in A1 may contribute to the enhanced pathogenicity of this subgenotype and its association with high rates of HCC in sub-Saharan Africa (Kew et al., 2005; Kramvis and Kew, 2007). In contrast, the predominance of genotype H among the Mexican population is associated with low prevalence of HCC (Roman and Panduro, 2013). Consistently, genotype H showed the lowest rates of T1753V and A1762T/G1764A and no Pre-S deletions, and was identified as the HBV genotype with fewer HCC-associated mutant sequences.

Although NA therapies have been successful in sustained viral suppression, long-term use of treatments with a low barrier to resistance contributes to the emergence of resistant HBV mutants. Therefore, such agents (LAM, LdT and ADV) are no longer recommended as first-line therapy (Easl, 2017; Terrault et al., 2018). However, management of NA failure remains a crucial issue, especially in low-resource countries where ETV, TDF, and TAF are not available for naïve patients or treatment-experienced patients. Consistently, the overall rates of mutations inducing resistance to LAM and LdT were higher than those for other NAs (LAM, 7.3%; LdT, 7.2%; ADV, 0.5%; ETV, 0.2%). In addition, very low rates of mutants with reduced susceptibility to TDF and TAF (0.05%, 3/6479) were observed.

The link between genotype and response to treatment with NA-based regimens is still unclear. Results from the current study showed significant differences in the rates of LAM-, LdT-, ADV-, and ETV-resistant mutants among HBV genotypes and subgenotypes. Genotype G showed the highest rates of LAM- and LdT-resistant mutants, whereas genotype E had the lowest ones (32.5 and 0.3% for both NAs, respectively, $P < 0.001$). However, these differences may not be exclusively due to intrinsic characteristics of HBV genotypes. Notably, genotype G has been frequently found in HIV-infected men who have sex with men (MSM) (Sanchez et al., 2007; Osiowy et al., 2008; Araujo et al., 2013; Cornelissen et al., 2016), suggesting a strong association with this risk group. In addition, LAM has been widely used in the treatment of both HBV and HIV viruses. This may lead to a stronger selective pressure due to the more frequent use of LAM on genotype G isolates than other genotypes, ultimately resulting in higher frequencies of LAM-resistant mutants (and LdT due to cross-resistance). Likewise, genotype E is highly endemic in most sub-Saharan Africa

TABLE 4 | Clinically relevant mutations in HBV/C subgenotypes.

	Subgenotypes, n (%)							
Sites	**C1 (*n* = 451)**	**QS-C2 (*n* = 1518)**	**C3 (*n* = 28)**	**C4 (*n* = 13)**	**C5 (*n* = 19)**	**C6–C15 (*n* = 103)**	**Clinical issue**	***P***
P120							Immune escape	
L	0	0	0	0	0	0		–
Q	0	0	0	0	0	0		–
S	0	2 (0.1)	0	0	0	0		0.976
T	1 (0.2)	7 (0.5)	0	0	0	0		0.945
Total	1 (0.2)	9 (0.6)	0	0	0	0		0.868
I/T126							Immune escape	
A	0	0	0	0	0	0		–
N	11 (2.4)	9 (0.6)	0	0	0	1 (1.0)		0.026
S	7 (1.6)	78 (5.1)	0	0	0	1 (1.0)		0.005
Total	18 (4.0)	87 (5.7)	0	0	0	2		0.166
Q129							Immune escape	
H	0	2 (0.1)	0	0	0	0		0.976
N	1 (0.2)	3 (0.2)	0	0	0	0		0.997
R	1 (0.2)	5 (0.3)	0	0	0	0		0.986
Total	2 (0.4)	10 (0.7)	0	0	0	0		0.936
G130							Immune escape	
N	1 (0.2)	24 (1.6)	0	0	0	0		0.177
R	1 (0.2)	4 (0.3)	0	0	0	0		0.994
Total	2 (0.4)	28 (1.8)	0	0	0	0		0.190
M133							Immune escape	
I	0	3 (0.2)	0	0	0	0		0.943
L	0	0	0	0	0	1 (1.0)		0.001
T	7 (1.6)	33 (2.2)	0	0	0	0		0.534
Total	7 (1.6)	36 (2.4)	0	0	0	1 (1.0)		0.674
K141							Immune escape	
E	0	1 (0.1)	0	0	0	0		0.995
I	0	2 (0.1)	0	0	0	0		0.976
Total	0	3 (0.2)	0	0	0	0		0.943
P142							Immune escape	
L	1 (0.2)	0	0	0	0	0		0.589
S	0	0	0	0	0	0		–
Total	1 (0.2)	0	0	0	0	0		0.589
S/T143L	1 (0.2)	0	0	0	0	0	Immune escape	0.589
D144							Immune escape	
A	1 (0.2)	0	0	0	0	0		0.589
E	1 (0.2)	0	1 (3.6)	0	0	0		<0.001
G	0	1 (0.1)	0	0	0	0		0.995
Total	2 (0.4)	1 (0.1)	1 (3.6)	0	0	0		0.001
G145							Immune escape	
A	3 (0.7)	22 (1.4)	1 (3.6)	0	0	0		0.445
R	4 (0.9)	7 (0.5)	0	0	0	35 (34)		<0.001
Total	7 (1.6)	29 (1.9)	1 (3.6)	0	0	35 (34)		<0.001
Escape mutants[a]	37 (8.2)	165 (10.9)	2 (7.1)	0	0	38 (36.9)	Immune escape	<0.001
C1653T	36 (8.0)	205 (13.5)	1 (3.6)	0	0	15 (14.6)	HCC	0.005
T1753V	99 (22)	262 (17.3)	2 (7.1)	1 (7.7)	0	19 (18.4)	HCC	0.026
A1762T/ G1764A	185 (41)	725 (47.8)	6 (21.4)	2 (15.4)	3 (15.8)	58 (56.3)	HCC	<0.001
Pre-S deletions	10 (2.2)	56 (3.7)	3 (10.7)	1 (7.7)	0	1 (1.0)	HCC	0.067
HCC-associated mutants[b]	201 (44.6)	782 (51.5)	7 (25)	3 (23.1)	3 (15.8)	59 (57.3)	HCC	<0.001
rtL180M	10 (2.2)	75 (4.9)	0	0	2 (10.5)	1 (1.0)	Antiviral resistance	0.023

(Continued)

TABLE 4 | Continued

Sites	Subgenotypes, n (%)						Clinical issue	*P*
	C1 (*n* = 451)	QS-C2 (*n* = 1518)	C3 (*n* = 28)	C4 (*n* = 13)	C5 (*n* = 19)	C6–C15 (*n* = 103)		
rtA181							Antiviral resistance	
T	1 (0.2)	2 (0.1)	0	0	0	0		0.994
V	0	15 (1.0)	0	0	0	0		0.296
Total	1 (0.2)	17 (1.1)	0	0	0	0		0.434
rtT184							Antiviral resistance	
G	0	0	0	0	0	0		–
S	0	0	0	0	0	1 (1.0)		0.001
Total	0	0	0	0	0	1 (1.0)		0.001
rtS202							Antiviral resistance	
G	1 (0.2)	4 (0.3)	0	0	0	0		0.994
I	0	0	0	0	0	0		–
Total	1 (0.2)	4 (0.3)	0	0	0	0		0.994
rtM204							Antiviral resistance	
I	4 (0.9)	121 (8.0)	0	0	0	0		<0.001
V	7 (1.6)	50 (3.3)	0	0	2 (10.5)	1 (1.0)		0.062
Total	11 (2.4)	171 (11.3)	0	0	2 (10.5)	1 (1.0)		<0.001
rtN236T	0	4 (0.3)	0	0	0	0	Antiviral resistance	0.899
rtM250V	0	3 (0.2)	0	0	0	0	Antiviral resistance	0.943
LAM-resistant mutants[c]	11 (2.4)	171 (11.3)	0	0	2 (10.5)	1 (1.0)	Antiviral resistance	<0.001
LdT-resistant mutants[d]	10 (2.2)	168 (11.1)	0	0	2 (10.5)	1 (1.0)	Antiviral resistance	<0.001
ADV-resistant mutants[e]	1 (0.2)	19 (1.3)	0	0	0	0	Antiviral resistance	0.343
ETV-resistant mutants[f]	0	6 (0.4)	0	0	0	1 (1.0)	Antiviral resistance	0.671
TDF/TAF-reduced susceptibility mutants[g]	0	2 (0.1)	0	0	0	0	Antiviral resistance	0.976

The P-value was listed as "–" when the specific mutation was found in no genotype (statistical analysis not conducted).
LAM, lamivudine; LdT, telbivudine; ADV, adefovir; ETV, entecavir; TDF, tenofovir disoproxil fumarate; TAF, tenofovir alafenamide; QS, quasi-subgenotype.
[a]*Sequences with at least one immune escape mutation.*
[b]*Sequences with at least one HCC-associated mutation.*
[c]*Amino acid substitution profile for LAM resistance: rtM204V/I.*
[d]*Amino acid substitution profile for LdT resistance: rtM204I or L180M + M204V.*
[e]*Amino acid substitution profile for ADV resistance: rtA181T/V or rtN236T.*
[f]*Amino acid substitution profile for ETV resistance: rtL180M + rtM204I/V ± rtT184S/G ± rtS202I/G ± rtM250V.*
[g]*Amino acid substitution profile for TDF/TAF-reduced susceptibility: rtA181T/V + rtN236T.*

(Andernach et al., 2013) where therapy is not always accessible to HBV chronic patients (Spearman et al., 2017). Consequently, weaker antiviral selection pressure on HBV isolates may occur in this region than in others where NA-based regimens are widely available, restricting the expansion of genotype E resistant mutants throughout populations.

Interestingly, significant differences in mutation profiles of same (sub)genotype among geographic regions were observed. In this regard, the differences in the rates of HCC-associated mutants between genotype C isolates from Asia and Oceania, and genotype F isolates from North America and South America, may be due to the circulation of distinct subgenotypes in these regions. On the other hand, subgenotype A2 isolates from North America showed significantly higher rates of immune escape and HCC-associated mutants than those from Western Europe. Likewise, genotype E isolates from Western Africa had significantly more HCC mutations than those from Middle Africa. Moreover, the highest rates of LAM-resistant mutants within (sub)genotypes A2, B, D, and G were observed in North America. Altogether, these findings suggest that the same (sub)genotype may have different potentials for immune escape, oncogenicity and treatment response, depending on the selective pressures (e.g., host genetic background, prophylactic and therapeutic regimens) applied in the geographic locations.

According to Carman's study (Carman, 1997), viral variants occur naturally and may have been selected over centuries, while viral mutants have been selected over a short period by human intervention (therapy or prophylaxis). It is not always clear that a sequence substitution is a functional mutation (mutant) or simply a characteristic (variant) of a specific (sub)genotype. Variants are expected to be more frequent than mutants within (sub)genotypes. Interestingly, the antiviral resistance- and immune escape mutations analyzed in our study showed much lower frequencies within (sub)genotypes than HCC-associated mutations. This might be due to the fact that therapy and prophylaxis are recent selective forces, while HCC-associated mutations have been selected during much longer time by host

TABLE 5 | Clinically relevant mutations in HBV/D subgenotypes.

	Subgenotypes, n (%)							
Sites	**D1 (*n* = 516)**	**D2 (*n* = 291)**	**D3 (*n* = 189)**	**D4 (*n* = 21)**	**D5 (*n* = 21)**	**D6 (*n* = 7)**	**Clinical issue**	***P***
P120							Immune escape	
L	0	0	1 (0.5)	0	0	0		0.475
Q	1 (0.2)	0	0	0	0	0		0.960
S	13 (2.5)	0	1 (0.5)	0	0	0		0.052
T	2 (0.4)	1 (0.3)	0	4 (19)	1 (4.8)	0		<0.001
Total	16 (3.1)	1 (0.3)	2 (1.1)	4 (19)	1 (4.8)	0		<0.001
I/T126							Immune escape	
A	0	0	0	0	0	0		–
N	1 (0.2)	1 (0.3)	0	0	0	0		0.976
S	2 (0.4)	0	0	0	0	0		0.842
Total	3 (0.6)	1 (0.3)	0	0	0	0		0.918
Q129							Immune escape	
H	2 (0.4)	0	0	0	0	0		0.006
N	0	0	0	0	0	0		–
R	9 (1.7)	0	0	4 (19)	1 (4.8)	0		<0.001
Total	11 (2.1)	0	0	4 (19)	1 (4.8)	0		<0.001
G130							Immune escape	
N	3 (0.6)	0	0	0	0	0		0.687
R	0	0	0	0	0	0		–
Total	3 (0.6)	0	0	0	0	0		0.687
M133							Immune escape	
I	2 (0.4)	1 (0.3)	0	0	0	0		0.970
L	0	1 (0.3)	0	0	0	0		0.762
T	1 (0.2)	0	0	0	0	0		0.960
Total	3 (0.6)	2 (0.7)	0	0	0	0		0.910
K141							Immune escape	
E	0	0	0	0	0	0		–
I	0	0	0	0	0	0		–
P142							Immune escape	
L	1 (0.2)	1 (0.3)	0	0	0	0		0.976
S	0	0	0	0	0	0		–
Total	1 (0.2)	1 (0.3)	0	0	0	0		0.976
S/T143L	13 (2.5)	0	6 (3.2)	0	0	0	Immune escape	0.085
D144							Immune escape	
A	0	0	0	0	0	0		–
E	2 (0.4)	2 (0.7)	0	0	0	0		0.899
G	1 (0.2)	0	0	0	0	0		0.960
Total	3 (0.6)	2 (0.7)	0	0	0	0		0.910
G145							Immune escape	
A	2 (0.4)	0	0	0	0	0		0.842
R	1 (0.2)	24 (8.2)	1 (0.5)	0	0	0		<0.001
Total	3 (0.6)	24 (8.2)	1 (0.5)	0	0	0		<0.001
Escape mutants[a]	51 (9.9)	30 (10.3)	9 (4.8)	4 (19)	2 (9.5)	0	Immune escape	0.139
C1653T	31 (6.0)	38 (13.1)	18 (9.5)	1 (4.8)	2 (9.5)	0	HCC	0.023
T1753V	118 (22.9)	36 (12.4)	14 (7.4)	0	0	1 (14.3)	HCC	<0.001
A1762T/ G1764A	146 (28.3)	42 (14.4)	29 (15.3)	3 (14.3)	4 (19)	1 (14.3)	HCC	<0.001
Pre-S deletions	0	0	0	0	0	0	HCC	–
HCC-associated mutants[b]	188 (36.4)	82 (28.2)	41 (21.7)	3 (14.3)	4 (19)	1 (14.3)	HCC	0.001
rtL180M	12 (2.3)	48 (16.5)	0	0	0	0	Antiviral resistance	<0.001
rtA181							Antiviral resistance	

(Continued)

TABLE 5 | Continued

Sites	Subgenotypes, n (%)						Clinical issue	*P*
	D1 (*n* = 516)	D2 (*n* = 291)	D3 (*n* = 189)	D4 (*n* = 21)	D5 (*n* = 21)	D6 (*n* = 7)		
T	0	0	0	0	0	0		–
V	1 (0.2)	0	0	0	0	0		0.960
Total	1 (0.2)	0	0	0	0	0		0.960
rtT184							Antiviral resistance	
G	0	0	0	0	0	0		–
S	1 (0.2)	0	0	0	0	0		0.960
Total	1 (0.2)	0	0	0	0	0		0.960
rtS202							Antiviral resistance	
G	0	1 (0.3)	0	0	0	0		0.762
I	0	1 (0.3)	0	0	0	0		0.762
Total	0	2 (0.7)	0	0	0	0		0.393
rtM204							Antiviral resistance	
I	10 (1.9)	25 (8.6)	14 (7.4)	0	1 (4.8)	0		<0.001
V	5 (1.0)	22 (7.6)	0	0	0	0		<0.001
Total	15 (2.9)	47 (16.2)	14 (7.4)	0	1 (4.8)	0		<0.001
rtN236T	0	0	0	0	0	0	Antiviral resistance	–
rtM250V	0	0	0	0	0	0	Antiviral resistance	–
LAM-resistant mutants[c]	15 (2.9)	47 (16.2)	14 (7.4)	0	1 (4.8)	0	Antiviral resistance	<0.001
LdT-resistant mutants[d]	15 (2.9)	47 (16.2)	14 (7.4)	0	1 (4.8)	0	Antiviral resistance	<0.001
ADV-resistant mutants[e]	1 (0.2)	0	0	0	0	0	Antiviral resistance	0.960
ETV-resistant mutants[f]	1 (0.2)	1 (0.3)	0	0	0	0	Antiviral resistance	0.976
TDF/TAF-reduced susceptibility mutants[g]	0	0	0	0	0	0	Antiviral resistance	–

The P-value was listed as "–" when the specific mutation was found in no genotype (statistical analysis not conducted).
LAM, lamivudine; LdT, telbivudine; ADV, adefovir; ETV, entecavir; TDF, tenofovir disoproxil fumarate; TAF, tenofovir alafenamide.
[a]*Sequences with at least one immune escape mutation.*
[b]*Sequences with at least one HCC-associated mutation.*
[c]*Amino acid substitution profile for LAM resistance: rtM204V/I.*
[d]*Amino acid substitution profile for LdT resistance: rtM204I or L180M + M204V.*
[e]*Amino acid substitution profile for ADV resistance: rtA181T/V or rtN236T.*
[f]*Amino acid substitution profile for ETV resistance: rtL180M + rtM204I/V ± rtT184S/G ± rtS202I/G ± rtM250V.*
[g]*Amino acid substitution profile for TDF/TAF-reduced susceptibility: rtA181T/V + rtN236T.*

immune pressure. Here, the vast majority of the mutations showed low frequencies within (sub)genotypes, and therefore, do not seem to be variants of any specific group. However, we have demonstrated that some mutations were biased by (sub)genotype. It is possible that over the years of viral evolution, they will become variants within these groups. This have occurred with BCP mutations in genotype G (frequency rate $\geq$ 95%), which have been stably integrated into the genome. The analysis of such a large dataset provided in this study will be helpful for further studies in establishing the likelihood that a sequence substitution is actually a mutant or a variant of a specific HBV (sub)genotype.

Finally, a number of limitations of this study need to be considered. Our analyses were based on HBV sequences from public databases, which may not be entirely representative of the globally circulating strains. Additionally, each sequence was considered a single HBV isolate, not taking into account the occurrence of serial sequences representing the same single isolate (e.g., clones), which may have generated some bias in our results. However, the large number of sequences analyzed may have overcome these limitations, at least in part. Moreover, information on whether the sequence was obtained before or after antiviral treatment or HCC development was not obtained here, which would have contributed to an accurate analysis for the relationship between mutation and clinical outcome. Nevertheless, since obtaining such information from all sequences was not possible, only mutations that have been strongly associated with immune escape, antiviral resistance, and HCC development in previous studies were selected for analysis. Lastly, due to the large variation in the number of sequences available for each (sub)genotype, over- or underrepresentation of some (sub)genotypes may have influenced the results. Therefore, although some results corroborated previous data, others should be further confirmed.

CONCLUSION

We have comprehensively investigated HBV mutations with clinical implications and their associations with genotype,

TABLE 6 | Clinically relevant mutations in HBV/F subgenotypes.

	Subgenotypes, n (%)					
Sites	**F1 (*n* = 123)**	**F2 (*n* = 29)**	**F3 (*n* = 39)**	**F4 (*n* = 56)**	**Clinical issue**	***P***
P120					Immune Escape	
L	0	0	0	0		–
Q	0	0	0	1 (1.8)		0.331
S	0	0	0	0		–
T	0	0	0	0		–
Total	0	0	0	1 (1.8)		0.331
I/T126					Immune escape	
A	0	1 (3.4)	1 (2.6)	0		0.140
N	0	0	0	0		–
S	0	0	0	0		–
Total	0	1 (3.4)	1 (2.6)	0		0.140
Q129					Immune escape	
H	0	0	0	0		–
N	0	0	0	0		–
R	0	0	0	0		–
G130					Immune escape	
N	0	0	0	0		–
R	0	1 (3.4)	0	0		0.056
Total	0	1 (3.4)	0	0		0.056
M133					Immune escape	
I	0	0	0	0		–
L	0	0	0	0		–
T	0	0	0	0		–
K141					Immune escape	
E	0	0	0	0		–
I	0	0	0	0		–
P142					Immune escape	
L	0	0	0	0		–
S	0	0	0	0		–
S/T143L	0	0	0	0	Immune escape	–
D144					Immune escape	
A	0	0	0	0		–
E	1 (0.8)	0	0	0		0.798
G	0	0	0	0		–
Total	1 (0.8)	0	0	0		0.798
G145					Immune escape	
A	0	0	0	0		–
R	0	0	0	0		–
Escape mutants[a]	1 (0.8)	2 (6.9)	1 (2.6)	1 (1.8)	Immune escape	0.216
C1653T	3 (2.4)	2 (6.9)	0	1 (1.8)	HCC	0.319
T1753V	6 (4.9)	13 (44.8)	4 (10.3)	9 (16.1)	HCC	<0.001
A1762T/G1764A	36 (29.3)	8 (27.6)	6 (15.4)	6 (10.7)	HCC	0.027
Pre-S deletions	3 (2.4)	0	0	0	HCC	0.382
HCC-associated mutants[b]	38 (30.9)	15 (51.7)	7 (17.9)	15 (26.8)	HCC	0.024
rtL180M	5 (4.1)	2 (6.9)	0	0	Antiviral resistance	0.159
rtA181					Antiviral resistance	
T	0	0	0	0		–
V	0	0	0	0		–
rtT184					Antiviral resistance	
G	0	0	0	0		–
S	0	0	0	0		–

(Continued)

TABLE 6 | Continued

Sites	Subgenotypes, n (%)				Clinical issue	*P*
	F1 (*n* = 123)	F2 (*n* = 29)	F3 (*n* = 39)	F4 (*n* = 56)		
rtS202					Antiviral resistance	
G	0	0	0	0		–
I	0	0	0	0		–
rtM204					Antiviral resistance	
I	0	0	0	0		–
V	4 (3.3)	2 (6.9)	0	0		0.161
Total	4 (3.3)	2 (6.9)	0	0		0.161
rtN236T	0	0	0	0	Antiviral resistance	–
rtM250V	0	0	0	0	Antiviral resistance	–
LAM-resistant mutants[c]	4 (3.3)	2 (6.9)	0	0	Antiviral resistance	0.161
LdT-resistant mutants[d]	4 (3.3)	2 (6.9)	0	0	Antiviral resistance	0.161
ADV-resistant mutants[e]	0	0	0	0	Antiviral resistance	–
ETV-resistant mutants[f]	0	0	0	0	Antiviral resistance	–
TDF/TAF-reduced susceptibility mutants[g]	0	0	0	0	Antiviral resistance	–

The P-value was listed as "– " when the specific mutation was found in no genotype (statistical analysis not conducted).
LAM, lamivudine; LdT, telbivudine; ADV, adefovir; ETV, entecavir; TDF, tenofovir disoproxil fumarate; TAF, tenofovir alafenamide.
[a]*Sequences with at least one immune escape mutation.*
[b]*Sequences with at least one HCC-associated mutation.*
[c]*Amino acid substitution profile for LAM resistance: rtM204V/I.*
[d]*Amino acid substitution profile for LdT resistance: rtM204I or L180M + M204V.*
[e]*Amino acid substitution profile for ADV resistance: rtA181T/V or rtN236T.*
[f]*Amino acid substitution profile for ETV resistance: rtL180M + rtM204I/V ± rtT184S/G ± rtS202I/G ± rtM250V.*
[g]*Amino acid substitution profile for TDF/TAF-reduced susceptibility: rtA181T/V + rtN236T.*

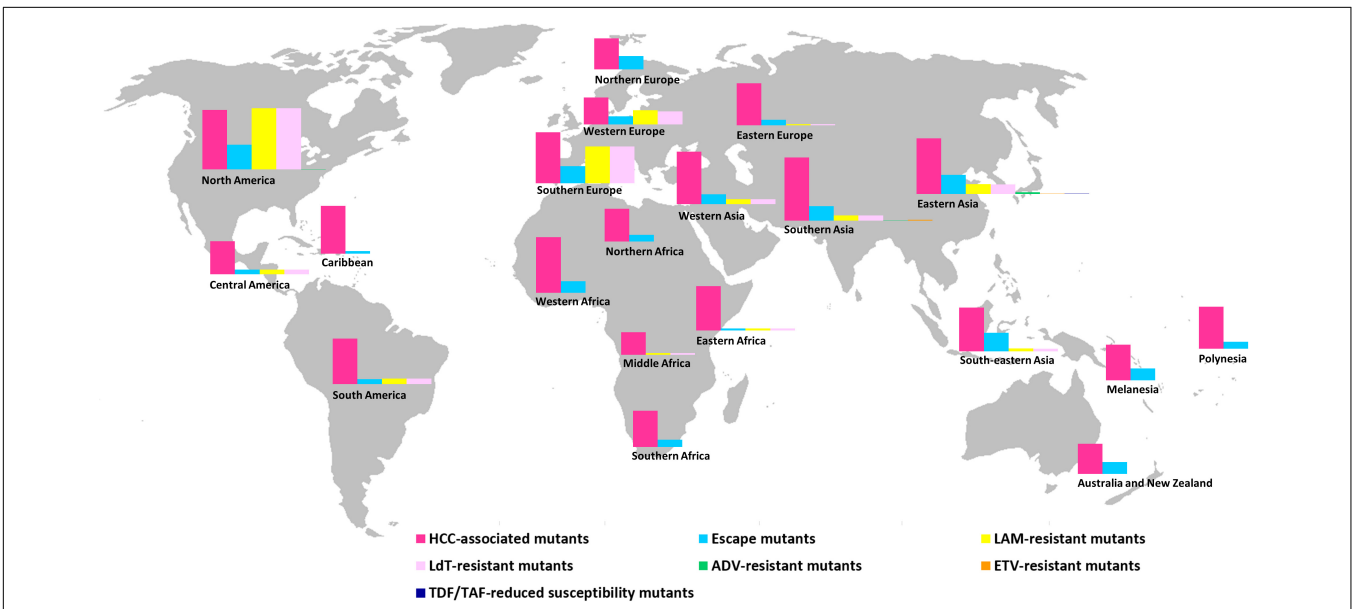

FIGURE 3 | Prevalence of HBV mutants in different geographic regions based on 6,207 complete genome sequences present in the HBVdb database. The map was reconstructed using Wikimedia Commons (https://commons.wikimedia.org/wiki/File:BlankMap-World-noborders.png), this figure is similar but not identical to the original image and is therefore for illustrative purposes only.

subgenotype and geographic region. We demonstrated heterogeneous prevalence of these mutations among HBV (sub)genotypes and locations, which may contribute to improvement of diagnostic procedures, immunization programs, therapeutic protocols, and disease prognosis. Since HBV (sub)genotypes and mutations are key determinants in the management and treatment of hepatitis B, reliable biomarkers and convenient methods need to be established to monitor viral factors and achieve a personalized treatment approach.

TABLE 7 | Clinically relevant HBV mutants by geographic region.

	Geographic Regions, n (%)																				
Clinical issue	**Australia and New Zealand (n = 51)**	**Caribbean (n = 61)**	**Central America (n = 70)**	**Central Asia (n = 13)**	**Eastern Africa (n = 69)**	**Eastern Asia (n = 3118)**	**Eastern Europe (n = 139)**	**Melanesia (n = 13)**	**Middle Africa (n = 96)**	**Northern Africa (n = 47)**	**North America (n = 443)**	**Northern Europe (n = 35)**	**Polynesia (n = 22)**	**South America (n = 342)**	**South-eastern Asia (n = 616)**	**Southern Africa (n = 106)**	**Southern Asia (n = 458)**	**Southern Europe (n = 54)**	**Western Africa (n = 174)**	**Western Asia (n = 126)**	**Western Europe (n = 154)**	**P**
LAM-resistant mutants[a]	0	0	2 (2.9)	0	1 (1.4)	199 (6.4)	1 (0.7)	0	1 (1.0)	0	181 (40.9)	0	0	12 (3.5)	12 (1.9)	0	15 (3.3)	13 (24)	0	4 (3.2)	14 (9.1)	< 0.001
LdT-resistant mutants[b]	0	0	2 (2.9)	0	1 (1.4)	196 (6.3)	1 (0.7)	0	1 (1.0)	0	181 (40.9)	0	0	12 (3.5)	11 (1.8)	0	15 (3.3)	13 (24)	0	4 (3.2)	13 (8.4)	< 0.001
ADV-resistant mutants[c]	0	0	0	0	0	33 (1.1)	0	0	0	0	1 (0.2)	0	0	0	0	0	1 (0.2)	0	0	0	0	0. 112
ETV-resistant mutants[d]	0	0	0	0	0	8 (0.3)	0	0	0	0	0	0	0	0	0	0	2 (0.4)	0	0	0	0	0. 991
TDF/TAF-reduced susceptibility mutants[e]	0	0	0	0	0	3 (0.1)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.000
Escape mutants[f]	4 (7.8)	1 (1.6)	2 (2.9)	0	1 (1.4)	385 (12.3)	5 (3.6)	1 (7.7)	0	2 (4.3)	72 (16.3)	3 (8.6)	1 (4.5)	11 (3.2)	74 (12)	5 (4.7)	43 (9.4)	6 (11.1)	13 (7.5)	8 (6.3)	8 (5.2)	< 0.001
HCC-associated mutants[g]	10 (19.6)	19 (31.1)	15 (21.4)	0	20 (29)	1139 (36.5)	38 (27.3)	3 (23.1)	14 (14.6)	10 (21.3)	164 (37)	7 (20)	6 (27.3)	101 (29.5)	175 (28.4)	25 (23.6)	188 (41)	18 (33.3)	63 (36.2)	43 (34.1)	27 (17.5)	<0.001

LAM. lamivudine; LdT. telbivudine; ADV. adefovir; ETV. entecavir; TDF. tenofovir disoproxil fumarate; TAF. tenofovir alafenamide.

[a]*Amino acid substitution profile for LAM resistance: rtM204V/I.*

[b]*Amino acid substitution profile for LdT resistance: rtM204I or L180M + M204V.*

[c]*Amino acid substitution profile for ADV resistance: rtA181T/V or rtN236T.*

[d]*Amino acid substitution profile for ETV resistance: rtL180M + rtM204I/V ± rtT184S/G ± rtS202I/G ± rtM250V.*

[e]*Amino acid substitution profile for TDF/TAF-reduced susceptibility: rtA181T/V + rtN236T.*

[f]*Sequences with at least one immune escape mutation.*

[g]*Sequences with at least one HCC-associated mutation.*

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

NA and NS: conceptualization. NA, ST, and NS: methodology, formal Analysis, and writing – review and editing. NA: writing – original draft preparation. All authors read and approved the final manuscript.

FUNDING

NA was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), grant number 428676/2018-9. ST was supported by CNPq, grant number 304948/2018-7.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.616023/full#supplementary-material

Supplementary Figure 1 | Phylogenetic analysis of HBV complete genome sequences representative of genotypes A–J. Maximum likelihood phylogenetic tree of **(A)** 871 complete genome sequences of genotype A from HBVdb database plus 104 reference sequences; **(B)** 1755 complete genome sequences of genotype B from HBVdb database plus 104 reference sequences; **(C)** 2187 complete genome sequences of genotype C from HBVdb database plus 104 reference sequences; **(D)** 1059 complete genome sequences of genotype D from HBVdb database plus 104 reference sequences; **(E)** 249 complete genome sequences of genotype F from HBVdb database plus 104 reference sequences. (Sub)genotypes are indicated next to their respective clusters. Sequences whose subgenotype could not be determined were indicated as "not determined." The numbers in branches indicate the statistical support (aLRT value). Only aLRT values $\geq$ 80 are shown.

Supplementary Table 1 | HBV complete genome sequences dataset.

Supplementary Table 2 | Clinically relevant HBV mutants by (sub)genotype and geographic region.

REFERENCES

Andernach, I. E., Hunewald, O. E., and Muller, C. P. (2013). Bayesian inference of the evolution of HBV/E. *PLoS One* 8:e81690. doi: 10.1371/journal.pone.0081690

Anisimova, M., and Gascuel, O. (2006). Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. *Syst. Biol.* 55, 539–552. doi: 10.1080/10635150600755453

Araujo, N. M. (2015). Hepatitis B virus intergenotypic recombinants worldwide: an overview. *Infect. Genet. Evol.* 36, 500–510. doi: 10.1016/j.meegid.2015.08.024

Araujo, N. M., Araujo, O. C., Silva, E. M., Villela-Nogueira, C. A., Nabuco, L. C., Parana, R., et al. (2013). Identification of novel recombinants of hepatitis B virus genotypes F and G in human immunodeficiency virus-positive patients from Argentina and Brazil. *J. Gen. Virol.* 94(Pt. 1), 150–158. doi: 10.1099/vir.0.047324-0

Araujo, N. M., Waizbort, R., and Kay, A. (2011). Hepatitis B virus infection from an evolutionary point of view: how viral, host, and environmental factors shape genotypes and subgenotypes. *Infect. Genet. Evol.* 11, 1199–1207. doi: 10.1016/j.meegid.2011.04.017

Avellon, A., and Echevarria, J. M. (2006). Frequency of hepatitis B virus 'a' determinant variants in unselected Spanish chronic carriers. *J. Med. Virol.* 78, 24–36. doi: 10.1002/jmv.20516

Carman, W. F. (1997). The clinical significance of surface antigen variants of hepatitis B virus. *J. Viral Hepat.* 4(Suppl. 1), 11–20. doi: 10.1111/j.1365-2893.1997.tb00155.x

Carman, W. F., Zanetti, A. R., Karayiannis, P., Waters, J., Manzillo, G., Tanzi, E., et al. (1990). Vaccine-induced escape mutant of hepatitis B virus. *Lancet* 336, 325–329. doi: 10.1016/0140-6736(90)91874-a

Chan, H. L., Hui, A. Y., Wong, M. L., Tse, A. M., Hung, L. C., Wong, V. W., et al. (2004). Genotype C hepatitis B virus infection is associated with an increased risk of hepatocellular carcinoma. *Gut* 53, 1494–1498. doi: 10.1136/gut.2003.033324

Chayanupatkul, M., Omino, R., Mittal, S., Kramer, J. R., Richardson, P., Thrift, A. P., et al. (2017). Hepatocellular carcinoma in the absence of cirrhosis in patients with chronic hepatitis B virus infection. *J. Hepatol.* 66, 355–362. doi: 10.1016/j.jhep.2016.09.013

Chou, Y. C., Yu, M. W., Wu, C. F., Yang, S. Y., Lin, C. L., Liu, C. J., et al. (2008). Temporal relationship between hepatitis B virus enhancer II/basal core promoter sequence variation and risk of hepatocellular carcinoma. *Gut* 57, 91–97. doi: 10.1136/gut.2006.114066

Cooreman, M. P., Leroux-Roels, G., and Paulij, W. P. (2001). Vaccine- and hepatitis B immune globulin-induced escape mutations of hepatitis B virus surface antigen. *J. Biomed. Sci.* 8, 237–247. doi: 10.1007/BF02256597

Cornelissen, M., Zorgdrager, F., Bruisten, S. M., Bakker, M., Berkhout, B., and van der Kuyl, A. C. (2016). Widespread hepatitis B virus genotype G (HBV-G) infection during the early years of the HIV epidemic in the Netherlands among men who have sex with men. *BMC Infect. Dis.* 16:268. doi: 10.1186/s12879-016-1599-7

Dao, D. Y., Balko, J., Attar, N., Neak, E., Yuan, H. J., Lee, W. M., et al. (2011). Hepatitis B virus genotype G: prevalence and impact in patients co-infected with human immunodeficiency virus. *J. Med. Virol.* 83, 1551–1558. doi: 10.1002/jmv.22160

Di Lello, F. A., Ridruejo, E., Martinez, A. P., Perez, P. S., Campos, R. H., and Flichman, D. M. (2019). Molecular epidemiology of hepatitis B virus mutants associated with vaccine escape, drug resistance and diagnosis failure. *J. Viral Hepat.* 26, 552–560. doi: 10.1111/jvh.13052

Easl. (2017). EASL 2017 Clinical Practice Guidelines on the management of hepatitis B virus infection. *J. Hepatol.* 67, 370–398. doi: 10.1016/j.jhep.2017.03.021

Hayer, J., Jadeau, F., Deleage, G., Kay, A., Zoulim, F., and Combet, C. (2013). HBVdb: a knowledge database for Hepatitis B Virus. *Nucl. Acids Res.* 41, D566–D570. doi: 10.1093/nar/gks1022

He, C., Nomura, F., Itoga, S., Isobe, K., and Nakai, T. (2001). Prevalence of vaccine-induced escape mutants of hepatitis B virus in the adult population in China: a prospective study in 176 restaurant employees. *J. Gastroenterol. Hepatol.* 16, 1373–1377. doi: 10.1046/j.1440-1746.2001.02654.x

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., and Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14, 587–589. doi: 10.1038/nmeth.4285

Kao, J. H., Chen, P. J., and Chen, D. S. (2010). Recent advances in the research of hepatitis B virus-related hepatocellular carcinoma: epidemiologic and molecular biological aspects. *Adv. Cancer Res.* 108, 21–72. doi: 10.1016/B978-0-12-380888-2.00002-9

Kay, A., and Zoulim, F. (2007). Hepatitis B virus genetic variability and evolution. *Virus Res.* 127, 164–176. doi: 10.1016/j.virusres.2007.02.021

Kew, M. C., Kramvis, A., Yu, M. C., Arakawa, K., and Hodkinson, J. (2005). Increased hepatocarcinogenic potential of hepatitis B virus genotype A in Bantu-speaking sub-saharan Africans. *J. Med. Virol.* 75, 513–521. doi: 10.1002/jmv.20311

Kim, B. K., Revill, P. A., and Ahn, S. H. (2011). HBV genotypes: relevance to natural history, pathogenesis and treatment of chronic hepatitis B. *Antivir Ther.* 16, 1169–1186. doi: 10.3851/IMP1982

Kowalec, K., Minuk, G. Y., Borresen, M. L., Koch, A., McMahon, B. J., Simons, B., et al. (2013). Genetic diversity of hepatitis B virus genotypes B6, D and F among circumpolar indigenous individuals. *J. Viral Hepat.* 20, 122–130. doi: 10.1111/j.1365-2893.2012.01632.x

Kramvis, A. (2014). Genotypes and genetic variability of hepatitis B virus. *Intervirology* 57, 141–150. doi: 10.1159/000360947

Kramvis, A., and Kew, M. C. (2005). Relationship of genotypes of hepatitis B virus to mutations, disease progression and response to antiviral therapy. *J. Viral Hepat.* 12, 456–464. doi: 10.1111/j.1365-2893.2005.00624.x

Kramvis, A., and Kew, M. C. (2007). Molecular characterization of subgenotype A1 (subgroup Aa) of hepatitis B virus. *Hepatol. Res.* 37, S27–S32. doi: 10.1111/j.1872-034X.2007.00100.x

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096

Lacombe, K., Massari, V., Girard, P. M., Serfaty, L., Gozlan, J., Pialoux, G., et al. (2006). Major role of hepatitis B genotypes in liver fibrosis during coinfection with HIV. *AIDS* 20, 419–427. doi: 10.1097/01.aids.0000200537.86984.0e

Lazarevic, I. (2014). Clinical implications of hepatitis B virus mutations: recent advances. *World J. Gastroenterol.* 20, 7653–7664. doi: 10.3748/wjg.v20.i24.7653

Lazarevic, I., Banko, A., Miljanovic, D., and Cupic, M. (2019). Immune-escape hepatitis b virus mutations associated with viral reactivation upon immunosuppression. *Viruses* 11:778. doi: 10.3390/v11090778

Levrero, M., and Zucman-Rossi, J. (2016). Mechanisms of HBV-induced hepatocellular carcinoma. *J. Hepatol.* 64 1(Suppl.), S84–S101. doi: 10.1016/j.jhep.2016.02.021

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

Liu, S., Zhang, H., Gu, C., Yin, J., He, Y., Xie, J., et al. (2009). Associations between hepatitis B virus mutations and the risk of hepatocellular carcinoma: a meta-analysis. *J. Natl. Cancer Inst.* 101, 1066–1082. doi: 10.1093/jnci/djp180

Livingston, S. E., Simonetti, J. P., McMahon, B. J., Bulkow, L. R., Hurlburt, K. J., Homan, C. E., et al. (2007). Hepatitis B virus genotypes in Alaska Native people with hepatocellular carcinoma: preponderance of genotype F. *J. Infect. Dis.* 195, 5–11. doi: 10.1086/509894

Ma, Q., and Wang, Y. (2012). Comprehensive analysis of the prevalence of hepatitis B virus escape mutations in the major hydrophilic region of surface antigen. *J. Med. Virol.* 84, 198–206. doi: 10.1002/jmv.23183

Malagnino, V., Bottero, J., Miailhes, P., Lascoux-Combe, C., Girard, P. M., Zoulim, F., et al. (2019). Hepatitis B virus genotype G and liver fibrosis progression in chronic hepatitis B and human immunodeficiency virus coinfection. *J. Med. Virol.* 91, 630–641. doi: 10.1002/jmv.25360

McMahon, B. J. (2009). The influence of hepatitis B virus genotype and subgenotype on the natural history of chronic hepatitis B. *Hepatol. Int.* 3, 334–342. doi: 10.1007/s12072-008-9112-z

Nguyen, L. T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300

Osiowy, C., Gordon, D., Borlang, J., Giles, E., and Villeneuve, J. P. (2008). Hepatitis B virus genotype G epidemiology and co-infection with genotype A in Canada. *J. Gen. Virol.* 89(Pt. 12), 3009–3015. doi: 10.1099/vir.0.2008/005124-0

Paraskevis, D., Angelis, K., Magiorkinis, G., Kostaki, E., Ho, S. Y., and Hatzakis, A. (2015). Dating the origin of hepatitis B virus reveals higher substitution rate and adaptation on the branch leading to F/H genotypes. *Mol. Phylogenet. Evol.* 93, 44–54. doi: 10.1016/j.ympev.2015.07.010

Parkin, D. M. (2006). The global health burden of infection-associated cancers in the year 2002. *Int. J. Cancer* 118, 3030–3044. doi: 10.1002/ijc.21731

Pollicino, T., Isgro, G., Di Stefano, R., Ferraro, D., Maimone, S., Brancatelli, S., et al. (2009). Variability of reverse transcriptase and overlapping S gene in hepatitis B virus isolates from untreated and lamivudine-resistant chronic hepatitis B patients. *Antivir. Ther.* 14, 649–654.

Puche, M. L., Kay-Valero, S., Michelli, P., Oropeza, M. D., Loureiro, C. L., Devesa, M., et al. (2016). Genetic diversity of hepatitis B virus and mutations associated to hepatocellular carcinoma in patients from Venezuela, with different stages of liver disease. *Invest. Clin.* 57, 38–46.

Pujol, F., Jaspe, R. C., Loureiro, C. L., and Chemin, I. (2020). Hepatitis B virus American genotypes: pathogenic variants ? *Clin. Res. Hepatol. Gastroenterol.* 44, 825-835. doi: 10.1016/j.clinre.2020.04.018

Qin, Y., and Liao, P. (2018). Hepatitis B virus vaccine breakthrough infection: surveillance of S gene mutants of HBV. *Acta Virol.* 62, 115–121. doi: 10.4149/av_2018_210

Rajoriya, N., Combet, C., Zoulim, F., and Janssen, H. L. A. (2017). How viral genetic variants and genotypes influence disease and treatment outcome of chronic hepatitis B. Time for an individualised approach? *J. Hepatol.* 67, 1281–1297. doi: 10.1016/j.jhep.2017.07.011

Revill, P. A., Tu, T., Netter, H. J., Yuen, L. K. W., Locarnini, S. A., and Littlejohn, M. (2020). The evolution and clinical impact of hepatitis B virus genome diversity. *Nat. Rev. Gastroenterol. Hepatol.* 17, 618-634. doi: 10.1038/s41575-020-0296-6

Roman, S., and Panduro, A. (2013). HBV endemicity in Mexico is associated with HBV genotypes H and G. *World J. Gastroenterol.* 19, 5446–5453. doi: 10.3748/wjg.v19.i33.5446

Sanchez, L. V., Tanaka, Y., Maldonado, M., Mizokami, M., and Panduro, A. (2007). Difference of hepatitis B virus genotype distribution in two groups of mexican patients with different risk factors. High prevalence of genotype H and G. *Intervirology* 50, 9–15. doi: 10.1159/000096307

Sarin, S. K., Kumar, M., Lau, G. K., Abbas, Z., Chan, H. L., Chen, C. J., et al. (2016). Asian-Pacific clinical practice guidelines on the management of hepatitis B: a 2015 update. *Hepatol. Int.* 10, 1–98. doi: 10.1007/s12072-015-9675-4

Sayan, M., Senturk, O., Akhan, S. C., Hulagu, S., and Cekmen, M. B. (2010). Monitoring of hepatitis B virus surface antigen escape mutations and concomitantly nucleos(t)ide analog resistance mutations in Turkish patients with chronic hepatitis B. *Int. J. Infect. Dis.* 14(Suppl. 3), e136–e141. doi: 10.1016/j.ijid.2009.11.039

Seeger, C., Zoulim, F., and Mason, W. S. (2013). "Hepadnaviruses," in *Fields virology*, 6th Edn, eds D. M. Knipe, and P. M. Howley, (Philadelphia, PA: Lippincott Williams & Wilkins), 2185–2221.

Spearman, C. W., Afihene, M., Ally, R., Apica, B., Awuku, Y., Cunha, L., et al. (2017). Hepatitis B in sub-Saharan Africa: strategies to achieve the 2030 elimination targets. *Lancet Gastroenterol. Hepatol.* 2, 900–909. doi: 10.1016/S2468-1253(17)30295-9

Spitz, N., Mello, F. C. A., Moreira, A. S., Gusatti, C. S., Martins, R. M. B., Gomes, S. A., et al. (2019). Reconstruction of the spatial and temporal dynamics of hepatitis B virus genotype D in the Americas. *PLoS One* 14:e0220342. doi: 10.1371/journal.pone.0220342

Terrault, N. A., Lok, A. S. F., McMahon, B. J., Chang, K. M., Hwang, J. P., Jonas, M. M., et al. (2018). Update on prevention, diagnosis, and treatment of chronic hepatitis B: AASLD 2018 hepatitis B guidance. *Hepatology* 67, 1560–1599. doi: 10.1002/hep.29800

Tseng, T. C., Liu, C. J., Yang, H. C., Chen, C. L., Yang, W. T., Tsai, C. S., et al. (2015). Higher proportion of viral basal core promoter mutant increases the risk of liver cirrhosis in hepatitis B carriers. *Gut* 64, 292–302. doi: 10.1136/gutjnl-2014-306977

WHO, (2020). *World Health Organization - Hepatitis B Key facts* [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/hepatitis-b (Accessed August 25, 2020).

Wong, G. L., Chan, H. L., Yiu, K. K., Lai, J. W., Chan, V. K., Cheung, K. K., et al. (2013). Meta-analysis: the association of hepatitis B virus genotypes and hepatocellular carcinoma. *Aliment Pharmacol. Ther.* 37, 517–526. doi: 10.1111/apt.12207

Yan, B., Lv, J., Feng, Y., Liu, J., Ji, F., Xu, A., et al. (2017). Temporal trend of hepatitis B surface mutations in the post-immunization period: 9 years of surveillance (2005-2013) in eastern China. *Sci. Rep.* 7:6669. doi: 10.1038/s41598-017-07085-z

Zehender, G., Ebranati, E., Gabanelli, E., Sorrentino, C., Lo Presti, A., Tanzi, E., et al. (2014). Enigmatic origin of hepatitis B virus: an ancient travelling companion or a recent encounter? *World J. Gastroenterol.* 20, 7622–7634. doi: 10.3748/wjg.v20.i24.7622

Zhang, Z. H., Wu, C. C., Chen, X. W., Li, X., Li, J., and Lu, M. J. (2016). Genetic variation of hepatitis B virus and its significance for pathogenesis. *World J. Gastroenterol.* 22, 126–144. doi: 10.3748/wjg.v22.i1.126

Zoulim, F. (2011). Hepatitis B virus resistance to antiviral drugs: where are we going? *Liver Int.* 31(Suppl. 1), 111–116. doi: 10.1111/j.1478-3231.2010.02399.x

Zoulim, F., and Locarnini, S. (2009). Hepatitis B virus resistance to nucleos(t)ide analogues. *Gastroenterology* 137, 1593-1608.e1-2.

Zuckerman, J. N., and Zuckerman, A. J. (2003). Mutations of the surface protein of hepatitis B virus. *Antiviral Res.* 60, 75–78. doi: 10.1016/j.antiviral.2003.08.013

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 15 January 2021
doi: 10.3389/fmicb.2020.602296

Analysis of Hepatitis B Virus Genotype D in Greenland Suggests the Presence of a Novel Quasi-Subgenotype

Adriano de Bernardi Schneider[1]*, *Carla Osiowy*[2], *Reilly Hostager*[1], *Henrik Krarup*[3,4,5], *Malene Børresen*[6], *Yasuhito Tanaka*[7], *Taylor Morriseau*[2†] *and Joel O. Wertheim*[1]

[1] Department of Medicine, University of California San Diego, San Diego, CA, United States, [2] National Microbiology Laboratory, Public Health Agency of Canada, Winnipeg, MB, Canada, [3] Department of Molecular Diagnostics, Aalborg University Hospital, Aalborg, Denmark, [4] Department of Medical Gastroenterology, Aalborg University Hospital, Aalborg, Denmark, [5] Clinical Institute, Aalborg University, Aalborg, Denmark, [6] Department of Epidemiological Research, Statens Serum Institut, Copenhagen, Denmark, [7] Department of Virology & Liver, Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan

OPEN ACCESS

Edited by:
Manuela Sironi, Eugenio Medea (IRCCS), Italy

Reviewed by:
Mahmoud Reza Pourkarim, KU Leuven, Belgium
Eleonora Cella, University of Central Florida, United States

***Correspondence:**
Adriano de Bernardi Schneider
adeberna@ucsd.edu

†Present address:
Taylor Morriseau, Department of Pharmacology and Therapeutics, University of Manitoba, Winnipeg, MB, Canada

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *02 September 2020*
Accepted: *14 December 2020*
Published: *15 January 2021*

Citation:
de Bernardi Schneider A, Osiowy C, Hostager R, Krarup H, Børresen M, Tanaka Y, Morriseau T and Wertheim JO (2021) Analysis of Hepatitis B Virus Genotype D in Greenland Suggests the Presence of a Novel Quasi-Subgenotype. Front. Microbiol. 11:602296. doi: 10.3389/fmicb.2020.602296

A disproportionate number of Greenland's Inuit population are chronically infected with Hepatitis B virus (HBV; 5–10%). HBV genotypes B and D are most prevalent in the circumpolar Arctic. Here, we report 39 novel HBV/D sequences from individuals residing in southwestern Greenland. We performed phylodynamic analyses with ancient HBV DNA calibrators to investigate the origin and relationship of these taxa to other HBV sequences. We inferred a substitution rate of 1.4×10^{-5} [95% HPD 8.8×10^{-6}, 2.0×10^{-5}] and a time to the most recent common ancestor of 629 CE [95% HPD 37–1138 CE]. The Greenland taxa form a sister clade to HBV/D2 sequences, specifically New Caledonian and Indigenous Taiwanese sequences. The Greenland sequences share amino acid signatures with subgenotypes D1 and D2 and ∼97% sequence identity. Our results suggest the classification of these novel sequences does not fit within the current nomenclature. Thus, we propose these taxa be considered a novel quasi-subgenotype.

Keywords: evolution, HBV, phylogenetics, phylogenomics, hepatitis

1. INTRODUCTION

Hepatitis B virus (HBV) is an ancient virus that remains a substantial public health burden. Transmitted by bodily fluid, it is one of the most prevalent viruses in the world, infecting over 2 billion people, 257 million of whom are chronically infected (WHO, 2019).

HBV is classified into nine distinct genotypes (A through I) and a putative 10th genotype (J) (Kramvis, 2014). Precedent sorting of genotype identity is calculated by a distance-based sequence threshold of 8% (Okamoto et al., 1988; Zhou and Holmes, 2007; Locarnini et al., 2013; Araujo, 2015; Lin and Kao, 2015; Littlejohn et al., 2016). Subgenotypes are traditionally assigned based on a 4–7.5% threshold across the full genome (McNaughton et al., 2020), though new phylogenetic inference methods to determine subgenotype have been suggested (Yin et al., 2019). McNaughton et al. (2020) notes that all genotypes have subgenotype sequence divergence rates of 3–8%, with the exception of genotype D, which is 2–4%, indicating a lack of long evolutionary branches separating well-defined clades. Recent suggestions for HBV taxonomic nomenclature include assigning the term "quasi-subgenotype" to a lineage that does not meet the criteria nor clearly belongs to a defined subgenotype based on <4% nucleotide divergence among complete genomes following detailed phylogenetic analysis (Pourkarim et al., 2010).

HBV was considered hyperendemic in the western circumpolar Arctic (Alaska, Canada, and Greenland) before the introduction of HBV vaccination programs (Lavanchy, 2008) with the burden of viral hepatitis disproportionately borne by Indigenous populations. Prior to the introduction of HBV vaccination programs in the Canadian Arctic in the 1990s, HBV prevalence was ∼3% among Inuit populations (Osiowy et al., 2013); moreover, the risk of lifetime exposure to HBV was 25% or five times higher than the non-Indigenous population (Tulisov et al., 2007). Approximately 5–10% of Greenland's Inuit population remains chronically infected with HBV (Børresen et al., 2011; Rex et al., 2015). Since 1965, serosurveys of HBV exposure among Greenland residents have demonstrated high rates of exposure among adults and chronic infection rates 14–40 times higher than the US and northern Europe (Tulisov et al., 2007). Genotypes B and D predominate in the North American circumpolar Arctic (Osiowy et al., 2013). Western Greenland's infections are represented by two genotypes: a majority genotype D and a minority genotype A (Osiowy et al., 2013). While infancy is regarded as the most receptive stage of HBV infection, the incidence of active infection increases in adolescence (Tulisov et al., 2007). This highlights transmission by sexual contact, which is the primary method for HIV transmission in Greenland (Bjorn-Mortensen et al., 2013).

The most prevalent and endemic subgenotype, B5 (formerly B6), within the Alaska Native and Inuit populations of the western circumpolar Arctic may have originated coincident with the rapid movement of the Thule from regions of Alaska to the eastern Arctic (Bouckaert et al., 2017) ∼1,000 years before present (YBP). Greenland Inuit harboring genotype D HBV (HBV/D) infections, such as those discussed in this study, may represent another endemic lineage with high rates of chronic infection and, importantly, a unique coalescent origin from other defined HBV/D subgenotypes. Known presence of HBV/D in the Arctic is limited to Greenland (Langer et al., 1997), Western Canada, and Alaska (Livingston et al., 2007; Osiowy et al., 2011). Association of HBV/D with First Nation Dene living in the western Arctic has also been observed, suggesting a separate introduction from Greenlandic Inuit (Osiowy et al., 2011).

Additionally, Inuit residing in the Canadian Arctic infected with subgenotype HBV/B5 experience less severe clinical outcomes typical of non-Indigenous people infected with other HBV/B subgenotypes (Minuk and Uhanova, 2003). This reduced risk phenomenon for Inuit infected with HBV/D is unknown. Historically, genotype D has poorer clinical prognoses and a lower response to therapy (McMahon, 2008; Yin et al., 2019). Specific basal core promoter (BCP) nucleotide mutations, T1762 and A1764, are important markers associated with liver disease progression and development of hepatocellular carcinoma (HCC) (Yang et al., 2016), and they are highly associated with HBV/D (Lin and Kao, 2015).

Despite its broad geographical distribution and large volume of sequence data, the evolutionary history of HBV remains ambiguous. The difference in substitution rates depending on whether the infection is acute, chronic, or within-vs.-between hosts (Vrancken et al., 2014) is compounded by discrepant calibrations (i.e., fossil-based or tip-dated). Ancient DNA (aDNA) calibration for exogenous viruses has been shown to tease apart some of the evolutionary mystery of HBV (Mühlemann et al., 2018), revealing a surprising lack of temporal genetic change in the last half millennium (Krause-Kyora et al., 2018). HBV is a suitable candidate for aDNA external calibration because of its high viremic levels, even during periods of prolonged infection, and covalently closed circular DNA (cccDNA) genome structure (Kramvis, 2014).

In this study, we report 39 novel Greenland taxa and investigate their phylogenetic relationships and the time to the most recent common ancestor (TMRCA) of HBV/D in Greenland. To understand how and when HBV reached Inuit populations, we analyzed novel sequence data combined with publicly available datasets and ancestral sequences. Here, we investigate the Greenland clade's genetic composition and existing genotype D heterogeneity and present the description of a novel quasi-subgenotype.

2. METHODS

2.1. Data Collection

Thirty-nine serum specimens were collected from 25 individuals infected with HBV from five settlements in Western Greenland: Aasiaat, Itilleq, Nuuk, Sarfannguaq, and Sisimiut (**Figure 1**), as previously described (Kowalec et al., 2013; Bouckaert et al., 2017) (**Supplementary Table 1**). In certain cases, consecutive paired sera separated 5–10 years were collected from individuals for the purpose of investigating the intrapatient genetic diversity of HBV (Kowalec et al., 2013). All specimens were collected on dates ranging from 1998 to 2017 (**Supplementary Table 1**). Viral genomic DNA was extracted from 200 μL sera and amplified as previously described (Bouckaert et al., 2017). Amplicons were sequenced with an AB 3730 XL DNA analyzer using Big Dye 3.1 terminator chemistry (Thermo Fisher Scientific, Burlington, ON, Canada). Sequences were assembled and analyzed using DNA sequence analysis software (Lasergene software suite v 15.0.0, DNASTAR, Madison, WI, USA).

2.2. Taxon Sampling

We created two datasets for the analysis of the Greenland taxa. The first dataset (dataset 1) was 3,567 nucleotides in length and included genomes larger than 2,900 nucleotides for a total of 1,921 sequences from all HBV genotypes. This dataset was created in order to assess the clustering of the Greenland taxa within existing HBV diversity. Dataset 1 was subsampled to contain only complete genomes as indicated in their NCBI record (dataset 1.1), resulting in a dataset with 1,777 sequences and 3,510 nucleotides in length from genotypes A–F. Dataset 1 was further subsampled for a genetic distance analysis (dataset 1.2), where 900 sequences consisting of complete genomes were used (see Methods subsection 2.4 for details).

The second dataset (dataset 2) was 3,182 nucleotides in length and included 93 sequences of genotype D sequences and Greenland sequences in order to assess the diversity of the novel sequences in a more granular way within genotype D, following the evidence from dataset 1 that these novel sequences fall within the genotype D clade. We kept two

FIGURE 1 | Sampling locations of Hepatitis B virus (HBV) isolates within Western Greenland. The map was generated using Tableau Desktop 2009.1 (https://www.tableau.com).

identical sequences which were serially collected from a single Greenland patient with Accession IDs JN792909 sampled in 1998, and JN792910 sampled in 2004. Out of the 39 Greenland sequences, 29 are novel and are published here for the first time (**Supplementary Table 1**). Seventeen sequences cover the full genome, while the remaining twelve are partial or nearly complete, though at least 1,253 nucleotides in length. We selected 54 sequences to represent HBV/D history and calibrate our data (**Supplementary Table 1**). Forty-two sequences serve as global representation of all subgenotypes (D1, $n = 10$; D2, $n = 1$; D3, $n = 5$; D4, $n = 6$; D5, $n = 5$; and D6, $n = 5$). An additional four are "novel quasi-subgenotype D2 of hepatitis B virus" full-length genome sequences from Taiwanese Indigenous peoples (Tran et al., 2014 AB555496, AB555497, AB555500, and AB555501), two are from a New Caledonia (HQ700511) and Argentina (JN688695) outgroup to the Indigenous taxa, and six are partial or nearly complete ancient HBV (aHBV) DNA sequences at least 943 nucleotides in length. Four aHBV sequences came from the Mühlemann et al. (2018) study (LT992438, LT992439, LT992444, and LT992454) and two aHBVs came from Krause-Kyora et al. (2018) study, a German medieval tooth and an Italian mummy available at PRJEB24921 of the European Nucleotide Archive. The ages of these ancient sequences ranged from 281 Before Common Era (BCE) to 1569 Common Era (CE).

The GenBank accession numbers for novel and previously published HBV genomes for dataset 2 are included in **Supplementary Table 1**.

2.3. Molecular Sequence Analyses

We computed multiple sequence alignments using MAFFT (Katoh and Standley, 2013) under default settings. We visualized the alignments and trimmed ragged edges with AliView (Larsson, 2014). We screened our datasets for recombination using RDP4 (Martin et al., 2015) applying six algorithms: Geneconv, Bootscan, MaxChi, Chimera, SiScan, and 3Seq.

The dataset 2 nucleotide alignment was translated to specific open reading frames for the polymerase and preS2 coding regions to determine the presence of signature amino acids associated with HBV/D subgenotypes within the Greenland sequences. We evaluated the amino acid residues for positions 39 (Pre-S2), 100 and 128 (POL Spacer), and 126 (RT) for all Greenland sequences and compared the observed residues to the predominant amino acids of subgenotypes D1 and D2, according to Yousif and Kramvis (2013). In addition, we mapped three distinct nucleotide mutations G1896A, A1762T, and G1764A across all Greenland taxa, as these are important markers associated with disease prognosis.

2.4. Genetic Distance

To calculate the genetic distance between the Greenland clade and other subgenotypes we estimated the evolutionary divergence over full-length genome sequence between groups as implemented in MEGA X (Kumar et al., 2016, 2018). The analyses were conducted using the Maximum Composite Likelihood model (Tamura et al., 2004). This analysis involved 900 nucleotide sequences from subgenotypes D1 (579), D2 (293), and Greenland (28) from dataset 1.2 (see **Supplementary Material**). Codon positions included were 1st + 2nd + 3rd + Non-coding. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There were a total of 3,353 positions in the final dataset.

2.5. Phylogenetic Tree Search

To assign the novel HBV sequences from Greenland to the correct genotype we conducted a phylogenetic tree search on dataset 1 under the optimality criterion of maximum likelihood with substitution model testing as available in IQ-TREE (Nguyen et al., 2015). We executed a 1,000-replicate SH-aLRT measure of support (Guindon et al., 2010). "*The SH-aLRT is an approximation of the likelihood ratio, a direct measure of how much the evidence supports the hypothesis. SH-aLRT is an approximation of the ratio of the log-likelihood of the optimal hypothesis and the best contradictory hypothesis*" (Schneider et al., 2020).

In order to avoid any artifact or bias in the analysis, we tested dataset 1 against its subset dataset 1.1, and no major changes of topology were found. We also tested dataset 2 of the novel Greenland taxa against two subsets of it: (1) only full genomic sequences; (2) only full genomic sequences and unique individuals (oldest sequence from each individual was selected). We built a total of three ML trees using IQ-Tree and compared them. Neither subset tree had a change in the overall tree topology or affected the monophyly of the Greenland clade.

We used BEAST v1.10.4 (Suchard et al., 2018) to calculate the TMRCA of the Greenland taxa on dataset 2 (genotype D). Based on the partial fragmentation of our alignment (see **Supplementary Material**), we elected to treat the alignment as a single substitution partition, contrary to the eight-partition approach used on genotype B by Bouckaert et al. (2017). We selected the GTR+Γ_4 substitution model with four empirical base frequencies, a Skyride coalescent tree prior (Minin et al., 2008), and an uncorrelated relaxed lognormal clock (URL). We also tested a strict clock; however, we did not observe convergence in Tracer (Rambaut et al., 2018), confirming the URL clock as the appropriate selection. We ran the model in duplicate with a Markov chain Monte Carlo length of 100 million. Convergence was determined by Effective Sample Sizes of at least 200 per statistic. The final maximum clade credibility tree with metadata annotation was rendered using FigTree (Rambaut, 2012).

3. RESULTS

3.1. Maximum Likelihood Tree Indicates Greenland Lineage Independence

Model testing prior to IQ-TREE building identified the best fit model from Akaike and Bayesian Information Criterion as GTR + F + R10. The maximum likelihood tree from dataset 1.1 has the Greenland taxa in a monophyletic group sister to genotype D sequences with 89.9% SH-aLRT-like support (**Figure 2**). Thus, we assigned them to genotype D for our downstream analyses. All subsequent references to our data refer to dataset 2 containing only genotype D sequences.

3.2. Recombination

We excluded no taxa based on recombination. RDP identified two potential recombinants (see **Supplementary Material**). The first we interpret to be false identification of the recombinant region from nucleotide position 1720–2579 in 11 D4/D6 sequences containing an 859 base pair section from D2 sequence AB267090. Based on the location of the areas internal and external to the recombination site in UPGMA trees, we see no evidence for inter-subgenotype recombination. An additional sequence, JN792912, which was labeled as mostly recombinant with unknown origin, was removed in addition to the eleven sequences and RDP was re-run. The same region was identified in all D3 sequences instead of D4/D6 with a supposed D1 parent. We hypothesize this is an artifact given the high nucleotide conservation of this region. The UPGMA tree topologies were not suspicious of recombinant template-switching in these taxa.

The second recombinant region flagged by RDP was in an Indian D5 sequence, GQ205382, with a 269 base pair region of a D3 sequence (KX827292) from nucleotides 1765 to 2034. This region has no mutations between the two sequences. Instead, the D3-specific single nucleotide polymorphisms (SNPs) are found in this D5 sequence, potentially representing a recombination event that cannot be excluded. However, the D5 sequence still clusters with other Indian D5 sequences in a clade with 100% SH-aLRT-like support in the ML tree. Neither of these taxa produced interference on ML tree topology. Thus, we elected to keep all of the 93 taxa. We found no recombination between D1, D2, and the Greenland sequences.

3.3. Amino Acid Signatures

The Greenland HBV/D amino acid signatures on Pre-S2 aa39, POL Spacer aa100 and aa128, and RT aa126 introduces a paradigm in relation to D1 and D2 signatures (**Table 1**). The Greenland sequences do not "fit" into either D1 or D2 subgenotypes amino acid signatures predominantly observed with these subgenotype sequences according to Yousif and Kramvis (2013). Instead, they share characteristics of D2 at PreS2 aa39 and POL aa100, and D1 at RT aa126 and POL aa128.

3.4. Mutations

We observed mutations in our sequences at positions 1762 (wild type A; mutant T), 1764 (wild type G; mutant A), and 1896 (wild type G; mutant A). Twenty two of the 39 Greenland sequences had a G to A mutation at nucleotide position 1896. Of the 13 consecutively sampled taxa, only two pairs reflected this substitution, whereas most (n = 5) had a mutant A at position 1896 in both samples. Mutations A1762T and G1764A were not as prevalent, present in only 9/39 and 14/39 Greenland sequences, respectively (**Table 2**).

3.5. Genetic Distance

The Greenland clade had an estimated evolutionary divergence of ∼3% with both subgenotypes D1 and D2. Both D1 and D2 also shown a divergence of ∼3% between each other (**Table 3**). We used the pool of complete genome sequences of the Greenland clade and compared to the pool of D1 and D2 sequences belonging to sister clades on the ML tree.

3.6. Substitution Rate and TMRCA of HBV/D in Greenland Populations

We inferred on the molecular clock analysis a substitution rate of 1.398×10^{-5} [95% HPD 8.837×10^{-6}, 1.952×10^{-5}],

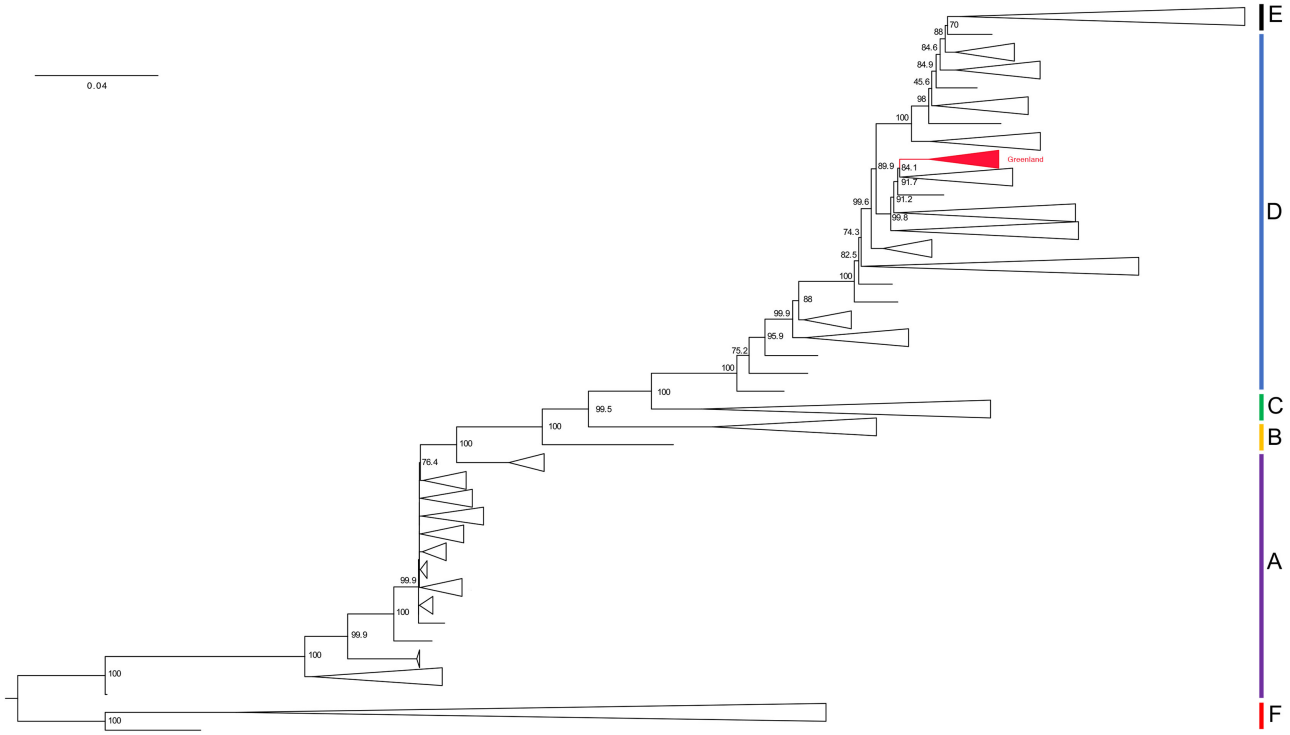

FIGURE 2 | Phylogenetic hypothesis of dataset 1.1. Maximum-likelihood tree. Branch lengths represent an estimation of the average number of nucleotide substitutions per site. Node labels indicate SH-aLRT support. Clade names correlate to their genotype assignment and are represented by the vertical bars on the right of the tree. Red collapsed clade represents the Greenland taxa. See **Supplementary Material** for a full version of this tree.

TABLE 1 | Amino acid signatures of HBV subgenotype D1, D2 and Greenland taxa.

AA Signature	D1	D2	Greenland
Pre-S2 aa39	V/a	A/v	A
POL Spacer aa100	S/a, t	A/t,v	A
POL Spacer aa128	S/g	G/s	S
RT aa126	H	R/h	H

Adapted from Yousif and Kramvis (2013).

similar to that obtained by Mühlemann et al. (2018): 1.18×10^{-5} [9.21×10^{-6}, -1.45×10^{-5}]. The age of the root, representing genotype D, was 675 BCE (1221 BCE, 281 BCE). The age of the Greenland taxa was 629 CE (95% HPD interval 37 CE, 1138 CE).

4. DISCUSSION

Studies on HBV phylogenetics provide a means of understanding the history of the disease spread in an epidemiological context. In this study, we reveal a novel quasi-subgenotype of HBV genotype D isolated from Greenland. Interestingly, this novel clade, sister to subgenotype D2, shares genetic characteristics of both subgenotypes D1 and D2 and shares slightly reduced genetic divergence with subgenotype D1 (3.1%) compared to subgenotype D2 (3.2%) based on our data. Inference of all Greenland taxa as monophyletic in maximum likelihood and Bayesian frameworks supports the hypothesis of a single origin event giving rise to a distinct and isolated lineage on the west coast of Greenland.

Phylogeography in the Greenland dataset was not used given the limited geographic scope of the Greenland taxa and lack of logistical purpose exploring global movement given our reductive representation of genotype D taxa. The 1762 and 1764 site mutations are associated with a significantly increased risk of liver disease progression and hepatocellular carcinoma development (Yang et al., 2016; Lin and Kao, 2017). The mutation on site 1896 is associated with HBeAg negativity and seroconversion to anti-HBeAg positivity by introducing a stop codon to the HBeAg reading frame (Kramvis et al., 2018). Twenty two of the Greenlandic sequences shared a non-exclusive G to A substitution at nucleotide position 1896, associated with seroconversion of positive HBeAg reactivity to anti-HBeAg positivity and historically most present in genotype D (Zhand et al., 2015). These taxa likely had higher rates of evolution (Kramvis et al., 2018), though our methods were unable to tease apart their contribution to the substitution rate. Of our consecutively sampled pairs/triplet, the lack of precore mutation ($n = 3$) or substitution from the mutant to the wild type ($n = 3$) indicates active infection occurring over multiple years, though our lack of sampling only offers sporadic glances into

TABLE 2 | Greenland HBV/D taxa and consecutive paired sera.

Patient ID	Taxon ID(s)	Collection date	Community
01	MT603386$^{\dagger}{}_{nodata}$	2017	Aasiaat
02	MT603402	2004	Itilleq
03	JN792909; JN792910	1998; 2004	Itilleq
04	JN792911†,•,∗; JN792912†,•,∗; MT603404•,∗	1998; 2004; 2008	Itilleq
05	MT603401; MT603403	2004; 2009	Itilleq
06	MT603400; MT603385	2004; 2009	Itilleq
07	MT603383†,•,∗; MT603384†,•,∗	2004; 2009	Itilleq
08	MT603382†,$*_{mixed}$	2017	Nuuk
09	MT603376†	2017	Nuuk
10	MT603378$^{\dagger}{}_{nodata}$	2017	Nuuk
11	MT603396$^{\dagger}{}_{mixed}$	2017	Nuuk
12	JN792904†; JN792903	1998; 2004	Sarfannguaq
13	JN792908†; JN792907	1998; 2004	Sarfannguaq
14	JN792906†; JN792905†	1998; 2004	Sarfannguaq
15	MT603397†	1998	Sisimiut
16	MT603399†	2017	Sisimiut
17	MT603380$^{\dagger}{}_{mixed}$	2017	Sisimiut
18	MT603381†,∗	2017	Sisimiut
19	MT603377†,$\bullet_{mixed}$	2017	Sisimiut
20	MT603379$^{\dagger}{}_{mixed}$	2017	Sisimiut
21	MT603392†,∗; MT603389†,∗	1998; 2008	Sisimiut
22	MT603391†; MT603390†	1998; 2008	Sisimiut
23	MT603388†,•,∗; MT603387†,•,∗	1998; 2008	Sisimiut
24	MT603398•,∗; MT603393†,•,∗	1998; 2008	Sisimiut
25	MT603394∗; MT603395†,∗	1998; 2008	Sisimiut

Paired sera is indicated by multiple taxon IDs in a single cell separated by semicolon. Significant mutant nucleotides (bolded) are indicated by symbol, opposed to unmarked wild type nucleotides. † *= G1896A,* • *= A1762T,* ∗ *= G1764A.*

TABLE 3 | Estimates of evolutionary divergence over sequence pairs between Greenland, D1 and D2 clades.

	Greenland	D1	D2
Greenland	0	0.0309	0.0322
D1	0.0309	0	0.0303
D2	0.0322	0.0303	0

The number of base substitutions per site from averaging over all sequence pairs between groups are shown.

infection dynamics. The identical pair of consecutively sampled sera separated by 6 years is interesting, as we expected this individual to seroconvert over this period. These findings reflect the slow and tangled nature of HBV evolution.

As the Greenland sequences are monophyletic, this suggests the presence of a lineage with a new amino acid signature. For instance, though no recombination was inferred between the Greenland sequences and subgenotype D1/D2, the overlapping serotype signatures (**Table 1**), suggest that the Greenland taxa cannot be neatly fit into the amino acid composition of either subgenotype.

It is a well-known observation that subgenotype D1 shows intergroup divergence with D2 <4–7.5% divergence agreed upon as the definition of an HBV subgenotype (Yousif and Kramvis, 2013; McNaughton et al., 2020). The justification for classifying D1 and D2 as subgenotypes is based upon well-resolved phylogenetic clustering, supported by high bootstrap values (>85–90%), different serological subtypes, and different geographic distribution (Yousif and Kramvis, 2013). As observed in this study, the Greenland HBV clade is well-resolved and clusters completely separately with high confidence by aLRT-SH and posterior probability, and is unique to the geographic region of Greenland. However, in keeping with the suggestion of McNaughton et al. (2020), the Greenland clade may be considered a quasi-subgenotype due to its monophyletic clustering following maximum-likelihood analysis but with a genetic distance separating it from D2 and D1 of ∼3%. Based on tree topology, this quasi-subgenotype clusters most closely with HBV subgenotype D2. The Greenland taxa's inferred sister clade, which includes HBV subgenotype D2 from Taiwanese Indigenous Peoples, New Caledonia, and Argentina, suggests a single origin between the Greenland and D2 taxa that is not inclusive of D1, with 94% posterior support (**Figure 3**).

Taxa are likely to be geographically related, preserving the complication that may be faced with pure geographic assignment (de Bernardi Schneider et al., 2019; Mojsiejczuk et al., 2020). Rather than focusing on endemicity, location, or solely genetic distance, the best way to characterize HBV may be by observing evolutionarily independent lineages

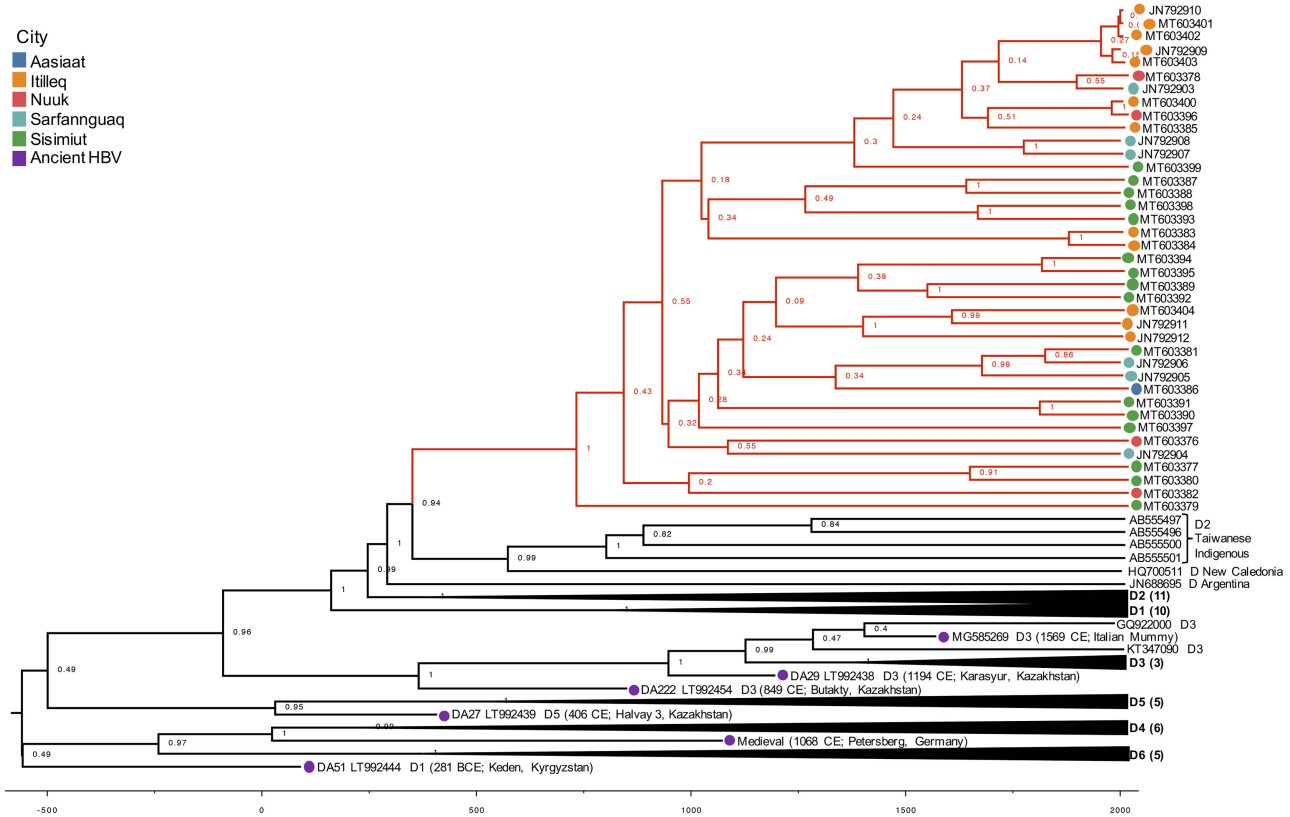

FIGURE 3 | Phylogenetic hypothesis of dataset 2. HBV/D Maximum Clade Credibility (MCC) tree from combined run trees with posterior support indicated on nodes. Greenland lineages are shaded in red and corresponding location indicated by colored circle on branch tips. X-axis represents the age of nodes.

through phylogenetic analysis, as is argued by Peterson (2014). Phylogenetic analysis circumvents the use of genetic distance as a sole metric for sub-/geno-type naming. Further, while genetic distances between the Greenland taxa and subgenotypes D1 and D2 (3.1 and 3.2%, respectively) are within what McNaughton et al. (2020) cites as the majority of genotype D pairwise distance (i.e., 2–4%, compared to the typical subgenotype divergence of 3–8%), they are also distinctly monophyletic across our methods. Here, we present our inference of 39 Greenland HBV/D taxa as a newly characterized quasi-subgenotype exemplified by this topologically independent lineage.

The Greenland taxa's TMRCA of 629 CE translates to a coalescence point about 1,390 years before present (YBP), preceding the only other Greenland HBV lineage (B5) studied (Bouckaert et al., 2017). Their study estimated that B5 was introduced to Greenland during coastal route movement of the Thule in the last 1,000 years (Bouckaert et al., 2017), calibrated by the date associated with Thule migration to the Eastern Arctic (647–953 YBP or 1370–1064 CE). Nonetheless, they are distinct genotypes evolving at different rates, thus it is feasible that these genotypes arrived in Western Greenland following separate introductions.

The more closely related subgenotype D2 may have an origin from the Middle East (Kostaki et al., 2018) or Southern Europe (Spitz et al., 2019). The use of ancient HBV DNA allowed us to estimate a Greenland HBV/D TMRCA range of 37–1138 CE. Similar to what has been described with HBV genotype B-infected indigenous people in the circumpolar Arctic (Bouckaert et al., 2017), this result suggests that the virus was introduced at some point to the Thule, but this ancestor no longer exists. However, ancestral remnants aside from the Taiwanese Indigenous and New Caledonian samples which formed an adjacent clade to the Greenland taxa are necessary for improved confidence in inferring the route which brought HBV/D to these populations.

Integrating novel sequences with published HBV/D data in this study has demonstrated a strongly inferred and geographically independent monophyletic lineage from existing HBV/D subgenotype architecture. We reveal that HBV/D samples from 25 Western Greenland residents, 13 of whom were consecutively sampled over 5–10 years, form an evolutionarily unique clade distinct from subgenotypes D1 and D2.

Pairing these modern data with ancient DNA, we uncovered more detail in the story of HBV genotype D evolution. Such diachronic studies are necessary for the tough-to-interpret HBV (Mühlemann et al., 2018), where study results on rate and origin vary by orders of magnitude subjective to data and model limitations. Additional HBV/D sequences from ancient archeological remains in the Arctic are necessary to resolve the mystery of its origin and pattern of dispersal beyond speculation.

DATA AVAILABILITY STATEMENT

The taxa utilized in this study can be found in **Supplementary Table 1**. All supplementary data, which includes all trees from this study and metadata, can be found at GitHub (https://github.com/abschneider/Paper_HBV_Greenland).

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Statens Serum Institute (Copenhagen, Denmark) and the Commission for Scientific Research in Greenland (Approval number 505-99). Written informed consent was obtained from each participant. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

HK and MB contributed to the acquisition of study samples. AB, CO, and RH contributed substantially to the conception and design of the work and the drafting and critical revision of the manuscript. AB, CO, and JW contributed to the revised version of the manuscript. AB, RH, HK, MB, YT, TM, CO, and JW contributed to the analysis and interpretation of data and manuscript revision. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Institutes of Health (NIH) National Institute of Allergy and Infectious Diseases (grant numbers K01AI110181 and AI135992) to JW.

ACKNOWLEDGMENTS

The authors acknowledge and were grateful for the technical assistance of Ms. Elizabeth Giles. This manuscript has been released as a pre-print at medRxiv (de Bernardi Schneider et al., 2020).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.602296/full#supplementary-material

REFERENCES

Araujo, N. M. (2015). Hepatitis B virus intergenotypic recombinants worldwide: an overview. *Infect. Genet. Evol.* 36, 500–510. doi: 10.1016/j.meegid.2015.08.024

Bjorn-Mortensen, K., Ladefoged, K., Obel, N., and Helleberg, M. (2013). The HIV epidemic in greenland-a slow spreading infection among adult heterosexual greenlanders. *Int. J. Circ. Health* 72:19558. doi: 10.3402/ijch.v72i0.19558

Børresen, M. L., Koch, A., Biggar, R. J., Andersson, M., Wohlfahrt, J., Ladefoged, K., et al. (2011). Hepatocellular carcinoma and other liver disease among greenlanders chronically infected with hepatitis B virus: a population-based study. *J. Natl. Cancer Inst.* 103, 1676–1685. doi: 10.1093/jnci/djr405

Bouckaert, R., Simons, B. C., Krarup, H., Friesen, T. M., and Osiowy, C. (2017). Tracing hepatitis B virus (HBV) genotype B5 (formerly B6) evolutionary history in the circumpolar arctic through phylogeographic modelling. *PeerJ* 5:e3757. doi: 10.7717/peerj.3757

de Bernardi Schneider, A., Hostager, R., Krarup, H., Borresen, M., Tanaka, Y., Morriseau, T., et al. (2020). Analysis of hepatitis B virus genotype d in greenland suggests presence of a novel subgenotype. *medRxiv*. doi: 10.1101/2020.06.06.20123968

de Bernardi Schneider, A., Ochsenreiter, R., Hostager, R., Hofacker, I. L., Janies, D., and Wolfinger, M. T. (2019). Updated phylogeny of chikungunya virus suggests lineage-specific RNA architecture. *Viruses* 11:798. doi: 10.1101/698522

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of phyml 3.0. *Syst. Biol.* 59, 307–321. doi: 10.1093/sysbio/syq010

Katoh, K., and Standley, D. M. (2013). Mafft multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010

Kostaki, E.-G., Karamitros, T., Stefanou, G., Mamais, I., Angelis, K., Hatzakis, A., et al. (2018). Unravelling the history of hepatitis B virus genotypes A and D infection using a full-genome phylogenetic and phylogeographic approach. *Elife* 7:e36709. doi: 10.7554/eLife.36709

Kowalec, K., Minuk, G., Børresen, M., Koch, A., McMahon, B., Simons, B., et al. (2013). Genetic diversity of hepatitis B virus genotypes B6, D and F among circumpolar indigenous individuals. *J. Viral Hepat.* 20, 122–130. doi: 10.1111/j.1365-2893.2012.01632.x

Kramvis, A. (2014). Genotypes and genetic variability of hepatitis B virus. *Intervirology* 57, 141–150. doi: 10.1159/000360947

Kramvis, A., Kostaki, E.-G., Hatzakis, A., and Paraskevis, D. (2018). Immunomodulatory function of hbeag related to short-sighted evolution, transmissibility, and clinical manifestation of hepatitis B virus. *Front. Microbiol.* 9:2521. doi: 10.3389/fmicb.2018.02521

Krause-Kyora, B., Susat, J., Key, F. M., Kühnert, D., Bosse, E., Immel, A., et al. (2018). Neolithic and medieval virus genomes reveal complex evolution of hepatitis B. *Elife* 7:e36666. doi: 10.7554/eLife.36666

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). Mega X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096

Kumar, S., Stecher, G., and Tamura, K. (2016). Mega7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054

Langer, B., Frösner, G., and Von Brunn, A. (1997). Epidemiological study of viral hepatitis types A, B, C, D and E among inuits in West Greenland. *J. Viral Hepat.* 4, 339–349. doi: 10.1046/j.1365-2893.1997.00063.x

Larsson, A. (2014). Aliview: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* 30, 3276–3278. doi: 10.1093/bioinformatics/btu531

Lavanchy, D. (2008). Chronic viral hepatitis as a public health issue in the world. *Best Pract. Res. Clin. Gastroenterol.* 22, 991–1008. doi: 10.1016/j.bpg.2008.11.002

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

Lin, C.-L., and Kao, J.-H. (2017). Natural history of acute and chronic hepatitis B: the role of hbv genotypes and mutants. *Best Pract. Res. Clin. Gastroenterol.* 31, 249–255. doi: 10.1016/j.bpg.2017.04.010

Littlejohn, M., Locarnini, S., and Yuen, L. (2016). Origins and evolution of hepatitis B virus and hepatitis D virus. *Cold Spring Harbor Perspect. Med.* 6:a021360. doi: 10.1101/cshperspect.a021360

Livingston, S. E., Simonetti, J. P., McMahon, B. J., Bulkow, L. R., Hurlburt, K. J., Homan, C. E., et al. (2007). Hepatitis B virus genotypes in Alaska native people with hepatocellular carcinoma: preponderance of genotype F. *J. Infect. Dis.* 195, 5–11. doi: 10.1086/509894

Locarnini, S., Littlejohn, M., Aziz, M. N., and Yuen, L. (2013). Possible origins and evolution of the hepatitis B virus (HBV). *Semin. Cancer Biol.* 23, 561–575. doi: 10.1016/j.semcancer.2013.08.006

Martin, D. P., Murrell, B., Golden, M., Khoosal, A., and Muhire, B. (2015). RDP4: detection and analysis of recombination patterns in virus genomes. *Virus Evol.* 1:vev003. doi: 10.1093/ve/vev003

McMahon, B. J. (2008). Natural history of chronic hepatitis B-clinical implications. *Medsc. J. Med.* 10:91.

McNaughton, A. L., Revill, P. A., Littlejohn, M., Matthews, P. C., and Ansari, M. A. (2020). Analysis of genomic-length HBV sequences to determine genotype and subgenotype reference sequences. *J. Gen. Virol.* 101, 271–283. doi: 10.1099/jgv.0.001387

Minin, V. N., Bloomquist, E. W., and Suchard, M. A. (2008). Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. *Mol. Biol. Evol.* 25, 1459–1471. doi: 10.1093/molbev/msn090

Minuk, G. Y., and Uhanova, J. (2003). Viral hepatitis in the Canadian Inuit and first nations populations. *Can. J. Gastroenterol. Hepatol.* 17, 707–712. doi: 10.1155/2003/350175

Mojsiejczuk, L., Torres, C., Flichman, D., and Campos, R. H. (2020). Long-term evolution of hepatitis B virus genotype F: strong association between viral diversification and the prehistoric settlement of central and South America. *J. Viral Hepat.* 27, 620–630. doi: 10.1111/jvh.13273

Mühlemann, B., Jones, T. C., de Barros Damgaard, P., Allentoft, M. E., Shevnina, I., Logvin, A., et al. (2018). Ancient hepatitis B viruses from the bronze age to the medieval period. *Nature* 557:418. doi: 10.1038/s41586-018-0097-z

Nguyen, L.-T., Schmidt, H. A., Von Haeseler, A., and Minh, B. Q. (2015). IQ-tree: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300

Okamoto, H., Tsuda, F., Sakugawa, H., Sastrosoewignjo, R. I., Imai, M., Miyakawa, Y., et al. (1988). Typing hepatitis B virus by homology in nucleotide sequence: comparison of surface antigen subtypes. *J. Gen. Virol.* 69, 2575–2583. doi: 10.1099/0022-1317-69-10-2575

Osiowy, C., Larke, B., and Giles, E. (2011). Distinct geographical and demographic distribution of hepatitis B virus genotypes in the Canadian arctic as revealed through an extensive molecular epidemiological survey. *J. Viral Hepat.* 18, e11–e19. doi: 10.1111/j.1365-2893.2010.01356.x

Osiowy, C., Simons, B. C., and Rempel, J. D. (2013). Distribution of viral hepatitis in indigenous populations of north america and the circumpolar arctic. *Antivir. Ther.* 18, 467–473. doi: 10.3851/IMP2597

Peterson, A. T. (2014). Defining viral species: making taxonomy useful. *Virol. J.* 11:131. doi: 10.1186/1743-422X-11-131

Pourkarim, M. R., Amini-Bavil-Olyaee, S., Lemey, P., Maes, P., and Van Ranst, M. (2010). Are hepatitis B virus "subgenotypes" defined accurately? *J. Clin. Virol.* 47, 356–360. doi: 10.1016/j.jcv.2010.01.015

Rambaut, A. (2012). *FigTree v1. 4. Molecular Evolution, Phylogenetics and Epidemiology*. Edinburgh. Available online at: http://tree.bio.ed.ac.uk/software/figtree

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., and Suchard, M. A. (2018). Posterior summarization in Bayesian phylogenetics using tracer 1.7. *Syst. Biol.* 67, 901–904. doi: 10.1093/sysbio/syy032

Rex, K. F., Andersen, S., and Krarup, H. B. (2015). Hepatitis B among inuit: a review with focus on Greenland Inuit. *World J. Hepatol.* 7:1265. doi: 10.4254/wjh.v7.i9.1265

Schneider, A. d. B., Jacob Machado, D., Guirales, S., and Janies, D. A. (2020). FLAVi: an enhanced annotator for viral genomes of *Flaviviridae*. *Viruses* 12:892. doi: 10.3390/v12080892

Spitz, N., Mello, F. C., Moreira, A. S., Gusatti, C. S., Martins, R. M., Gomes, S. A., et al. (2019). Reconstruction of the spatial and temporal dynamics of hepatitis B virus genotype D in the americas. *PLoS ONE* 14:e0220342. doi: 10.1371/journal.pone.0220342

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

Tamura, K., Nei, M., and Kumar, S. (2004). Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc. Natl. Acad. Sci. U.S.A.* 101, 11030–11035. doi: 10.1073/pnas.0404206101

Tran, H., Yu, M.-L., Dai, C.-Y., Lin, I.-L., Yeh, M.-L., Chuang, W.-L., et al. (2014). Novel quasi-subgenotype D2 of hepatitis B virus identified in Taiwanese aborigines. *Virus Genes* 49, 30–37. doi: 10.1007/s11262-014-1072-x

Tulisov, A., McMahon, B. J., Koch, A., Minuk, G., Chulanov, V., Bruce, M. G., et al. (2007). Viral hepatitis in the arctic. A review from a circumpolar workshop on viral hepatitis, ICCH 13. *Alaska Med.* 49:193.

Vrancken, B., Rambaut, A., Suchard, M. A., Drummond, A., Baele, G., Derdelinckx, I., et al. (2014). The genealogical population dynamics of HIV-1 in a large transmission chain: bridging within and among host evolutionary rates. *PLoS Comput. Biol.* 10:e1003505. doi: 10.1371/journal.pcbi.1003505

WHO (2019). *Hepatitis B*. World Health Organization. Available online at: https://www.who.int/en/news-room/fact-sheets/detail/hepatitis-b

Yang, Z., Zhuang, L., Lu, Y., Xu, Q., Tang, B., and Chen, X. (2016). Naturally occurring basal core promoter A1762T/G1764A dual mutations increase the risk of HBV-related hepatocellular carcinoma: a meta-analysis. *Oncotarget* 7:12525. doi: 10.18632/oncotarget.7123

Yin, Y., He, K., Wu, B., Xu, M., Du, L., Liu, W., et al. (2019). A systematic genotype and subgenotype re-ranking of hepatitis B virus under a novel classification standard. *Heliyon* 5:e02556. doi: 10.1016/j.heliyon.2019.e02556

Yousif, M., and Kramvis, A. (2013). Genotype D of hepatitis B virus and its subgenotypes: an update. *Hepatol. Res.* 43, 355–364. doi: 10.1111/j.1872-034X.2012.01090.x

Zhand, S., Karami, C., Adli, A. H., Tabarraei, A., Khodabakhshi, B., and Moradi, A. (2015). Correlation between hepatitis B G1896A precore mutations and HBEAG in chronic HBV patients. *Jundishapur J. Microbiol.* 8:e17126. doi: 10.5812/jjm.17126

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

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 28 January 2021
doi: 10.3389/fmicb.2020.634430

Classification of the Zoonotic Hepatitis E Virus Genotype 3 Into Distinct Subgenotypes

Florence Nicot[1], Chloé Dimeglio[1,2], Marion Migueres[1,2], Nicolas Jeanne[1], Justine Latour[1], Florence Abravanel[1,2,3], Noémie Ranger[1], Agnès Harter[1], Martine Dubois[1], Sonia Lameiras[4], Sylvain Baulande[4], Sabine Chapuy-Regaud[1,2,3], Nassim Kamar[2,3,5], Sébastien Lhomme[1,2,3] and Jacques Izopet[1,2,3]**

[1] CHU de Toulouse, Hôpital Purpan, Laboratoire de Virologie, Centre National de Référence du Virus de l'Hépatite E, Toulouse, France, [2] INSERM, U1043, Toulouse, France, [3] Department of Virology, Université Toulouse III Paul-Sabatier, Toulouse, France, [4] Institut Curie Genomics of Excellence Platform, Institut Curie Research Center, Paris, France, [5] CHU de Toulouse, Hôpital Rangueil, Service de Néphrologie, Dialyse et Transplantation d'Organe, Toulouse, France

OPEN ACCESS

Edited by:
Lilly Yuen, Victorian Infectious Diseases Reference Laboratory, Australia

Reviewed by:
Antonio Rivero-Juarez, Maimonides Biomedical Research Institute of Cordoba (IMIBIC), Spain
Anna Rosa Garbuglia, Istituto Nazionale per le Malattie Infettive Lazzaro Spallanzani (IRCCS), Italy

***Correspondence:**
Florence Nicot
nicot.f@chu-toulouse.fr
Jacques Izopet
izopet.j@chu-toulouse.fr

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *27 November 2020*
Accepted: *30 December 2020*
Published: *28 January 2021*

Citation:
Nicot F, Dimeglio C, Migueres M, Jeanne N, Latour J, Abravanel F, Ranger N, Harter A, Dubois M, Lameiras S, Baulande S, Chapuy-Regaud S, Kamar N, Lhomme S and Izopet J (2021) Classification of the Zoonotic Hepatitis E Virus Genotype 3 Into Distinct Subgenotypes. Front. Microbiol. 11:634430. doi: 10.3389/fmicb.2020.634430

Hepatitis E virus (HEV) genotype 3 is the most common genotype linked to HEV infections in Europe and America. Three major clades (HEV-3.1, HEV-3.2, and HEV-3.3) have been identified but the overlaps between intra-subtype and inter-subtype p-distances make subtype classification inconsistent. Reference sequences have been proposed to facilitate communication between researchers and new putative subtypes have been identified recently. We have used the full or near full-length HEV-3 genome sequences available in the Genbank database (April 2020; n = 503) and distance analyses of clades HEV-3.1 and HEV-3.2 to determine a p-distance cut-off (0.093 nt substitutions/site) in order to define subtypes. This could help to harmonize HEV-3 genotyping, facilitate molecular epidemiology studies and investigations of the biological and clinical differences between HEV-3 subtypes.

Keywords: HEV-3, subtype, classification, diversity, full-length genome, phylogeny

INTRODUCTION

The hepatitis E virus (HEV) is a significant human pathogen causing viral hepatitis worldwide. Most of the strains that infect humans belong to two species, *Orthohepevirus A* (8 genotypes; HEV 1–8) and *Orthohepevirus C* (Purdy et al., 2017; Smith and Simmonds, 2018; Sridhar et al., 2018; Primadharsini et al., 2019). The most prevalent genotype in industrialized countries at least in Europe and America is HEV genotype 3 (HEV-3). It is transmitted zoonotically by direct contact with infected animals, eating contaminated food, or via the environment. HEV-3 infection is frequently asymptomatic but it can result in severe acute hepatitis in patients with chronic liver disease and lead to chronic hepatitis and cirrhosis in immunocompromised patients (Kamar et al., 2017). Extra-hepatic manifestations have been also described in patients with acute and chronic hepatitis E (Kamar et al., 2017).

Hepatitis E virus genotype 3 variants have been assigned to one of several subtypes based on analysis of a limited number of complete genome sequences and subgenomic regions (Lu et al., 2006). Despite the increasing number of full-length or near full-length genomes deposited in the NCBI database, it is difficult to provide consistent criteria that identify viruses that are members of the same subtype due to overlaps between the intra-subtype and inter-subtype p-distances

commonly used for classification (Smith et al., 2015, 2016). Nevertheless, HEV-3 viruses can be classified into three major clades based on phylogenetic grouping. Clade 3.1 includes HEV-3 subtypes a, b, c, h, i, and j; clade 3.2 includes HEV-3 subtypes e, f, and g, and clade 3.3 contains rabbit strains corresponding to the HEV-3ra subtype (Oliveira-Filho et al., 2013; Ijaz et al., 2014; Vina-Rodriguez et al., 2015). A standard reference set of genome sequences including 17 that are full-length or near full-length HEV-3 genomes was proposed in 2016 using a conservative pragmatic approach (Smith et al., 2016). Subsequently, new potential subtypes have been proposed: 3k (Miura et al., 2017), 3l (De Sabato et al., 2018), 3chi-new (Lhomme et al., 2019), and 3s (Wist et al., 2018; Sahli et al., 2019). The standard reference set of genome sequence was recently updated identifying 3k, 3l, and 3m (previously named 3chi-new) as new subtypes (Smith et al., 2020). Recently, an automated partition of phylogenetic trees constructed from 250 full-length HEV-3 genome sequences has been used to classify more than 99% of the complete genome sequences into subtypes (Nicot et al., 2018).

This study was done to determine a distance cut-off that can be used to assign HEV-3 sequences to a subtype using the new full-length HEV-3 sequences available in NCBI and the recently defined new subtypes.

MATERIALS AND METHODS

Sequencing of HEV Complete Genome Sequence

Stored plasma samples from HEV-infected patients consecutively tested for HEV RNA between 2017 and 2019 in the laboratory of Virology at Toulouse University Hospital, National Reference Center for HEV, with viral load of HEV-3 >100,000 copies/mL were selected for PacBio single molecule real-time sequencing. HEV-RNA extraction and F1 and F2 amplifications were realized for 188 samples as previously described (Nicot et al., 2018). SMRT bell library was constructed by pooling 96 barcoded samples according the manufacturer instructions for SMRTbell Barcoded Adapter Prep kit. Sequencing was performed by using chemistry v3.0 on a PacBio Sequel sequencer available at ICGex, Institut Curie Research Center, Paris, France. Bioinformatics analysis and complete genome reconstruction were realized with an in-house developed pipeline. From demultiplexed.bam files provided by ICGex, CCS were constructed using min passes = 3 and min RQ 0.999 parameters. Reads were mapped to a reference sequence (Minimap 2 2.17) to retain HEV reads and remove chimeric reads. Non-identical reads were subsequently combined using a medoïd-based clustering (cluster-fast from USEARCH 11.0.667) into clusters at 99% genetic identity. A consensus sequence was generated for each cluster. F1 and F2 sequences were assembled with Megamerger (EMBOSS 6.6.0). The consensus sequence with the higher number of reads was used as complete genome sequence, annotated as previously described (Nicot et al., 2018) and submitted to Genbank with accession numbers MW355217–MW355404.

Nucleotide Sequences and Phylogenetic Analysis

The 188 complete genome sequences obtained by SMRT sequencing and all full or near full-length genome sequences of genotype 3 (n = 315) of human or animal origin available in the Genbank database on April 2020 were included (**Supplementary Table 1**). Duplicate sequences from a single individual and six recombinant sequences (D11092, MG783571, KJ013414, KJ013415, KT633715, and DQ450072) were removed. We also included 29 complete genome sequences of genotypes 1, 2, 4, 5, 6, 7, and 8 (Smith et al., 2020) (accession numbers HEV-1: FJ457024, MH918640; HEV-1a: M73218; HEV-1b: L088816; HEV-1c: X98292; HEV-1d: AY230202; HEV-1e: AY204877; HEV-1f: JF443721; HEV-1g: LC225387; HEV-2a: KX578717; HEV-2b: MH809516; HEV-4: MK410048, AB369688; HEV-4a: AB197673; HEV-4b: DQ279091; HEV-4c: AB074915; HEV-4d: AJ272108; HEV–4e: AY723745; HEV-4f: AB220974; HEV-4g: AB108537; HEV-4h: GU119961; HEV-4i: AB369690; HEV-5a: AB573435; HEV-6: AB856243; HEV-6a: AB602441; HEV-7: KJ496144; HEV-7a: KJ496143; and HEV-8: MH410174; HEV-8a: KX387865). The 532 sequences (503 HEV-3 and 29 HEV non-3 genotype) were aligned with MUSCLE v.3.8.31 and a bootstrapped tree (100 replicates) using the maximum likelihood (ML) method with the general time reversible model (GTR + I + G) was constructed on phyML v3.3. Interactive Tree Of Life (iTOL) v3 was used to visualize the whole large tree.

Automated Partition of Phylogenetic Tree

The ML phylogenetic tree was partitioned and strain clusters within genotype 3 were identified using a method adapted from Prosperi et al. (2011). Briefly, the topology of the ML tree was analyzed with a depth first search by considering the number of subtrees with a node reliability $\geq$70% and an associated number of leaves with at least two distinct patients. A subtree was identified as a cluster if the median value of the subtree distance distribution was below a t-percentile threshold of the whole-tree distance distribution. If a node satisfied this condition, the search was stopped at that node, children nodes were ignored, and other sibling nodes were analyzed. The threshold t was evaluated over the (5th, 50th) percentile range of the whole tree distance distribution with steps of 1 between the 5th and 15th and 5 between the 15th and 50th percentiles.

Distance Cut-Off for Identification of HEV Genotype 3 Subtypes

The clusters of sequences identified by automated partition of ML phylogenetic tree were used to analyze the intra- and inter-subtype nucleotide pairwise distances. The distances were estimated on MEGA X using the maximum composite likelihood (MCL) method and a gamma distribution to model evolutionary rate differences among sites (four categories). All distances were analyzed (to determine a cut-off that identified subtypes) by generating boxplots in Matlab R2018B software. Based on the algorithm described in **Supplementary Figure 1**, a sequence X_i can be considered to be a new subtype if all the intra-subtype and inter-subtype distances $(S_{i,j})_{j\in[1,N]}$ are above a cut-off α. If at least

one distance $d(X_i, X_{Vj})$, $j \in [1, N]$ between the new sequence X_i and one of the known sequences X_{Vj} is below the cut-off α, the sequence X_i is assigned to the subtype containing the sequence with the shorter distance.

Statistical Analysis

Continuous variables were tested with Student's *t*-test on STATA 14.0 software. *p*-Values of <0.05 were considered to be significant.

RESULTS

Identification of Clusters

We identified clusters among genotype 3 sequences using automated partition of a ML phylogenetic tree. Clusters of known subtypes were identified with a threshold of 13% (median distance <0.126 nt substitutions/site). This threshold assigns 99.1% of the 483 sequences belonging to clade 3.1 and 3.2 to one of the following known subtypes: 3aj (n = 26), 3b (n = 29), 3c (n = 117), 3h (n = 17), 3k (n = 4), 3i (n = 6), 3l (n = 6), 3m (n = 6) for clade 3.1 and 3e (n = 40), 3f (n = 228), 3g (n = 1) for clade 3.2 (**Figure 1**). The complete genome of HEV-3j was included in the HEV-3a subtype. The sequences proposed as HEV-3s were classified as HEV-3h subtypes. Only four sequences belonging to the group HEV-3chi (MK390370, MK390371, LC260517, and MF959765) were not classified (**Figure 1**). The 13% threshold also defined two clusters within clade 3.3 (HEV-3ra) composed of six (HEV-3ra1) and six sequences (HEV-3ra2) while seven sequences were outside the HEV-3ra1 and HEV-3ra2 clusters (**Figure 1**).

Determination of a Distance Cut-Off for Subtype Discrimination

The pairwise distances were estimated on MEGA X using the MCL method. Means for intra-subtype and inter-subtype distances, standard deviation, 95% confidence interval (CI) and 99% CI for each subtype are shown in **Table 1**. The mean intra-subtype distances obtained for HEV-3ra strains from clade 3.3 (0.116 ± 0.002 nt substitutions/site) was significantly greater than each of the mean intra-subtype distances obtained for the other HEV-3 subtypes ($p < 0.01$ for each subtype). Therefore, the HEV-3ra sequences are too heterogeneous to be used to determine the distance cut-off discriminating between subtypes. Analysis of sequences from clades 3.1 and 3.2 indicated that the overall mean intra-subtype distances was 0.064 (95% CI: 0.03–0.09) whereas the overall mean inter-subtype distances was 0.142 (95% CI: 0.106–0.181). The 95% CI upper limit of intra-subtype distance for each subtype was lower than the 95% CI lower limit of the inter-subtype distance (**Table 1**). Similarly, the 99% CI upper limit of intra-subtype distance was lower than the 99% CI lower limit of inter-subtype distance, except for subtypes 3aj and 3b. Analysis of the intra- and inter-subtype distances for each subtype indicated that 0.093 can be used as a distance cut-off for assigning a sequence to a subtype (**Figure 2**). We therefore designed an algorithm based on this cut-off distance that would assign sequences to a subtype (**Supplementary Figure 1**) and used it to assign 99.1% of the HEV-3 sequences to a defined subtype (3aj, 3b, 3c, 3e, 3f, 3h, 3i, 3k, 3l, 3m). Only four sequences were not assigned (MK390370, MK390371, LC260517, and MF959765), in agreement with the data from automated partition of the ML phylogenetic tree. The distance between sequences MK390370 and MK390371, 0.006 nt substitutions/site, assigned these sequences to the same cluster. LC260517 and MF959765 were isolated sequences (minimum intersubtype-distance: 0.099 nt substitutions/site for both sequences).

Classification of HEV-3 Sequences

We used 11 subtypes to classifying sequences in clades 3.1 (3aj, 3b, 3c, 3h, 3i, 3k, 3l, 3m) and 3.2 (3e, 3f, 3g), based on the criteria proposed by Smith et al. (2020) for subtype assignment and the results of the automated partition and distance cut-off methods (**Table 2**). Each complete genome sequence used in our study was assigned to a subtype (**Supplementary Table 1**). New complete genome sequences can be assigned to an existing subtype provided there is at least one distance less than 0.093. Otherwise, the new sequence may be a new subtype, which then needs to be confirmed (Smith et al., 2016). Complete genome sequences of subtype 3a, 3b, and 3ra were detected worldwide (Asia, Europe, and America), subtypes 3c, 3e, 3f, and 3h were detected in Asia and Europe, subtypes 3i, 3l, and 3m were detected only in Europe and subtype 3k only in Japan (**Table 2** and **Supplementary Table 1**).

DISCUSSION

Hepatitis E virus genotype 3 viruses display considerable diversity and have been classified into subtypes with no clear criteria based on distance and phylogenetic methods for demarcation. The set of reference sequences proposed by Smith et al. (2016) has enabled common subtypes to be assigned, but more than 10% of HEV-3 strains were not classified (Nicot et al., 2018). Automated partition of a ML phylogenetic tree using 503 HEV-3 sequences and distance analysis confirmed the classification of 250 HEV-3 sequences (Nicot et al., 2018) and supported the existence of several new post-2016 subtypes included in update classification (Smith et al., 2020). Subtype assignment using our new analysis method is automated and allow the classification of sequences not classified by Smith et al. (2020). It is important to classify a majority of sequences within a subtype and to have an objective method of classification.

The putative subtypes 3k (Miura et al., 2017), 3l (De Sabato et al., 2018), 3m (Nicot et al., 2018; Lhomme et al., 2019), and 3s (Wist et al., 2018; Sahli et al., 2019) have been described. Evidence for subtypes 3k, 3l, and 3m was provided by the automated partition of phylogenetic trees and the distance cut-off of 0.093. HEV-3k has been found in humans and pigs in Japan (Miura et al., 2017) while the HEV-3l subtype was first described in pigs in Northern Italy (De Sabato et al., 2018). Our analyses indicate that HEV-3l subtype also occurs in humans in France (sequences MF444121, MF444131, and HESQL113). The first strain of subtype 3m was detected in a Spanish patient in 2011 (Munoz-Chimeno et al., 2016) and has

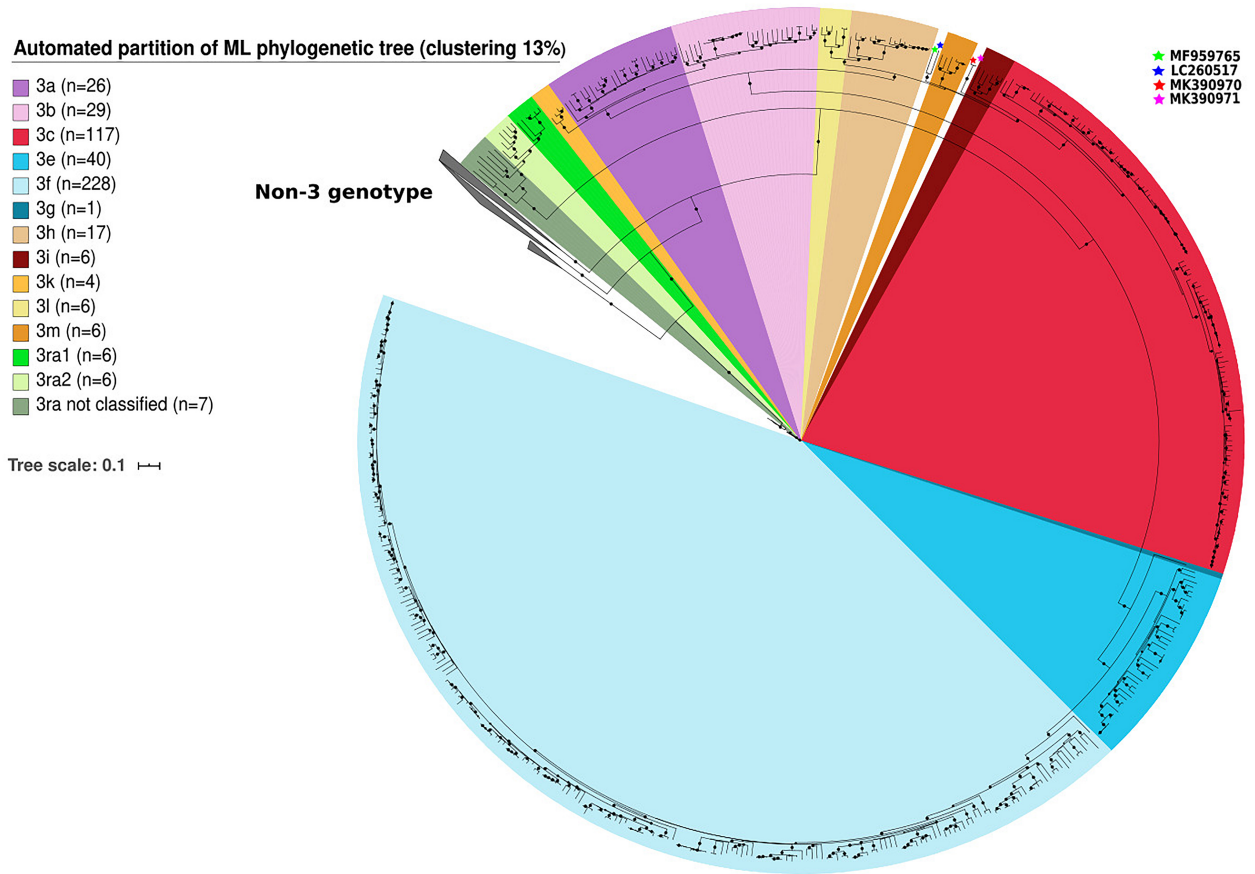

FIGURE 1 | Automated partition of genotype 3 (*n* = 503) clusters within full-length or near full-length genomes. HEV-3 sequences not assigned to a cluster in clades HEV-3.1 are indicated with stars.

since been detected in France, Belgium, Netherlands and the United Kingdom (Ijaz et al., 2014; Nicot et al., 2018; Lhomme et al., 2019). A recent study showed that this subtype circulates in wild boar in Spain and also in human in Sweden, suggesting that it is transmitted via consumption of contaminated meat or water or direct contact with wild boar (Wang et al., 2019). In contrast, all the putative subtype 3s sequences were assigned to subtype 3h by the automated partition of ML phylogenetic tree, and distance analysis indicated that 3s sequences should be assigned to subtype 3h, all with distances below the cut-off of 0.093. In addition, these strains, which have been found in both humans and animals in Switzerland, form a cluster that is transmitted by the consumption of locally produced pork meat (Sahli et al., 2019). They cannot be assigned to a new subtype because they are epidemiologically related (Smith et al., 2016). Our analyses also indicate that HEV-3i, described in boar in Germany (Adlhoch et al., 2009) or human in Sweden (Norder et al., 2018), occurs in human in France (HESQL053 and HESQL059).

The sequence AY115488 classified 3j with Smith criteria (Smith et al., 2020) was obtained from the feces of pigs housed in Canada (Pei and Yoo, 2002). This sequence was classified among subtype 3a with our analysis. Indeed, the minimum intra-subtype distance observed between AY115488 and AB089824 (0.083 nucleotide subtitutions/site) in subtype 3aj is much lower than the cut-off value of 0.093 for assigning a sequence to a different subtype. The four sequences (MK390370, MK390371, LC260517, and MF95765) were not classified according Smith et al. (2020) criteria and our analysis. They could be consider as three potential new subtypes, considering MK390370 and MK390371 are assigned to the same cluster. However, in the absence of at least three complete genome sequences epidemiologically unrelated (Smith et al., 2016), these new subtypes could not be confirmed.

Hepatitis E virus genotype 3 is found worldwide and is the predominant genotype in Europe and America. The majority of Asian and North American strains of HEV-3 belong to subtypes 3a and 3b (Zehender et al., 2014). Subtype 3b is indigenous to Japan, although 3b strains have occasionally been identified in Europe (Legrand-Abravanel et al., 2009; Vina-Rodriguez et al., 2015). Subtypes 3k strains have been described only in Asia (Miura et al., 2017). The majority of European strains belong to subtypes 3c, 3f, and 3e. Changes in the distribution of variants within genotypes have been highlighted. A switch from clade 3.2 (mainly 3f and 3e) to clade 3.1 (mainly 3c) infections was observed in France and the United Kingdom after 2010 (Nicot et al., 2018; Oeser et al., 2019) and a similar switch occurred more recently in Belgium; the subtype 3f strains that were predominant

TABLE 1 | Nucleotide p-distances calculated for each subtype (intra and inter-subtype distances).

Subtype	*N*	Distances	Observations	p-distances			
				Mean	SD	95% CI	99% CI
3aj	26	Intra	325	0.067	< 0.001	0.066 – 0.069	0.01 – 0.096
		Inter	11,804	0.134	< 0.001	0.134 – 0.135	0.089 – 0.157
3b	29	Intra	406	0.068	< 0.001	0.066 – 0.07	0.003 – 0.089
		Inter	13,079	0.134	< 0.001	0.134 – 0.135	0.089 – 0.157
3c	117	Intra	6786	0.055	< 0.001	0.055 – 0.056	0.004 – 0.085
		Inter	42,471	0.14	< 0.001	0.139 – 0.14	0.1 – 0.157
3e	40	Intra	780	0.076	< 0.001	0.075 – 0.077	0.002 – 0.091
		Inter	17,600	0.129	< 0.001	0.128 – 0.129	0.101 – 0.157
3f	228	Intra	25,878	0.065	< 0.001	0.065 – 0.065	0.019 – 0.099
		Inter	57,456	0.142	< 0.001	0.142 – 0.142	0.103 – 0.157
3h	17	Intra	136	0.05	0.003	0.045 – 0.055	0.002 – 0.095
		Inter	7871	0.132	< 0.001	0.132 – 0.133	0.098 – 0.157
3i	6	Intra	15	0.073	0.005	0.062 – 0.084	0.025 – 0.09
		Inter	2844	0.127	< 0.001	0.126 – 0.128	0.094 – 0.152
3k	4	Intra	6	0.033	0.006	0.019 – 0.048	0.019 – 0.048
		Inter	1914	0.132	< 0.001	0.131 – 0.133	0.088 – 0.156
3l	6	Intra	15	0.064	0.006	0.051 – 0.077	0.007 – 0.081
		Inter	2844	0.134	< 0.001	0.134 – 0.134	0.093 – 0.157
3m	6	Intra	15	0.054	0.006	0.042 – 0.069	0.004 – 0.069
		Inter	2844	0.131	< 0.001	0.13 – 0.132	0.1 – 0.156
3ra	17	Intra	171	0.116	0.002	0.111 – 0.121	0.026 – 0.161
		Inter	8626	0.181	< 0.001	0.181 – 0.181	0.166 – 0.193

p-distances were calculated with the maximum composite likelihood method (MCL) on MEGA X.

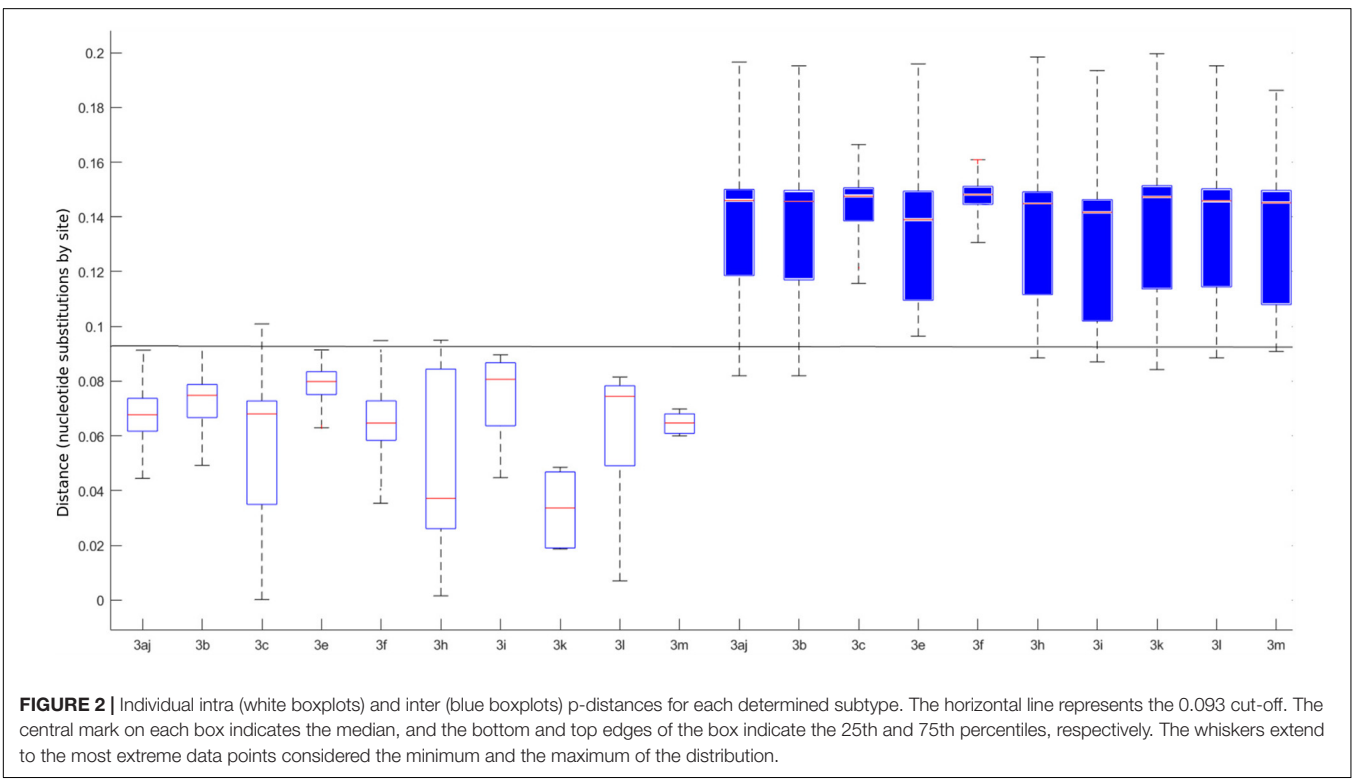

FIGURE 2 | Individual intra (white boxplots) and inter (blue boxplots) p-distances for each determined subtype. The horizontal line represents the 0.093 cut-off. The central mark on each box indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points considered the minimum and the maximum of the distribution.

TABLE 2 | HEV-3 subtype reference sequences based on full or near full-length genomes.

HEV3 subtype according to Smith et al. (2020)	Genbank accession number	Strain	Classification based on clustering or distance analyses	Geographical origin	Comment
3a	AF082843	Meng	3aj	Canada, China, France, Japan, Korea, Mexico, Singapore, Thailand, United Kingdom, United States	
3b	AP003430	JRA1	3b	Canada, China, France, Japan	
3c	FJ705359	wbGER27	3c	France, Germany, Netherlands, Sweden, Thailand, United Kingdom	
3e	AB248521	swJ8-5	3e	Germany, Japan, France, Hungary, Italy, United Kingdom	
3f	AB369687	E116-YKH98C	3f	Denmark, France, Germany, Japan, Singapore, Spain, Sweden, Thailand, United Kingdom	
3g	AF455784	Osh	3g	Kyrgyzstan	Only one complete genome sequence for this subtype
3h	JQ013794	TR19	3h	France, Mongolia, Switzerland	
3i	FJ998008	BB02	3i	France, Germany, Sweden	
3j	AY115488	Arkell	–	Canada	Isolated from pooled stools (Pei and Yoo, 2002), classified 3a
3k	AB369689	E088-STM04C	3k	Japan	
3l	JQ953664	FR-SHEV3c-like	3l	France, Italy	
3m	KU513561	IC2011	3m	France, Spain	
3	AB290313	swMN06-C1056	3f	Mongolia	
3	MF959765	WB/HEV/NA21ITA15	3	Italy	
3	LC260517	swHE1606845	3	Japan	
3	MK390971	17RS1920	3	Italy	
3	MF959764	WB/HEV/NA17ITA15	3i	Italy	
3	KP294371	MWP_2010	3i	Germany	
3ra	FJ906895	GDC9	3ra	China, France, Germany, Japan, United States	

before 2015 were replaced by subtype 3c strains after 2015 (Suin et al., 2019). The reason for these changes in subtype distribution is uncertain, it could reflect the distributions of HEV-3 subtypes in the pig reservoirs of different countries. Both locally produced and imported pigs or pork meat could be involved. Phylogenetic and coalescence analyses based on full-length sequences of HEV-3 from acute hepatitis patients, domestic pigs and wild boars provide evidence that HEV-3e strains were introduced from Europe into Japan through importation of pigs in the 1960s (Nakano et al., 2013). Transmission of HEV-3e strains from pigs to wild boars has been also suggested in Japan (Nakano et al., 2013). HEV-3f subtypes were recently detected in humans, domestic pigs and wild mammals in Japan, but indigenous Japanese HEV-3 strains belong to subtypes 3a, 3b, and 3e (Nakano et al., 2018). These new HEV-3f strains may have entered Japan from Europe in this way because the proportion of pork meat imported from Europe has increased in the past decade, leading to cases of hepatitis due to eating pork meat. The changes in HEV-3 subtype distribution are probably the result of changes in the origin of pork meat.

The clinical significance of infection with different HEV-3 subtypes has been discussed. Most studies have shown that asymptomatic blood donors and patients with symptomatic hepatitis E had similar genotype distributions and neither the severity of symptoms nor liver enzyme activities were significantly associated with clades 3.1 or 3.2 (Smith et al., 2015; Lhomme et al., 2019). However, two recent studies from Belgium and France found that the risk of HEV-3-infected patients being hospitalized varied with the subtype (Subissi et al., 2019; Abravanel et al., 2020). Patients infected with subtype 3c were at lower risk of hospitalization than those infected with subtypes 3f or 3e (Subissi et al., 2019; Abravanel et al., 2020). Larger studies are now needed to clarify the influence of host factors and virus diversity on HEV-3 pathogenesis.

A limitation of the present study is the relatively limited number of full length HEV-3 genome sequences available worldwide and the way diversity varies within HEV-3 subtypes. Despite very good groupings of sequences, there are outliers for most subtypes indicating that subtype assignment can be ambiguous. In addition, HEV-3ra strains are particularly heterogeneous. Nevertheless, all HEV-3ra sequences have a common signature, a 93-nucleotide insertion within the macrodomain of the HEV genome (Izopet et al., 2012).

Our findings suggest that the strains in clades 3.1 and 3.2 can be assigned to 1 of 11 subtypes, each represented by a full length or near full-length reference sequence. We have proposed a cut-off value for assigning subtypes. Update of the reference sequences (Smith et al., 2020) could help harmonize HEV-3 classification, which would be useful for comparing strains circulating in humans and the animal reservoir, for tracing the source of an individual infection and for investigating the pathogenicity of HEV-3 subtypes.

DATA AVAILABILITY STATEMENT

The sequences have been submitted in Genbank database. The accession numbers corresponding the sequences data are MW355217–MW355404.

AUTHOR CONTRIBUTIONS

FN and JI designed the project. FN, CD, and JI analyzed the results and wrote the manuscript. NJ, JL, and FN performed the bio-informatics analyses. CD performed the statistics analysis. NK provided the plasma samples. FA, SLh, SC-R, NR, AH, and MD carried out the experiments. SLa and SB performed Sequel sequencing. All authors contributed to the article and approved the submitted version.

FUNDING

High-throughput sequencing has been performed by the ICGex NGS platform of the Institut Curie supported by the grants ANR-10-EQPX-03 (Equipex) and ANR-10-INBS-09-08 (France Génomique Consortium) from the Agence Nationale de la Recherche ("Investissements d'Avenir" program), by the Canceropole Ile-de-France, and by SiRIC-Curie program-SiRIC Grant "INCa-DGOS-4654."

ACKNOWLEDGMENTS

The English text was checked by Owen Parkes.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.634430/full#supplementary-material

REFERENCES

Abravanel, F., Dimeglio, C., Castanier, M., Peron, J. M., Kamar, N., Lhomme, S., et al. (2020). Does HEV-3 subtype play a role in the severity of acute hepatitis E? *Liver Int.* 40, 333–337. doi: 10.1111/liv.14329

Adlhoch, C., Wolf, A., Meisel, H., Kaiser, M., Ellerbrok, H., and Pauli, G. (2009). High HEV presence in four different wild boar populations in east and west germany. *Vet. Microbiol.* 139, 270–278. doi: 10.1016/j.vetmic.2009.06.032

De Sabato, L., Lemey, P., Vrancken, B., Bonfanti, L., Ceglie, L., Vaccari, G., et al. (2018). Proposal for a new subtype of the zoonotic genotype 3 Hepatitis E virus: HEV-3l. *Virus Res.* 248, 1–4. doi: 10.1016/j.virusres.2018.02.010

Ijaz, S., Said, B., Boxall, E., Smit, E., Morgan, D., and Tedder, R. S. (2014). Indigenous hepatitis E in England and wales from 2003 to 2012: evidence of an emerging novel phylotype of viruses. *J. Infect Dis.* 209, 1212–1218. doi: 10.1093/infdis/jit652

Izopet, J., Dubois, M., Bertagnoli, S., Lhomme, S., Marchandeau, S., Boucher, S., et al. (2012). Hepatitis E virus strains in rabbits and evidence of a closely related strain in humans, France. *Emerg. Infect. Dis.* 18, 1274–1281. doi: 10.3201/eid1808.120057

Kamar, N., Izopet, J., Pavio, N., Aggarwal, R., Labrique, A., Wedemeyer, H., et al. (2017). Hepatitis E virus infection. *Nat. Rev. Dis. Primers* 3:17086. doi: 10.1038/nrdp.2017.86

Legrand-Abravanel, F., Mansuy, J. M., Dubois, M., Kamar, N., Peron, J. M., Rostaing, L., et al. (2009). Hepatitis E virus genotype 3 diversity, France. *Emerg. Infect. Dis.* 15, 110–114. doi: 10.3201/eid1501.080296

Lhomme, S., Gallian, P., Dimeglio, C., Assal, A., Abravanel, F., Tiberghein, P., et al. (2019). Viral load and clinical manifestations of hepatitis E virus genotype 3 infections. *J Viral Hepat* 26, 1139–1142. doi: 10.1111/jvh.13128

Lu, L., Li, C., and Hagedorn, C. H. (2006). Phylogenetic analysis of global hepatitis E virus sequences: genetic diversity, subtypes and zoonosis. *Rev. Med. Virol.* 16, 5–36. doi: 10.1002/rmv.482

Miura, M., Inoue, J., Tsuruoka, M., Nishizawa, T., Nagashima, S., Takahashi, M., et al. (2017). Full-length genomic sequence analysis of new subtype 3k hepatitis E virus isolates with 99.97% nucleotide identity obtained from two consecutive acute hepatitis patients in a city in northeast Japan. *J. Med. Virol.* 89, 1116–1120. doi: 10.1002/jmv.24743

Munoz-Chimeno, M., Forero, J. E., Echevarria, J. M., Munoz-Bellido, J. L., Vazquez-Lopez, L., Morago, L., et al. (2016). Full coding hepatitis E virus genotype 3 genome amplification method. *J. Virol. Methods* 230, 18–23. doi: 10.1016/j.jviromet.2016.01.004

Nakano, T., Takahashi, K., Arai, M., Okano, H., Kato, H., Ayada, M., et al. (2013). Identification of European-type hepatitis E virus subtype 3e isolates in Japanese wild boars: molecular tracing of HEV from swine to wild boars. *Infect. Genet. Evol.* 18, 287–298. doi: 10.1016/j.meegid.2013.06.004

Nakano, T., Takahashi, M., Takahashi, K., Nagashima, S., Suzuki, Y., Nishigaki, Y., et al. (2018). Hepatitis E virus subtype 3f strains isolated from Japanese hepatitis patients with no history of travel to endemic areas - the origin analyzed by molecular evolution. *Virology* 513, 146–152. doi: 10.1016/j.virol.2017.08.008

Nicot, F., Jeanne, N., Roulet, A., Lefebvre, C., Carcenac, R., Manno, M., et al. (2018). Diversity of hepatitis E virus genotype 3. *Rev. Med. Virol.* 28:e1987. doi: 10.1002/rmv.1987

Norder, H., Galli, C., Magnil, E., Sikora, P., Ekvarn, E., Nystrom, K., et al. (2018). Hepatitis E virus genotype 3 genomes from RNA-positive but serologically negative plasma donors have CUG as the start Codon for ORF3. *Intervirology* 61, 96–103. doi: 10.1159/000491926

Oeser, C., Vaughan, A., Said, B., Ijaz, S., Tedder, R., Haywood, B., et al. (2019). Epidemiology of hepatitis E in England and wales: a 10-year retrospective surveillance study, 2008-2017. *J. Infect. Dis.* 220, 802–810. doi: 10.1093/infdis/jiz207

Oliveira-Filho, E. F., Konig, M., and Thiel, H. J. (2013). Genetic variability of HEV isolates: inconsistencies of current classification. *Vet. Microbiol.* 165, 148–154. doi: 10.1016/j.vetmic.2013.01.026

Pei, Y., and Yoo, D. (2002). Genetic characterization and sequence heterogeneity of a canadian isolate of swine hepatitis E virus. *J. Clin. Microbiol.* 40, 4021–4029. doi: 10.1128/jcm.40.11.4021-4029.2002

Primadharsini, P. P., Nagashima, S., and Okamoto, H. (2019). Genetic variability and evolution of hepatitis E virus. *Viruses* 11:27. doi: 10.3390/v11050456

Prosperi, M. C., Ciccozzi, M., Fanti, I., Saladini, F., Pecorari, M., Borghi, V., et al. (2011). A novel methodology for large-scale phylogeny partition. *Nat. Commun.* 2:321. doi: 10.1038/ncomms1325

Purdy, M. A., Harrison, T. J., Jameel, S., Meng, X. J., Okamoto, H., Van der Poel, W. H. M., et al. (2017). ICTV virus taxonomy profile: hepeviridae. *J. Gen. Virol.* 98, 2645–2646. doi: 10.1099/jgv.0.000940

Sahli, R., Fraga, M., Semela, D., Moradpour, D., and Gouttenoire, J. (2019). Rabbit HEV in immunosuppressed patients with hepatitis E acquired in Switzerland. *J. Hepatol.* 70, 1023–1025. doi: 10.1016/j.jhep.2019.01.025

Smith, D. B., Ijaz, S., Tedder, R. S., Hogema, B., Zaaijer, H. L., Izopet, J., et al. (2015). Variability and pathogenicity of hepatitis E virus genotype 3 variants. *J. Gen. Virol.* 96, 3255–3264. doi: 10.1099/jgv.0.000264

Smith, D. B., Izopet, J., Nicot, F., Simmonds, P., Jameel, S., Meng, X. J., et al. (2020). Update: proposed reference sequences for subtypes of hepatitis E virus (species *Orthohepevirus* A). *J. Gen. Virol.* 101, 692–698. doi: 10.1099/jgv.0.001435

Smith, D. B., and Simmonds, P. (2018). Classification and genomic diversity of enterically transmitted hepatitis viruses. *Cold Spring Harb. Perspect. Med.* 8:a031880. doi: 10.1101/cshperspect.a031880

Smith, D. B., Simmonds, P., Izopet, J., Oliveira-Filho, E. F., Ulrich, R. G., Johne, R., et al. (2016). Proposed reference sequences for hepatitis E virus subtypes. *J. Gen. Virol.* 97, 537–542. doi: 10.1099/jgv.0.000393

Sridhar, S., Yip, C. C. Y., Wu, S., Cai, J., Zhang, A. J., Leung, K. H., et al. (2018). Rat hepatitis E virus as cause of persistent hepatitis after liver transplant. *Emerg. Infect. Dis.* 24, 2241–2250. doi: 10.3201/eid2412.180937

Subissi, L., Peeters, M., Lamoral, S., Klamer, S., Suin, V., and Van Gucht, S. (2019). Subtype-specific differences in the risk of hospitalisation among patients infected with hepatitis E virus genotype 3 in Belgium, 2010-2018. *Epidemiol. Infect.* 147:e224. doi: 10.1017/S0950268819001122

Suin, V., Klamer, S. E., Hutse, V., Wautier, M., Jacques, M., Abady, M., et al. (2019). Epidemiology and genotype 3 subtype dynamics of hepatitis E virus in belgium, 2010 to 2017. *Euro. Surveill.* 24:1800141. doi: 10.2807/1560-7917.ES.2019.24.10.1800141

Vina-Rodriguez, A., Schlosser, J., Becher, D., Kaden, V., Groschup, M. H., and Eiden, M. (2015). Hepatitis E virus genotype 3 diversity: phylogenetic analysis and presence of subtype 3b in wild boar in Europe. *Viruses* 7, 2704–2726. doi: 10.3390/v7052704

Wang, H., Castillo-Contreras, R., Saguti, F., Lopez-Olvera, J. R., Karlsson, M., Mentaberre, G., et al. (2019). Genetically similar hepatitis E virus strains infect both humans and wild boars in the Barcelona area, Spain, and Sweden. *Transbound. Emerg. Dis.* 66, 978–985. doi: 10.1111/tbed.13115

Wist, V., Kubacki, J., Lechmann, J., Steck, M., Fraefel, C., Stephan, R., et al. (2018). Complete genome sequence of a swiss hepatitis E virus isolate from the liver of a fattening pig. *Genome Announc.* 6:e113-18. doi: 10.1128/genomeA.00113-18

Zehender, G., Ebranati, E., Lai, A., Luzzago, C., Paladini, S., Tagliacarne, C., et al. (2014). Phylogeography and phylodynamics of European genotype 3 hepatitis E virus. *Infect. Genet. Evol.* 25, 138–143. doi: 10.1016/j.meegid.2014.04.016

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 28 January 2021
doi: 10.3389/fmicb.2020.617375

Prevalence of Naturally-Occurring NS5A and NS5B Resistance-Associated Substitutions in Iranian Patients With Chronic Hepatitis C Infection

Pooneh Rahimi[1], Heidar Sharafi[2]*, Golnaz Bahramali[1], FaridehSadat SajadianFard[1], Nafiseh Sadat Asadi[1], Seyed Moayed Alavian[2], Vahid Iranpur Mobarakeh[1] and Seyedeh Zahra Moravej[1]*

[1] Department of Hepatitis and AIDS, Pasteur Institute of Iran, Tehran, Iran, [2] Middle East Liver Diseases Center, Tehran, Iran

OPEN ACCESS

Edited by:
Lilly Yuen,
Victorian Infectious Diseases Reference Laboratory, Australia

Reviewed by:
Stephen J. Polyak,
University of Washington, United States
Jorge Quarleri,
Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina

***Correspondence:**
Pooneh Rahimi
Pooneh5376@yahoo.com
Heidar Sharafi
h.sharafi@meldcenter.com

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: 14 October 2020
Accepted: 30 December 2020
Published: 28 January 2021

Citation:
Rahimi P, Sharafi H, Bahramali G, SajadianFard FS, Asadi NS, Alavian SM, Iranpur Mobarakeh V and Moravej SZ (2021) Prevalence of Naturally-Occurring NS5A and NS5B Resistance-Associated Substitutions in Iranian Patients With Chronic Hepatitis C Infection. Front. Microbiol. 11:617375. doi: 10.3389/fmicb.2020.617375

Background: Hepatitis C virus (HCV), non-structural 5A (NS5A), and non-structural 5B (NS5B) resistance-associated substitutions (RASs) are the main causes of failure to direct-acting antiviral agents (DAAs). NS5A and NS5B RASs can occur in patients with HCV infection naturally and before exposure to DAAs.

Objectives: This study aimed to evaluate naturally-occurring NS5A and NS5B RASs in Iranian patients with HCV genotype 1a (HCV-1a) and -3a infections.

Methods: In this cross-sectional study, viral RNA was extracted from serum specimens. NS5A and NS5B regions were amplified using RT-PCR followed by DNA sequencing. The results of nucleotide sequences were aligned against reference sequences of HCV-1a and -3a and the amino acid substitutions were analyzed using geno2pheno [hcv] web application.

Results: Among 135 patients with hepatitis C, NS5A amino acid substitutions/RASs were identified in 26.4% and 15.9% of patients with HCV-1a and -3a infections, respectively. The identified amino acid substitutions/RASs in the NS5A region of patients with HCV-1a infection were M28T/V/I 11.1%, Q30R/H 4.2%, L31M 1.4%, and H58Y/P/C/D/Q/S/T 16.7%. Y93H substitution was not found in HCV-1a sequences. In patients with HCV-3a infection, NS5A amino acid substitutions/RASs were A30T/K 9.5%, L31F 1.6%, P58S/T/C 3.2%, Y93H 3.2%, and Y93N 3.2%. No resistance substitutions were identified in NS5B sequences from patients with HCV-1a and -3a infections.

Conclusion: In this study, baseline amino acid substitutions/RASs were only identified in the NS5A region in Iranian patients with HCV-1a and -3a infections, and the prevalence of these amino acid substitutions/RASs were in accordance with similar studies. There were no RASs in the HCV-1a and -3a NS5B region.

Keywords: HCV, Direct-acting antiviral agents, NS5A, NS5B, Resistance-associated substitution

INTRODUCTION

It has been estimated that approximately 71 million people worldwide are chronically infected with the hepatitis C virus (HCV). This viral infection is the leading cause of acute and chronic hepatitis with a spectrum of mild illness to end-stage liver disease such as hepatocellular carcinoma (HCC) with about 399,000 deaths each year (Bittar et al., 2010; Paolucci et al., 2013; Blach et al., 2017). The virus belongs to the Hepacivirus genus in the Flaviviridae family with a genome of positive-sense ssRNA with approximately 9.6 Kb encoding for a polyprotein precursor which is cleaved by the viral and host proteinases into the structural (core, E1, E2, and P7) and the non-structural proteins (NS2, NS3, NS4A, NS4B, NS5A, and NS5B) (Bittar et al., 2010; Paolucci et al., 2013; Blach et al., 2017). HCV is one of the blood-borne viruses; so, the most common modes of transmission are through exposure to small quantities of blood or its derivatives that make this infection one the worldwide public health problems (Paolucci et al., 2013; Khodabandehloo and Roshani, 2014; Taherkhani and Farshadpour, 2015; Blach et al., 2017; Mahmud et al., 2018). The prevalence of HCV infection in Iran is estimated between 0.3 and 0.5% among the general population to 32.1% among high-risk populations (Khodabandehloo and Roshani, 2014; Taherkhani and Farshadpour, 2015; Mahmud et al., 2018).

Up to now, seven genotypes with many subtypes have been identified for HCV and the distribution of these genotypes is varying between different populations (Murphy et al., 2007, 2015; Kau et al., 2008; Sarrazin and Zeuzem, 2010; Nakano et al., 2012; Smith et al., 2014). According to the molecular epidemiology studies, HCV genotype 1a (HCV-1a), HCV-3a, and with the lower frequency HCV-1b are the most prevalent HCV genotypes in Iran (Khodabandehloo and Roshani, 2014; Taherkhani and Farshadpour, 2015; Mahmud et al., 2018). Viral RNA-dependent RNA polymerase of HCV which lacks 3′ to 5′ exonuclease proofreading activity leads to a high genome replication rate with no fidelity, and enables HCV to escape from selective pressures of the immune response and antiviral therapies (Kau et al., 2008; Paolucci et al., 2013; Murphy et al., 2015; de Rueda et al., 2017). All these characteristics make HCV a challenging target for designing effective vaccine and antiviral drugs (Murphy et al., 2007; Gottwein et al., 2009; Nakano et al., 2012; Smith et al., 2014; Walker et al., 2015; Alavian and Sharafi, 2017; de Rueda et al., 2017; Mahmud et al., 2018). For many years, pegylated interferon (PegIFN) plus ribavirin (RBV) has been used for the treatment of chronic HCV infection. However, treatment with PegIFN and RBV had undesirable side-effects and suboptimal responses or even failure to treatment especially in patients with HCV-1 and -4 and those with cirrhosis (Kau et al., 2008; Sarrazin and Zeuzem, 2010). Recently, direct-acting antiviral agents (DAAs) have been developed against different functional proteins of HCV, which seem to be promising. However, the efficiency of these drugs could be impacted negatively by the existence of resistance-associated substitutions (RASs) both at the baseline before initiation of treatment or re-treatment following the previous failure to DAA-based regimens, infection with HCV genotypes that used to be known as hard-to-treat genotypes and cirrhosis (Mahmud et al., 2018; Ghany et al., 2019). Although using a combination of DAAs results in inhibition of HCV replication efficiently, the mutable nature of the HCV genome makes it necessary to put the occurrence of emerging substitutions associated with DAAs resistance in patients with HCV infection under precise surveillance. Herein, we report the naturally-occurring NS5A and NS5B RASs causing resistance to DAAs in a cohort of DAA-naïve patients with chronic HCV-1a and -3a infections.

MATERIALS AND METHODS

Study Population and Sample Collection

This cross-sectional study recruited patients managed at the hepatitis clinic of Digestive Disease Research Institute (DDRI), Tehran, Iran from 2015 to 2017. In this study, adults (>18 years old) with chronic HCV-1a and -3a infections were included. Hepatitis C chronicity was defined as being positive for HCVAb and HCV RNA for more than 6 months. Co-infection with HIV and/or HBV or any condition that leads to being immunocompromised, being diagnosed with HCC, and having a previous history of treatment with DAAs were considered as exclusion criteria. Cirrhosis was diagnosed based on clinical or histological measurements or with non-invasive assessment by transient elastography. The baseline HCV RNA level and HCV genotype were assessed before the initiation of HCV antiviral therapy and the results were available as the routine workup for diagnosis and management of HCV infection in patients' records. These tests were carried out in the diagnostic laboratories as commercial accredited services. The blood sampling procedures were explained to the patients clearly and the consent form for using their blood samples for further analysis was signed by each volunteer or their official custodian. This study was approved by the Ethics Committee of the Pasteur Institute of Iran (no: IR.PII.REC.1395.81) according to the standard biosecurity and institutional safety procedures. The current study was conducted according to the Helsinki Declaration of 1975, as revised in 2008.

HCV Genome Extraction and Amplification

Viral RNA was extracted from 140 μL of baseline serum samples using a commercially available kit (HighPure Viral Nucleic Acid kit, Roche, Germany). cDNA synthesis and first-round PCR amplifications were done in a 25 μL reaction mixture using PrimeScript^{TM} One-Step RT-PCR Kit (TaKaRa, Clontech, Japan). The following primers were used for the amplification of the HCV-1a NS5A region: FO1ans5a and RO1ans5a (**Supplementary Table 1**). NS5B region is a long region with approximately 1776 nucleotides. To study the most possible complete sequences of this region, we divided it into two sub-regions as NS5B1 and NS5B2 by designing specific primers for each region: FO1ans5b1 and RO1ans5b1 (NS5B1), and FO1ans5b2 and RO1ans5b2 (NS5B2; **Supplementary Table 1**). The amplification reaction was performed according to the manufacturer's instruction following the program: incubation at 50°C for 30 min and initial denaturation at 94°C for 2 min,

then 35 cycles of denaturation at 94°C for 30 s, annealing at 59°C (for NS5A outer primers), 61°C (for NS5B1 outer primers), and 59°C (for NS5B2 outer primers) for 30 s, extension at 72°C for 1 min and a final extension at 72°C for 10 min. Then, the second round of PCR was performed using TaKaRa Ex Taq^{TM} kit (TaKaRa, Clontech, Japan) and specific inner primers for each region resulted in specific PCR products as a 510 bp fragment for NS5A, and a 797 bp and a 928 bp PCR product for NS5B1 and NS5B2, respectively, (**Supplementary Table 1**). This amplification was performed in a 50 μL reaction mixture according to the manufacturer's instruction and the program was initiated with denaturation at 94°C for 5 min and 35 cycles including denaturation at 94°C for 1 min, annealing at 60°C (for NS5A inner primers), 59°C (for NS5B1 inner primers), and 58°C (for NS5B2 inner primers) for 30 s, extension at 72°C for 7 min and a final extension at 72°C for 10 min.

Similar procedures were executed for the amplification of NS5A and NS5B regions from HCV-3a using specific primers. Briefly, specific outer primers were used for the first round of NS5A and NS5B including FO3ans5a and RO3ans5a (NS5A), FO3ans5b1 and RO3ans5b1 (NS5B1), and FO3ans5b2 and RO3ans5b2 (NS5B2; **Supplementary Table 1**). The PCR program was performed in the same way according to the manufacturer's instruction as it was done for HCV-1a except for the annealing temperatures which were 61°C for NS5A, 59°C for NS5B1, and 63°C for NS5B2 amplification. Those first-round PCR products were used as templates for the nested round of PCR by using inner primers including FI3ans5a and RI3ans5a (NS5A), FI3ans5b1 and RI3ans5b1 (NS5B1), and FI3ans5b2 and RI3ans5b2 (NS5B2; **Supplementary Table 1**). The amplification procedure for the nested round of PCR was the same as the nested round of PCR for the amplification of these regions of HCV-1a. The annealing temperature for NS5A was 60°C and 58°C was used to amplify inner parts of both NS5B1 and NS5B2. The specific bands of HCV-3a NS5A, NS5B1, and NS5B2 were identified with 348, 691, and 768 bp of PCR products, respectively. The PCR products underwent electrophoresis on a 1.5% agarose gel containing safe stain (YTA, Iran) and the specific bands were purified using the Yekta Tajhiz Azma Gel and PCR purification kit (YTA, Iran). Then the purified PCR fragments accompanied with the related specific inner primers for each region were sent to GenFanavaran company to be sequenced (Sanger sequencing) in South Korea in both directions using a BigDye Terminator cycle sequencing kit (Perkin Elmer–Applied Biosystems Inc., CA, United States). The chromatogram was used to assess the nucleotide substitutions in whole sequences, and the nucleotide redundancy was considered when representing ≥15% of the sequence population.

Nucleotide sequences were first analyzed by Bioedit software (v.7.9.5). Then they were aligned against the HCV-1a reference sequences (GenBank accession numbers: EU255982.1, KX621456.1, JX463555.1, KJ747896.1, and HQ891277.1), and the HCV-3a reference sequences (GenBank accession numbers: GQ275355.1, KJ470615.1, HW121730.1, JN689927.1, and KF944665.1) separately using ClustalW integrated into MEGA 6 software. Amino acid positions, RASs, and RASs > 100X were selected using the geno2pheno [HCV] rules (GENAFOR, 2019; **Table 1**). Geno2pheno [HCV] (publically available at http://hcv.geno2pheno.org) is a basic tool for interpretation and evaluation of viral sequences for susceptibility to anti-HCV DAAs. The same analysis was done to investigate the RASs in the NS5B region for each HCV-1a and -3a, separately (Hernandez et al., 2013; Ghany et al., 2019, 2020). The HCV-1a and -3a sequences obtained by Sanger sequencing in this study have been deposited in the GenBank under accession numbers; HCV-1a NS5A; MT259593-MT259659, HCV-1a NS5B: MT454922-MT455006, HCV-3a NS5A: MT347782-MT347803, and HCV-3a NS5B: MT502204-MT502244, respectively.

Phylogenetic Analyses

The sequences of HCV-1a and -3a obtained from patients with HCV infection, as well as HCV-1a and -3a reference sequences from the NCBI nucleotide database, were used to build phylogenetic trees for each HCV genotype, separately. Phylogenetic trees have been constructed automatically using the Kimura 2-parameter approach implemented in the statistical method Neighbor-Joining and bootstrap values were calculated using 1000 bootstrap iterations.

Statistical Analysis

Data were analyzed using SPSS version (25.0.0.0. Chicago, IL, United States). The $P < 0.05$ was considered statistically significant.

RESULTS

Baseline Characteristics of Study Population

Both target regions (NS5A and NS5B) of HCV were successfully amplified in 72/72 and 63/63 patients who were infected with HCV-1a and -3a, respectively. Baseline demographic characteristics of 72 patients with HCV-1a infection and 63 with HCV-3a infection have been summarized in **Table 2**. The mean age was 41.6 ± 8.7 and 43.4 ± 9.2 years for HCV-1a, and -3a groups, respectively. In the HCV-1a group, 42/72 (58.3%) were males, and 30/72 (41.7%) patients were females. There were 60.3% (38/63) males, and 39.7% (25/63) females in patients with HCV-3a infection. Their mean HCV RNA level was 5.2 ± 3.8 Log IU/ml, and 5.8 ± 4.1 Log IU/mL for HCV-1a and -3a, respectively. The majority (56.9%) of patients in the HCV-1a group had cirrhosis, and 43.1% of patients had no evidence of cirrhosis. In patients with HCV-3a infection, cirrhosis had been diagnosed in 63.5% of them, and 36.5% were non-cirrhotic. All patients were DAA-naïve, however, 44.4% of patients with HCV-1a and 30.2% of patients with HCV-3a had a previous history of IFN-based treatment. Among HCV-1a and -3a patients, 12 (16.7%), and 17 (27%) had cirrhosis and also, a previous history of IFN-based treatment. Data has been depicted in detail in **Table 2**.

TABLE 1 | The amino acid positions, consensus amino acids, amino acid substitutions, resistance-associated substitutions (RASs), and the RASs with >100 resistance fold change.

	Amino acid position	HCV-1a			HCV-3a		
		Cons.[a]	RAS[b]	RAS > 100X[c]	Cons.[a]	RAS[b]	RAS > 100X[c]
NS5A RASs	28	M	A, G, T	A, G, T	M	–	–
	30	Q	D, E, G, H, K, N, R, T, Y	D, E, G, H, K, N, R, Y	A	K	–
	31	L	F, I, M, V	F, I, M, V	L	F, I, M, V	F, I, M, V
	32	P	L	L	P	–	–
	58	H	D	D	P	–	–
	93	Y	C, H, N, R, S, T, W	C, H, N, R, S, W	Y	H	H
NS5B RASs	282	S	T	–	S	T	–

[a]*Consensus amino acid was selected based on the alignment of the included sequences by each HCV genotype.* [b]*Amino acid substitutions conferring > 2 resistance fold change (GENAFOR, 2019).* [c]*Amino acid substitutions conferring > 100 resistance fold change (Sorbo et al., 2018). Abbreviations: Cons., consensus; RAS, resistance-associated substitution; A, alanine; G, glycine; T, threonine; V, valine; C, cysteine; E, glutamate; H, histidine; I, isoleucine; K, lysine; R, arginine; S, serine; Y, tyrosine; F, phenylalanine; M, methionine; D, aspartate; N, asparagine; W, tryptophan; L, leucine; Q, glutamine; and P, proline.*

TABLE 2 | Characteristics of direct-acting antiviral agent-naïve patients with hepatitis C virus infection (genotypes 1a and 3a).

		HCV-1a (*n* = 72)	HCV-3a (*n* = 63)
Sex, *n* (%)	Male	42 (58.3)	38 (60.3)
	Female	30 (41.7)	25 (39.7)
Age (years)	mean ± SD	41.6 ± 8.7	43.4 ± 9.2
HCV RNA level (log IU/mL)	mean ± SD	5.2 ± 3.8	5.8 ± 4.1
Cirrhosis condition, *n* (%)	Non-cirrhotic	31 (43.1)	23 (36.5)
	Cirrhotic	41 (56.9)	40 (63.5)
History of treatment with IFN-based regimens	Naïve	40 (55.6)	44 (69.8)
	IFN-experienced	32 (44.4)	19 (30.2)

Abbreviations: HCV, hepatitis C virus; IU/mL, international units per milliliter; and IFN, interferon.

Prevalence of NS5A Amino Acid Substitutions and RASs

The frequency of identified NS5A amino acid substitutions/RASs in males was higher than in females, however, the difference in both HCV genotypes (1a and 3a) was not statistically significant ($P > 0.05$). In the patients with HCV-1a infection, the prevalence of NS5A amino acid substitutions/RASs was 26.4% (19/72) from which, 12 patients had cirrhosis and 7 patients without cirrhosis. In these patients, the identified NS5A amino acid substitutions/RASs were 11.1% (8/72) in amino acid (aa) 28 (M28T = 2, M28V = 5, and M28I = 1), 4.2% (3/72) of aa 30 (Q30R = 2 and Q30H = 1), 1.4% (1/72) of aa 31 (L31M), and 16.7% (12/72) of aa 58 (H58Y = 2, H58C = 2, H58D = 2, H58P = 2, H58Q = 1, and H58S/T/C = 3). In HCV-1a sequences, the RASs > 100X were identified in 8.3% (6/72) patients including M28T (2 patients), Q30R and Q30H (2 and 1 patients, respectively), and L31M (1 patient). From these six clinically relevant RASs > 100X, five (M28T in 1 patient, Q30R in 2, Q30H in 1, and L31M in 1) were identified in patients with cirrhosis that had a previous history of treatment with IFN-based regimens, and one (M28T) was detected in a non-cirrhotic patient ($P < 0.05$). None of the HCV-1a NS5A sequences harbored the Y93H substitution. Paired NS5A amino acid substitutions were identified in 5.6% (4/72) patients (each in 1 patient) including M28V+H58S/T/C, L31M+H58C, M28V+Q30R, and M28T+Q30H. The results of detected RASs in HCV-1a NS5A have been shown in **Table 3**.

In patients with HCV-3a infection, the number of patients harboring known NS5A amino acid substitutions/RASs at baseline was 10 out of 63 patients (15.9%) which 7 patients had cirrhosis and previous history of IFN-based treatment, and the remaining 3 patients were non-cirrhotic with previous history IFN-based treatment. Six (9.5%) amino acid substitutions/RASs were identified in aa 30 (A30T/K), 1 (1.6%) in aa 31 (L31F), 2 (3.2%) in aa 58 (P58S/T/C), 4 (6.3%) in aa 93 including 2 (3.2%) Y93H, and 2 (3.2%) Y93N. From these amino acid substitutions, 3/63 (4.8%) were identified as RASs; 1 as RAS L31F and 2 as RAS Y93H. In patients with HCV-3a infection, multiple amino acid variants in a single position were detected totally in 6.3% (4/63) of patients including A30T/K in 2 patients, and P58S/T/C in 2 patients. The clinically important substitution (RAS > 100X) Y93H was detected in 2 (3.2%) patients with cirrhosis and previous history of IFN-based treatment. Paired NS5A amino acid substitutions/RASs were identified in 2 (3.2%) patients with cirrhosis: L31F+P58S/T/C+Y93H and A30T/K+P58S/T/C each in one patient. **Table 3** has presented the results of identified NS5A RASs in HCV-3a in this study.

Prevalence of NS5B RASs

In this study, those NS5B RASs conferring resistance to nucleotide analog inhibitors (Sofosbuvir) were not identified.

TABLE 3 | Baseline NS5A amino acid substitutions/resistance-associated substitutions in patients with HCV infection.

HCV genotype	RASs, Position/amino acid	Number and proportion	Paired RASs, Position/amino acid	Total number and proportion of paired RASs
1a	M28T/V/I	8/72 (11.1%)	M28V+H58S/T/C	1/72 (1.4%)
	Q30R/H	3/72 (4.2%)	M28V+Q30R	1/72 (1.4%)
	L31M	1/72 (1.4%)	M28T+Q30H	1/72 (1.4%)
	H58Y/P/C/D/Q/S/T	12/72 (16.7%)	L31M+H58C	1/72 (1.4%)
3a	A30T/K	6/63 (9.5%)	L31F+P58S/T/C+Y93H	1/63 (1.6%)
	L31F	1/63 (1.6%)	A30T/K+P58S/T/C	1/63 (1.6%)
	P58S/T/C	2/63 (3.2%)		
	Y93H/N	4/63 (6.3%)		

Phylogenetic Analysis

The genetic relationship among NS5A and NS5B sequences from each HCV-1a and -3a was investigated through the phylogenetic analysis. The results for NS5A phylogenetic analysis of HCV-1a and -3a have been shown in **Figures 1A,B**, respectively. The results for NS5B phylogenetic analysis of HCV-1a and -3a have been shown in **Figures 2A,B**, respectively. In all four phylogenetic trees, sequences of the same region and the same subtype were grouped closely, while considerable distance was found between those sequences and outgroups in each tree that confirm the accuracy of sequences from NS5A and NS5B regions of each HCV-1a and -3a patients. According to the phylogenetic trees, the HCV subtypes were the same as the HCV genotyping results included in the patients' clinical records. Briefly, the consistency between the HCV genotypes (1a and 3a) and the results of NS5A and NS5B phylogenetic analyses confirmed that no recombination event occurred between these regions.

DISCUSSION

Treatment of HCV has been developed substantially after approval of the first generation of DAAs by the United States Food and Drug Administration in 2011 (Ghany et al., 2019, 2020). These years, combinations of different DAAs have been used to treat HCV infection. They are highly efficacious and have fewer side effects with shorter treatment duration (usually 12 weeks), and are better tolerated than previous therapies (Hernandez et al., 2013; Kalaghatgi et al., 2016; Alavian and Sharafi, 2017; GENAFOR, 2019; Ghany et al., 2019, 2020). Although expectancy to achieve SVR even in patients with chronic HCV infection with failure to previous HCV antiviral therapy by using these new DAAs has been dramatically increased, RASs may emerge either before treatment with DAAs or following drug exposure (Fridell et al., 2011; Hernandez et al., 2013; Walker et al., 2015; Paolucci et al., 2017; Zeuzem et al., 2017; World Health Organization, 2017; Grandal et al., 2018; Sharafi et al., 2019). These drug resistance substitutions especially those to NS5A inhibitors lead to treatment failure so, they are considered a major challenge for HCV treatment and elimination (Fridell et al., 2011; Hernandez et al., 2013; Walker et al., 2015; Paolucci et al., 2017; Zeuzem et al., 2017; World Health Organization, 2017; Grandal et al., 2018; Sharafi et al., 2019).

Combinations of DCV/SOF and LDV/SOF are the most preferred regimens with above 95% cure rates and are strongly recommended by WHO (World Health Organization, 2017). Although NS5A inhibitors are effective in low concentrations, there are concerns about the existence of a low barrier to the selection of resistance mutations that decrease susceptibility to NS5A inhibitors. Generally, the NS5A RASs could result in >2 resistance fold-change, and some of them which cause >100 resistance fold-change are known as RASs > 100X (Issur and Götte, 2014; Sharafi and Alavian, 2018). For instance, just a single nucleotide substitution in codon 93 can change its related amino acid (from Y to H, C, and N) in NS5A protein and leads to resistance to most of the NS5A inhibitors (Issur and Götte, 2014; Nakamoto et al., 2014; Sharafi and Alavian, 2018). However, the viral genotype and subtype should be considered as an important factor to define the fold-change (Wang et al., 2013; Issur and Götte, 2014; Nakamoto et al., 2014; World Health Organization, 2017; Brandão et al., 2018; Hezode et al., 2018; Sharafi and Alavian, 2018; Ghany et al., 2019, 2020). In this study, baseline RASs were investigated in 72 HCV-1a patients, and 63 HCV-3a patients, which are the most prevalent HCV genotypes in Iran and were used to be known as "difficult-to-treat genotypes" (Bittar et al., 2010; Fridell et al., 2011; Hernandez et al., 2013; Walker et al., 2015; de Rueda et al., 2017; Paolucci et al., 2017; Zeuzem et al., 2017; Grandal et al., 2018). In our study, all of the patients were DAA-naïve. However, 44.4% of patients with HCV-1a infection, and 30.2% of patients with HCV-3a infection had a history of treatment with IFN-based regimens. There was a total of 41 (56.9%) and 40 (63.5%) cirrhotic patients infected with HCV-1a and HCV-3a, respectively. In cirrhotic patients with HCV-1a infection, 12 (16.7%) had a previous history of IFN-based treatment while it was 17 (27%) in cirrhotic patients with HCV-3a infection. Although the high rate of SVR (>90%) has led to a reduction in the number of studies on baseline resistance before initiating DAA-based treatments, many investigations of baseline RASs showed the impact of NS5A RASs on the virologic outcome for NS5A inhibitor-containing regimens such as DCV/SOF, and LDV/SOF (Wang et al., 2013; Brandão et al., 2018; Sharafi and Alavian, 2018; Sharafi et al., 2019). The European Association for the Study of Liver Diseases (EASL) recommended that the identification of baseline RASs to DAAs including LDV/SOF and DCV/SOF for HCV genotypes such as 1a could be useful to decide on extending treatment duration or if needed adding RBV (Pawlotsky et al., 2020). The

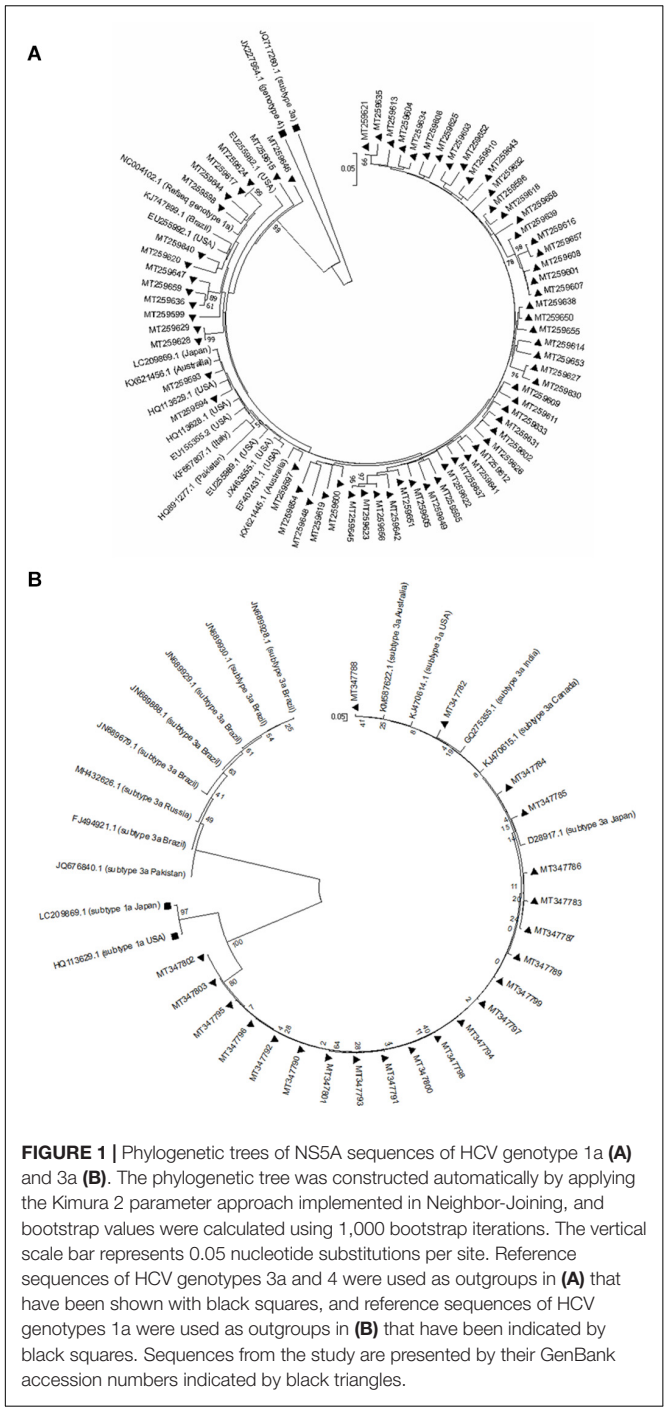

FIGURE 1 | Phylogenetic trees of NS5A sequences of HCV genotype 1a **(A)** and 3a **(B)**. The phylogenetic tree was constructed automatically by applying the Kimura 2 parameter approach implemented in Neighbor-Joining, and bootstrap values were calculated using 1,000 bootstrap iterations. The vertical scale bar represents 0.05 nucleotide substitutions per site. Reference sequences of HCV genotypes 3a and 4 were used as outgroups in **(A)** that have been shown with black squares, and reference sequences of HCV genotypes 1a were used as outgroups in **(B)** that have been indicated by black squares. Sequences from the study are presented by their GenBank accession numbers indicated by black triangles.

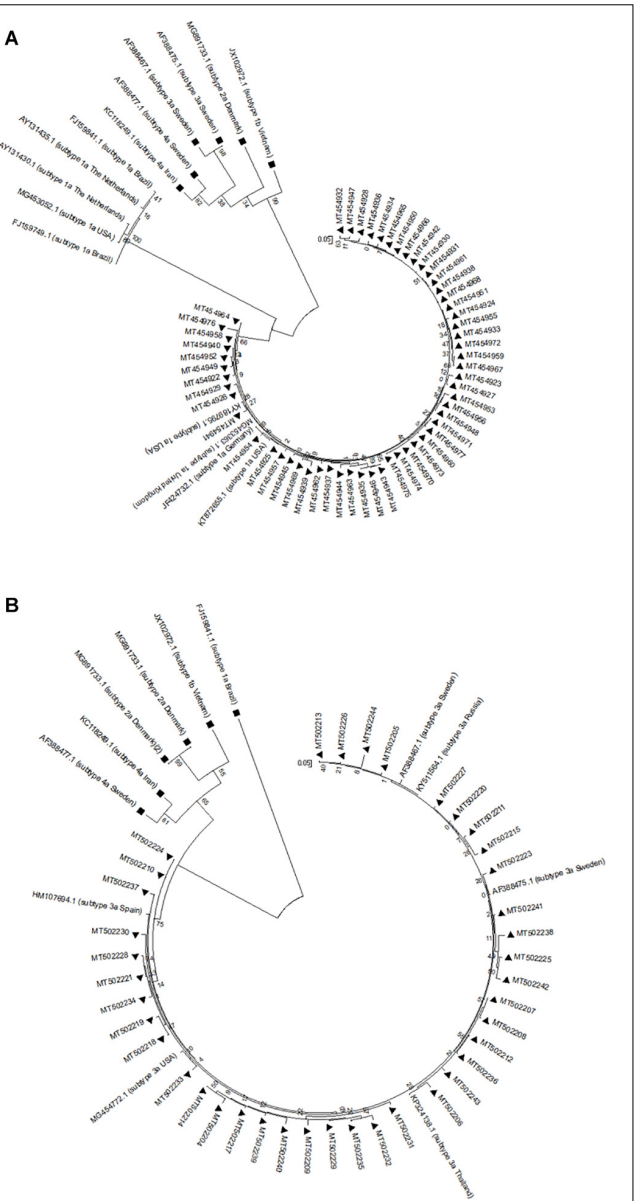

FIGURE 2 | Phylogenetic trees of NS5B sequences of HCV genotype 1a **(A)** and 3a **(B)**. The phylogenetic tree was constructed automatically by applying the Kimura 2 parameter approach implemented in Neighbor-Joining, and bootstrap values were calculated using 1,000 bootstrap iterations. The vertical scale bar represents 0.05 nucleotide substitutions per site. Reference sequences of HCV genotypes 1b, 2a, 3a, and 4a were used as outgroups in **(A)** that have been shown with black squares, and reference sequences of HCV genotypes 1a, 1b, 2a, and 4a were used as outgroups in **(B)** that have been indicated by black squares. Sequences from this study are presented by their GenBank accession numbers indicated by black triangles.

techniques which have been used for sequencings such as Sanger sequencing or deep sequencing are an important factor that affects the detection rate of RASs (Ghany et al., 2019, 2020; Pawlotsky et al., 2020). In this study, we could not perform/order deep sequencing because of budget limitations that we were challenged with during this project. Despite that limitation, both NS5A and NS5B regions of HCV-1a and -3a were successfully amplified and Sanger sequencing of 270 PCR products resulted in reliable and accurate detection of amino acid substitutions/RASs as were presented in details in the "Results" section. Many investigators suggested that both methods can be considered equivalent if a $\geq$15% cut point is used for determination of RASs by NGS due to recent studies that have shown the results of NGS at a 1% level of sensitivity often lead to the identification of additional RASs that are not associated with clinical failure

(Zeuzem et al., 2017; Perales et al., 2018; GENAFOR, 2019; Ghany et al., 2019, 2020; Chen et al., 2020). Herein, all the amplified NS5A and NS5B fragments were sequenced using the Sanger method, and NS5A amino acid substitutions/RASs were identified in 26.4%, and 15.9% of patients infected with HCV-1a and -3a, respectively. For instance, in one study, NS5A RASs were detected in 28.4% of Portuguese patients (Brandão et al., 2018; Hezode et al., 2018). Besides, several other studies showed that the overall proportion of NS5A RASs in baseline varies between <10% to >50% (Issur and Götte, 2014). Therefore, the detection rate of NS5A amino acid substitutions/RASs in this study is in line with those studies (Fridell et al., 2011; Hernandez et al., 2013; Paolucci et al., 2013, 2017; Dietz et al., 2015; de Rueda et al., 2017; Wyles and Luetkemeyer, 2017; Brandão et al., 2018; Grandal et al., 2018; Hezode et al., 2018; Ghany et al., 2019, 2020; Sharafi et al., 2019).

The genotype/subtype of HCV is one of the most effective factors, which should be considered in studying baseline RASs (Fridell et al., 2011; Paolucci et al., 2013, 2017; Wyles and Luetkemeyer, 2017). In addition, according to several studies the prevalence of NS5A amino acid substitutions/RASs in patients with HCV-1a was found at a higher rate than in those with HCV-3a infection, and in this study, the prevalence of NS5A amino acid substitutions and also the RASs > 100X in patients infected with HCV-1a was higher than patients with HCV-3a infection (26.4% vs. 15.9% and 8.3% vs. 4.8%) which was concordant with other studies (Sarrazin and Zeuzem, 2010; Nakamoto et al., 2014; Walker et al., 2015; Bagaglio et al., 2016; Kliemann et al., 2016; Pawlotsky, 2016; Malta et al., 2017; McPhee et al., 2017; Zeuzem et al., 2017; Grandal et al., 2018; Lu et al., 2019). Cirrhosis is considered a complicated condition in patients with chronic HCV infection as it indicates persistent viral infection accompanied by virus replication for a long time that could result in viral fitness (Nakamoto et al., 2014; Sarrazin, 2016; Bartolini et al., 2017; Pérez et al., 2017; Welzel et al., 2017; Zeuzem et al., 2017; Smith et al., 2019). In this study, five clinically relevant RASs > 100X, out of six identified RASs > 100X in HCV-1a infected patients were detected in patients with cirrhosis and previous history of IFN-based treatment. In patients with HCV-3a infection, the amino acid substitutions/RASs have detected in 10 patients, seven were cirrhotic with a previous history of IFN-based treatment, and the RAS > 100X Y93H was identified in 2 patients with cirrhosis. Our findings are in accordance with other studies that amino acid substitutions/RASs and especially the clinically relevant RASs (RASs > 100X) could be identified more frequently in cirrhotic patients than in non-cirrhotic patients (Nakamoto et al., 2014; Walker et al., 2015; Sarrazin, 2016; Malta et al., 2017; Paolucci et al., 2017; Welzel et al., 2017; Zeuzem et al., 2017; Brandão et al., 2018; Ghany et al., 2019, 2020).

In patients with HCV-1a infection, substitutions in aa 58 were the most commonly identified substitutions (16.7%) while; substitutions in aa 30 were the most commonly detected amino acid changes in HCV-3a (9.5%) infection and this was consistent with other studies (Sarrazin and Zeuzem, 2010; Hernandez et al., 2013; de Rueda et al., 2017; Ghany et al., 2019, 2020; Smith et al., 2019). Here, we detected substitutions of aa 58 in 12 patients with HCV-1a infection. Although the impact of substitutions in aa 58 is not clear, H58D is known as a resistance substitution and some studies reported that H58P confers the resistance to DCV (Bagaglio et al., 2016; Pawlotsky, 2016; Bartolini et al., 2017; Malta et al., 2017; Lu et al., 2019). Several studies found that when substitution in aa 58 was the only detected RAS, a very low inhibitory effect on resistance to NS5A inhibitors would be expected (Nakamoto et al., 2014; Bagaglio et al., 2016; Pawlotsky, 2016; Bartolini et al., 2017; Malta et al., 2017; Lu et al., 2019). In contrast, a combination of substitutions in aa 58 with other certain NS5A RASs (e.g., L31M, Q30R) could increase the resistance to NS5A inhibitors in patients with HCV-1a infection (Nakamoto et al., 2014; Bagaglio et al., 2016; Bartolini et al., 2017; Malta et al., 2017; Pérez et al., 2017; Welzel et al., 2017). In this investigation, only one combination of substitution in aa 58 with substitution in aa 28 was identified in HCV-1a; (M28V+H58S/T/C). A similar combination of substitutions in aa 58 was found in two patients with HCV-3a infection: (A30T/K+P58S/T/C) and (L31F+P58S/T/C+Y93H).

Multiple amino acid variations in specific locations such as aa 58 and aa 30 in NS5A of HCV-1a and HCV-3a have been identified with different frequencies in many studies (Bittar et al., 2010; Paolucci et al., 2013; Issur and Götte, 2014; Bagaglio et al., 2016; Kliemann et al., 2016; Pawlotsky, 2016; Malta et al., 2017; McPhee et al., 2017; Lu et al., 2019). Herein, the multiple amino acid changes were found in 4.2% of HCV-1a patients in aa 58 (H58S/T/C), and in 6.3% of HCV-3a sequences in aa 30 (A30T/K), and aa 58 (P58S/T/C). The presence of NS5A RASs at baseline has been known effective on the outcome of DAA-based treatments while each RAS has a specific impact on its DAA target. For example, Q30R/H and M28T have been considered as low-level resistance substitutions for HCV-1a by some researchers (Nakamoto et al., 2014; Bagaglio et al., 2016; Pawlotsky, 2016; Bartolini et al., 2017; Malta et al., 2017). In this study, Q30R/H and M28T/V/I substitutions were identified in 4.2% and 11.1% of patients with HCV-1a infection, respectively.

Generally, the presence of pre-treatment NS5A RASs has a negative effect on achieving SVR in comparison to those without baseline NS5A RASs (Bagaglio et al., 2016; Welzel et al., 2017). In one study, SVR in patients with baseline NS5A RASs has been estimated at 93.5%, while it was 98.4% in patients without baseline NS5A RASs (Sarrazin et al., 2016). Moreover, an increase in SVR failure could be observed when the NS5A RASs > 100-fold resistance exist (Bagaglio et al., 2016; de Rueda et al., 2017; Welzel et al., 2017; Sharafi and Alavian, 2018; Smith et al., 2019). In our study, the clinically significant NS5A RASs > 100X of genotype 1a were detected in 8.3% of patients (Q30R/H:3 patients, L31M:1 patient, and M28T:2 patients), without any NS5A RASs in aa 93 while, 3.2% of patients with HCV-3a infection harbored this substitution (Y93H:2 patients). Several studies reported that only Y93H should be considered as clinically significant NS5A RAS > 100X in HCV-3a (Nakamoto et al., 2014; Bagaglio et al., 2016; Kalaghatgi et al., 2016; Paolucci et al., 2017; Dietz et al., 2018; Perales et al., 2018; Smith et al., 2019). Herein, one of the identified substitution Y93H was paired with RASs in aa 31, and aa 58 (L31F+P58S/T/C+Y93H). According to some studies, RASs in aa 93 (Y93H) is considered a > 100 resistance

fold-change NS5A RAS, especially when it appeared as paired substitution with RASs in aa 31 in HCV-3a (Wang et al., 2013; Nakamoto et al., 2014; Paolucci et al., 2017; Dietz et al., 2018; Smith et al., 2019).

Although in some reports the rate of naturally-occurring > 100 resistance fold-change NS5A RASs in treatment-naïve patients has been estimated as less than 5%, other studies have shown the contrary results as in one study there were 14.6% NS5A RASs > 100X for HCV-1a, and 22.6% for HCV-3a (Sarrazin et al., 2016; Malta et al., 2017; Welzel et al., 2017; Zeuzem et al., 2017; Sharafi and Alavian, 2018). In another study, the rate of baseline NS5A RASs > 100X was reported as 9.9% (Hezode et al., 2018). This controversy might result from variation in the number of studied patients, genetic factors such as *IFNL3/4* polymorphisms, the difference in the prevalence of HCV genotypes in the area where the study was done, and the viral load (Nakamoto et al., 2014; Bagaglio et al., 2016; Pawlotsky, 2016; Bartolini et al., 2017; Malta et al., 2017).

Liver status such as cirrhosis is another important predictor of treatment response (Pawlotsky, 2016; Bartolini et al., 2017; Malta et al., 2017; Pérez et al., 2017; Welzel et al., 2017; Smith et al., 2019). Herein, 56.9% of patients with HCV-1a infection, and 63.9% of patients with HCV-3a infection were diagnosed with cirrhosis. In patients infected with HCV-1a, the clinically relevant substitutions (RASs > 100X) were detected in 6/72 patients, from them five substitutions were identified in cirrhosis: M28T in one patient, Q30R/H in 3 patients, and L31M in 1 patient. While one of the RAS > 100X (M28T) was detected in one non-cirrhotic HCV-1a patient. The clinically important substitution (RAS > 100X) Y93H was detected in 3.2% (2/63) of patients with HCV-3a infection, both diagnosed with cirrhosis. In addition, the identified two paired NS5A amino acid substitutions/RASs (L31F+P58S/T/C+Y93H) and (A30T/K+P58S/T/C) were detected each in one cirrhotic HCV-3 patient. In both HCV-1a and HCV-3a infected patients in this study, the high prevalence of RASs > 100X in cirrhotic patients in comparison with non-cirrhotic patients has been reported in many studies, and here our results conformed with them (Sulkowski et al., 2014; Bagaglio et al., 2016; Sarrazin et al., 2016; Bartolini et al., 2017; Malta et al., 2017; McPhee et al., 2017; Paolucci et al., 2017; Lu et al., 2019; Sayan et al., 2020).

Some investigators proposed that the SVR rate in patients infected with HCV-1a or -3a, could be strongly affected by the presence of RASs in companion with a history of previous IFN-based treatment, and this negative effect was irrespective of the cirrhosis status (Walker et al., 2015; Sarrazin, 2016; Sarrazin et al., 2016; Welzel et al., 2017; Zeuzem et al., 2017; Dietz et al., 2018; Esposito et al., 2019; Smith et al., 2019). However, many other researchers believed that achieving SVR is a multifactorial event that could be affected by host factors such as liver conditions (cirrhosis), the previous history of IFN-based treatment, the existence of some *IFNL3/4* polymorphisms, and consuming alcohol. Also, viral factors including infection with some genotypes (1a and 3), the presence of RASs, HIV/HCV co-infection, and viral load could affect the treatment response in a negative way. However, treatment strategy including its antiviral drugs and treatment duration could overcome other factors causing poor prognosis (Sarrazin and Zeuzem, 2010; Sarrazin, 2016; Sarrazin et al., 2016; Bartolini et al., 2017; Paolucci et al., 2017; Pérez et al., 2017; Zeuzem et al., 2017; Aldunate et al., 2018; Dietz et al., 2018; Esposito et al., 2019; Sayan et al., 2020). Sofosbuvir (PSI-7977 and GS-7977), is a very potent viral RdRp (NS5B) inhibitor with a pan-genotypic effect on HCV, without issues of viral resistance (Murakami et al., 2010; Sofia et al., 2010; Poordad and Dieterich, 2012; Stedman, 2014; Tong et al., 2014; Costantino et al., 2015; Gallego et al., 2016; Gane et al., 2017; Cory et al., 2018; Sorbo et al., 2018). It is a prodrug of 2′-deoxy-2′-fluoro-2′-C-methyluridine monophosphate which should be changed to its activated form (2‘- C-methyl group) in the liver and acts as a chain terminator (Murakami et al., 2010; Sofia et al., 2010; Poordad and Dieterich, 2012; Lawitz et al., 2014; Stedman, 2014; Tong et al., 2014; Costantino et al., 2015; Gallego et al., 2016; Bagaglio et al., 2018; Cory et al., 2018; Sorbo et al., 2018). Since its discovery, SOF has dramatically changed the treatment outcome of patients with HCV infection who were infected with hard-to-treat genotypes (1 and 3), cirrhosis, and previous history of treatment (Sofia et al., 2010; Lawitz et al., 2014; Stedman, 2014; Gallego et al., 2016; Gane et al., 2017). Despite being mutation prone which results in the selection of most fitted variants, especially when virus replication is under pressure due to anti-HCV drugs, SOF has shown the exception results as SOF-resistant variants may not be selected or even selectable (Sofia et al., 2010; Poordad and Dieterich, 2012; Lawitz et al., 2014; Tong et al., 2014; Costantino et al., 2015; Gallego et al., 2016; Gane et al., 2017; McPhee et al., 2017; Bagaglio et al., 2018; Sorbo et al., 2018).

CONCLUSION

In this study, NS5A RASs were identified at baseline in Iranian patients with HCV-1a and -3a infection. The prevalence of baseline RASs was in line with similar reports from other countries. Recently, new drugs with a high genetic barrier to resistance have been introduced to the pipeline as a combination of pan-genotypic regimens. As a result, promising anti-HCV treatment and even, elimination of HCV infection could be achieved.

DATA AVAILABILITY STATEMENT

The dataset is available upon request from PR (pooneh5376@yahoo.com). The submitted sequences in GenBank are available as a **Supplementary Material**.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee of the Pasteur Institute of Iran (no: IR.PII.REC.1395.81). The patients/participants provided written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

PR designed the research. PR, HS, GB, and SA analyzed results. PR and HS drafted the manuscript. FS, NA, and VI contributed to the clinical data collection and laboratory detection methods. All authors read and approved the submitted version.

FUNDING

This work was supported by a grant from the Pasteur Institute of Iran (Grant Number 943).

ACKNOWLEDGMENTS

We wish to thank the colleagues of the Department of Hepatitis and AIDS, Pasteur Institute of Iran, Tehran, Iran. We express our sincere gratitude to Dr. Shahin Merat for providing the study samples.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.617375/full#supplementary-material

REFERENCES

Alavian, S. M., and Sharafi, H. (2017). Update on recommendations for the clinical management of hepatitis C in Iran 2017. *Hepat. Mon.* 17, e63956.

Aldunate, F., Echeverría, N., Chiodi, D., López, P., Sánchez-Cicerón, A., Fajardo, A., et al. (2018). Pretreatment hepatitis C virus NS5A/NS5B resistance-associated substitutions in genotype 1 Uruguayan infected patients. *Dis. Mark.* 2018, 2514901.

Bagaglio, S., Andolina, A., Merli, M., Uberti-Foppa, C., and Morsica, G. (2016). Frequency of natural resistance within NS5a replication complex domain in hepatitis C genotypes 1a, 1b: possible implication of subtype-specific resistance selection in multiple direct acting antivirals drugs combination treatment. *Viruses* 8, 91. doi: 10.3390/v8040091

Bagaglio, S., Uberti-Foppa, C., Olgiati, A., Messina, E., Hasson, H., Ferri, C., et al. (2018). Natural polymorphisms in the resistance associated sites of HCV-G1 NS5B domain and correlation with geographic origin of HCV isolates. *Virol. J.* 15, 1–7.

Bartolini, B., Giombini, E., Taibi, C., Lionetti, R., Montalbano, M., Visco-Comandini, U., et al. (2017). Characterization of naturally occurring NS5A and NS5B polymorphisms in patients infected with HCV genotype 3a treated with direct-acting antiviral agents. *Viruses* 9, 212. doi: 10.3390/v9080212

Bittar, C., Jardim, A. C. G., Yamasaki, L. H., De Queiróz, A. T., Carareto, C. M., Pinho, J. R. R., et al. (2010). Genetic diversity of NS5A protein from hepatitis C virus genotype 3a and its relationship to therapy response. *BMC Infect. Dis.* 10:36. doi: 10.1186/1471-2334-10-36

Blach, S., Zeuzem, S., Manns, M., Altraif, I., Duberg, A.-S., Muljono, D. H., et al. (2017). Global prevalence and genotype distribution of hepatitis C virus infection in 2015: a modelling study. *Lancet Gastroenterol. Hepatol.* 2, 161–176.

Brandão, R., Marcelino, R., Gonçalves, F., Diogo, I., Carvalho, A., Cabanas, J., et al. (2018). Characterization of NS5A and NS5B resistance-associated substitutions from genotype 1 hepatitis C virus infected patients in a Portuguese cohort. *Viruses* 10, 223. doi: 10.3390/v10050223

Chen, Q., Perales, C., Soria, M. E., García-Cehic, D., Gregori, J., Rodríguez-Frías, F., et al. (2020). Deep-sequencing reveals broad subtype-specific HCV resistance mutations associated with treatment failure. *Antiviral Res.* 174, 104694.

Cory, T. J., Mu, Y., Gong, Y., Kodidela, S., and Kumar, S. (2018). Sofosbuvir+ velpatasvir+ voxilaprevir for the treatment of hepatitis C infection. *Expert Opin. Pharmacother.* 19, 749–757. doi: 10.1080/14656566.2018.1459567

Costantino, A., Spada, E., Equestre, M., Bruni, R., Tritarelli, E., Coppola, N., et al. (2015). Naturally occurring mutations associated with resistance to HCV NS5B polymerase and NS3 protease inhibitors in treatment-naïve patients with chronic hepatitis C. *Virol. J.* 12, 186.

de Rueda, P. M., Rodríguez, J. M. F., Pérez, R. Q., Medina, A. G., Álvarez, A. B. M., Ruíz, J. C., et al. (2017). Hepatitis C virus NS5A region mutation in chronic hepatitis C genotype 1 patients who are non-responders to two or more treatments and its relationship with response to a new treatment. *World J. Gastroenterol.* 23, 4538. doi: 10.3748/wjg.v23.i25.4538

Dietz, J., Susser, S., Berkowski, C., Perner, D., Zeuzem, S., and Sarrazin, C. (2015). Consideration of viral resistance for optimization of direct antiviral therapy of hepatitis C virus genotype 1-infected patients. *PLoS One* 10:e0134395. doi: 10.1371/journal.pone.0134395

Dietz, J., Susser, S., Vermehren, J., Peiffer, K.-H., Grammatikos, G., Berger, A., et al. (2018). Patterns of resistance-associated substitutions in patients with chronic HCV infection following treatment with direct-acting antivirals. *Gastroenterology* 154, 976–988.e974.

Esposito, I., Marciano, S., Haddad, L., Galdame, O., Franco, A., Gadano, A., et al. (2019). Prevalence and factors related to natural resistance-associated substitutions to direct-acting antivirals in patients with genotype 1 hepatitis C virus infection. *Viruses* 11, 3. doi: 10.3390/v11010003

Fridell, R. A., Wang, C., Sun, J. H., O'boyle, D. R., Nower, P., Valera, L., et al. (2011). Genotypic and phenotypic analysis of variants resistant to hepatitis C virus nonstructural protein 5A replication complex inhibitor BMS−790052 in humans: in vitro and in vivo correlations. *Hepatology* 54, 1924–1935. doi: 10.1002/hep.24594

Gallego, I., Sheldon, J., Moreno, E., Gregori, J., Quer, J., Esteban, J. I., et al. (2016). Barrier-independent, fitness-associated differences in sofosbuvir efficacy against hepatitis C virus. *Antimicrob. Agents Chemother.* 60, 3786–3793. doi: 10.1128/aac.00581-16

Gane, E. J., Metivier, S., Nahass, R., Ryan, M., Stedman, C. A., Svarovskaia, E. S., et al. (2017). The emergence of NS5B resistance associated substitution S282T after sofosbuvir−based treatment. *Hepatol. Commun.* 1, 538–549. doi: 10.1002/hep4.1060

GENAFOR (2019). *HCV*. Available online at: http://hcv.geno2pheno.org/index.php (accessed September 20, 2019)

Ghany, M. G., Marks, K. M., Morgan, T. R., Wyles, D. L., Aronsohn, A. I., Bhattacharya, D., et al. (2019). Hepatitis C guidance 2019 update: AASLD-IDSA recommendations for testing, managing, and treating hepatitis C virus infection. *Hepatology* 71, 686–721. doi: 10.1002/hep.31060

Ghany, M. G., Morgan, T. R., and Aasld-Idsa Hepatitis C Guidance Panel. (2020). Hepatitis C guidance 2019 update: American association for the study of liver diseases–infectious diseases society of america recommendations for testing, managing, and treating Hepatitis C virus infection. *Hepatology* 71, 686–721. doi: 10.1002/hep.31060

Gottwein, J. M., Scheel, T. K., Jensen, T. B., Lademann, J. B., Prentoe, J. C., Knudsen, M. L., et al. (2009). Development and characterization of hepatitis C virus genotype 1−7 cell culture systems: role of CD81 and scavenger receptor class B type I and effect of antiviral drugs. *Hepatology* 49, 364–377. doi: 10.1002/hep.22673

Grandal, M., Pernas, B., Tabernilla, A., Mariño, A., Álvarez, H., Valcarce, N., et al. (2018). Prevalence of NS5A resistance associated substitutions in patients with hepatitis C virus genotypes 1a and 3: Impact on current therapeutic strategies. *J. Med. Virol.* 90, 1094–1098. doi: 10.1002/jmv.25048

Hernandez, D., Zhou, N., Ueland, J., Monikowski, A., and Mcphee, F. (2013). Natural prevalence of NS5A polymorphisms in subjects infected with hepatitis C virus genotype 3 and their effects on the antiviral activity of NS5A inhibitors. *J. Clin. Virol.* 57, 13–18. doi: 10.1016/j.jcv.2012.12.020

Hezode, C., Reau, N., Svarovskaia, E. S., Doehle, B. P., Shanmugam, R., Dvory-Sobol, H., et al. (2018). Resistance analysis in patients with genotype 1–6 HCV

infection treated with sofosbuvir/velpatasvir in the phase III studies. *J. Hepatol.* 68, 895–903. doi: 10.1016/j.jhep.2017.11.032

Issur, M., and Götte, M. (2014). Resistance patterns associated with HCV NS5A inhibitors provide limited insight into drug binding. *Viruses* 6, 4227–4241. doi: 10.3390/v6114227

Kalaghatgi, P., Sikorski, A. M., Knops, E., Rupp, D., Sierra, S., Heger, E., et al. (2016). Geno2pheno [HCV]–a web-based interpretation system to support hepatitis C treatment decisions in the era of direct-acting antiviral agents. *PLoS One* 11:e0155869. doi: 10.1371/journal.pone.0155869

Kau, A., Vermehren, J., and Sarrazin, C. (2008). Treatment predictors of a sustained virologic response in hepatitis B and C. *J. Hepatol.* 49, 634–651. doi: 10.1016/j.jhep.2008.07.013

Khodabandehloo, M., and Roshani, D. (2014). Prevalence of hepatitis C virus genotypes in Iranian patients: a systematic review and meta-analysis. *Hepat. Monthly* 14, e22915.

Kliemann, D. A., Tovo, C. V., Da Veiga, A. B. G., De Mattos, A. A., and Wood, C. (2016). Polymorphisms and resistance mutations of hepatitis C virus on sequences in the European hepatitis C virus database. *World J. Gastroenterol.* 22, 8910. doi: 10.3748/wjg.v22.i40.8910

Lawitz, E., Poordad, F. F., Pang, P. S., Hyland, R. H., Ding, X., Mo, H., et al. (2014). Sofosbuvir and ledipasvir fixed-dose combination with and without ribavirin in treatment-naive and previously treated patients with genotype 1 hepatitis C virus infection (LONESTAR): an open-label, randomised, phase 2 trial. *Lancet* 383, 515–523. doi: 10.1016/s0140-6736(13)62121-2

Lu, J., Feng, Y., Chen, L., Zeng, Z., Liu, X., Cai, W., et al. (2019). Subtype-specific prevalence of hepatitis C virus NS5A resistance associated substitutions in Mainland China. *Front. Microbiol.* 10:535. doi: 10.3389/fmicb.2019.00535

Mahmud, S., Akbarzadeh, V., and Abu-Raddad, L. J. (2018). The epidemiology of hepatitis C virus in Iran: systematic review and meta-analyses. *Sci. Rep.* 8, 1–25.

Malta, F., Gaspareto, K. V., Lisboa-Neto, G., Carrilho, F. J., Mendes-Correa, M. C., and Pinho, J. R. R. (2017). Prevalence of naturally occurring NS5A resistance-associated substitutions in patients infected with hepatitis C virus subtype 1a, 1b, and 3a, co-infected or not with HIV in Brazil. *BMC Infect. Dis.* 17:716. doi: 10.1186/s12879-017-2817-7

McPhee, F., Hernandez, D., and Zhou, N. (2017). Effect of minor populations of NS5A and NS5B resistance-associated variants on HCV genotype-3 response to daclatasvir plus sofosbuvir, with or without ribavirin. *Antivir. Ther.* 22, 237–246. doi: 10.3851/imp3120

Murakami, E., Tolstykh, T., Bao, H., Niu, C., Steuer, H. M. M., Bao, D., et al. (2010). Mechanism of activation of PSI-7851 and its diastereoisomer PSI-7977. *J. Biol. Chem.* 285, 34337–34347. doi: 10.1074/jbc.m110.161802

Murphy, D. G., Sablon, E., Chamberland, J., Fournier, E., Dandavino, R., and Tremblay, C. L. (2015). Hepatitis C virus genotype 7, a new genotype originating from central Africa. *J. Clin. Microbiol.* 53, 967–972. doi: 10.1128/jcm.02831-14

Murphy, D. G., Willems, B., Deschênes, M., Hilzenrat, N., Mousseau, R., and Sabbah, S. (2007). Use of sequence analysis of the NS5B region for routine genotyping of hepatitis C virus with reference to C/E1 and 5′ untranslated region sequences. *J. Clin. Microbiol.* 45, 1102–1112. doi: 10.1128/jcm.02366-06

Nakamoto, S., Kanda, T., Wu, S., Shirasawa, H., and Yokosuka, O. (2014). Hepatitis C virus NS5A inhibitors and drug resistance mutations. *World J. Gastroenterol.* 20, 2902. doi: 10.3748/wjg.v20.i11.2902

Nakano, T., Lau, G. M., Lau, G. M., Sugiyama, M., and Mizokami, M. (2012). An updated analysis of hepatitis C virus genotypes and subtypes based on the complete coding region. *Liver Int.* 32, 339–345. doi: 10.1111/j.1478-3231.2011.02684.x

Paolucci, S., Fiorina, L., Mariani, B., Gulminetti, R., Novati, S., Barbarini, G., et al. (2013). Naturally occurring resistance mutations to inhibitors of HCV NS5A region and NS5B polymerase in DAA treatment-naive patients. *Virol. J.* 10, 355. doi: 10.1186/1743-422x-10-355

Paolucci, S., Premoli, M., Novati, S., Gulminetti, R., Maserati, R., Barbarini, G., et al. (2017). Baseline and breakthrough resistance mutations in HCV patients failing DAAs. *Sci. Rep.* 7, 1–9.

Pawlotsky, J.-M. (2016). Hepatitis C virus resistance to direct-acting antiviral drugs in interferon-free regimens. *Gastroenterology* 151, 70–86. doi: 10.1053/j.gastro.2016.04.003

Pawlotsky, J.-M., Negro, F., Aghemo, A., Berenguer, M., Dalgard, O., Dusheiko, G., et al. (2020). EASL recommendations on treatment of hepatitis C: final update of the series?. *J. Hepatol.* 73, 1170–1218. doi: 10.1016/j.jhep.2020.08.018

Perales, C., Chen, Q., Soria, M. E., Gregori, J., Garcia-Cehic, D., Nieto-Aponte, L., et al. (2018). Baseline hepatitis C virus resistance-associated substitutions present at frequencies lower than 15% may be clinically significant. *Infect. Drug Resist.* 11, 2207. doi: 10.2147/idr.s172226

Pérez, A. B., Chueca, N., and García, F. (2017). Resistance testing for the treatment of chronic hepatitis C with direct acting antivirals: when and for how long? *Germs* 7, 40. doi: 10.18683/germs.2017.1107

Poordad, F., and Dieterich, D. (2012). Treating hepatitis C: current standard of care and emerging direct−acting antiviral agents. *J. Viral Hepat.* 19, 449–464. doi: 10.1111/j.1365-2893.2012.01617.x

Sarrazin, C. (2016). The importance of resistance to direct antiviral drugs in HCV infection in clinical practice. *J. Hepatol.* 64, 486–504. doi: 10.1016/j.jhep.2015.09.011

Sarrazin, C., Dvory-Sobol, H., Svarovskaia, E. S., Doehle, B. P., Pang, P. S., Chuang, S.-M., et al. (2016). Prevalence of resistance-associated substitutions in HCV NS5A, NS5B, or NS3 and outcomes of treatment with ledipasvir and sofosbuvir. *Gastroenterology* 151, 501–512.e501.

Sarrazin, C., and Zeuzem, S. (2010). Resistance to direct antiviral agents in patients with hepatitis C virus infection. *Gastroenterology* 138, 447–462. doi: 10.1053/j.gastro.2009.11.055

Sayan, M., Yıldırım, F. S., Akhan, S., Yıldırım, A. A., Şirin, G., Cabalak, M., et al. (2020). NS5A resistance-associated substitutions in chronic Hepatitis C patients with direct acting antiviral treatment failure in Turkey. *Int. J. Infect. Dis.* 95, 84–89. doi: 10.1016/j.ijid.2020.03.061

Sharafi, H., and Alavian, S. M. (2018). Hepatitis C resistance to NS5A inhibitors: is it going to be a problem? *World J. Hepatol.* 10, 543. doi: 10.4254/wjh.v10.i9.543

Sharafi, H., Maleki, S., and Alavian, S. M. (2019). Prevalence of hepatitis C virus NS5A resistance-associated substitutions in chronic infection with genotype 1: a pooled analysis based on deposited sequences in GenBank. *Virus Res.* 259, 54–61. doi: 10.1016/j.virusres.2018.10.014

Smith, D., Magri, A., Bonsall, D., Ip, C. L., Trebes, A., Brown, A., et al. (2019). Resistance analysis of genotype 3 hepatitis C virus indicates subtypes inherently resistant to nonstructural protein 5A inhibitors. *Hepatology* 69, 1861–1872. doi: 10.1002/hep.29837

Smith, D. B., Bukh, J., Kuiken, C., Muerhoff, A. S., Rice, C. M., Stapleton, J. T., et al. (2014). Expanded classification of hepatitis C virus into 7 genotypes and 67 subtypes: updated criteria and genotype assignment web resource. *Hepatology* 59, 318–327. doi: 10.1002/hep.26744

Sofia, M. J., Bao, D., Chang, W., Du, J., Nagarathnam, D., Rachakonda, S., et al. (2010). Discovery of a β-d-2′-deoxy-2′-α-fluoro-2′-β-C-methyluridine nucleotide prodrug (PSI-7977) for the treatment of hepatitis C virus. *J. Med. Chem.* 53, 7202–7218. doi: 10.1021/jm100863x

Sorbo, M. C., Cento, V., Di Maio, V. C., Howe, A. Y., Garcia, F., Perno, C. F., et al. (2018). Hepatitis C virus drug resistance associated substitutions and their clinical relevance: update 2018. *Drug Resist. Updates* 37, 17–39. doi: 10.1016/j.drup.2018.01.004

Stedman, C. (2014). Sofosbuvir, a NS5B polymerase inhibitor in the treatment of hepatitis C: a review of its clinical potential. *Ther. Adv. Gastroenterol.* 7, 131–140. doi: 10.1177/1756283x13515825

Sulkowski, M. S., Gardiner, D. F., Rodriguez-Torres, M., Reddy, K. R., Hassanein, T., Jacobson, I., et al. (2014). Daclatasvir plus sofosbuvir for previously treated or untreated chronic HCV infection. *N. Engl. J. Med.* 370, 211–221.

Taherkhani, R., and Farshadpour, F. (2015). Epidemiology of hepatitis C virus in Iran. *World J. Gastroenterol.* 21, 10790. doi: 10.3748/wjg.v21.i38.10790

Tong, X., Le Pogam, S., Li, L., Haines, K., Piso, K., Baronas, V., et al. (2014). In vivo emergence of a novel mutant L159F/L320F in the NS5B polymerase confers low-level resistance to the HCV polymerase inhibitors mericitabine and sofosbuvir. *J. Infect. Dis.* 209, 668–675. doi: 10.1093/infdis/jit562

Walker, A., Siemann, H., Groten, S., Ross, R. S., Scherbaum, N., and Timm, J. (2015). Natural prevalence of resistance-associated variants in hepatitis C virus NS5A in genotype 3a-infected people who inject drugs in Germany. *J. Clin. Virol.* 70, 43–45. doi: 10.1016/j.jcv.2015.07.008

Wang, C., Sun, J.-H., O'boyle, D. R., Nower, P., Valera, L., Roberts, S., et al. (2013). Persistence of resistant variants in hepatitis C virus-infected patients treated with the NS5A replication complex inhibitor daclatasvir. *Antimicrob. Agents Chemother.* 57, 2054–2065. doi: 10.1128/aac.02494-12

Welzel, T. M., Bhardwaj, N., Hedskog, C., Chodavarapu, K., Camus, G., Mcnally, J., et al. (2017). Global epidemiology of HCV subtypes and resistance-associated

substitutions evaluated by sequencing-based subtype analyses. *J. Hepatol.* 67, 224–236. doi: 10.1016/j.jhep.2017.03.014

World Health Organization (2017). *Global Hepatitis Report 2017*. Geneva: WHO.

Wyles, D. L., and Luetkemeyer, A. F. (2017). Understanding hepatitis C virus drug resistance: clinical implications for current and future regimens. *Top. Antiviral Med.* 25, 103.

Zeuzem, S., Mizokami, M., Pianko, S., Mangia, A., Han, K.-H., Martin, R., et al. (2017). NS5A resistance-associated substitutions in patients with genotype 1 hepatitis C virus: prevalence and effect on treatment outcome. *J. Hepatol.* 66, 910–918. doi: 10.1016/j.jhep.2017.01.007

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 04 February 2021
doi: 10.3389/fmicb.2021.640325

Autochthonous and Travel Acquired Hepatitis E Virus in Australia

*Jacinta O'Keefe, Lilly Tracy, Lilly Yuen, Sara Bonanzinga, Xin Li, Brian Chong, Suellen Nicholson and Kathy Jackson**

Victorian Infectious Diseases Reference Laboratory, Royal Melbourne Hospital, Doherty Institute for Infection and Immunity, Melbourne, VIC, Australia

Background: Hepatitis E virus (HEV) is a common cause of acute viral hepatitis with significant morbidity and mortality, particularly in pregnant women. There are four major genotypes which can cause disease in humans. Genotypes 1 and 2 are usually associated with outbreaks and spread via facal/oral route or contaminated water. Genotypes 3 and 4 are zoonotic and usually associated with handling of pigs or consumption of contaminated pork. The strains circulating in Australia have never been characterized.

Rationale/Aims: The aims for this project are to identify the HEV genotypes found in Australia and link them to possible sources of transmission by phylogenetic analysis.

Materials and Methods: Between 2015 and 2020, 91 HEV isolates were sequenced and genotyped using an in-house PCR. Sixty-six of these were also sequenced by using the international HEVnet primers. Genotypes were determined using the BLASTn program. Relatedness to other strains in Australia was determined by phylogenetic analyses of the HEVnet sequences. Isolates were also stratified by state of origin, gender, age, predisposing factors and travel history (if known).

Results: Of the 91 HEV isolates sequenced, 55 (60.4%) were genotype 1. There were 34 (37.4%) genotype 3 strains and two genotype 4 (2.2%). At least 20 of the genotype 1 strains have been linked to travel in India, and another three with Pakistan. Five of the "Indian" strains were closely related and are suspected to have originated in Gujarat. Phylogenetic analysis also showed that 12 genotype 3 strains were genetically related and potentially acquired in/from New South Wales, Australia. The two genotype 4 strains may have originated in China.

Discussion: This is the first study to describe the HEV isolates identified in Australia. The results infer that HEV may be acquired during overseas travel as well as locally, presumably from consumption of pork or pork-related products. The phylogenetic analyses also reveal clusters of infection originating from India and Pakistan. This study provides some insight into the source and epidemiology of HEV infection in Australia which may be used to guide public health procedure and enable the implementation of measures to deal with potential outbreaks of infection.

Keywords: hepatitis E virus, Australia, phylogenetics analysis, autochthonous, zoonotic

OPEN ACCESS

Edited by:
Gkikas Magiorkinis, National and Kapodistrian University of Athens, Greece

Reviewed by:
Antonio Rivero-Juarez, Maimonides Biomedical Research Institute of Cordoba (IMIBIC), Spain
Sayed F. Abdelwahab, Minia University, Egypt

***Correspondence:**
Kathy Jackson
kathy.jackson@mh.org.au

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *11 December 2020*
Accepted: *18 January 2021*
Published: *04 February 2021*

Citation:
O'Keefe J, Tracy L, Yuen L, Bonanzinga S, Li X, Chong B, Nicholson S and Jackson K (2021) Autochthonous and Travel Acquired Hepatitis E Virus in Australia. Front. Microbiol. 12:640325. doi: 10.3389/fmicb.2021.640325

INTRODUCTION

Hepatitis E virus (HEV) is a major cause of acute hepatitis worldwide and is associated with significant morbidity and mortality in pregnant women and people with chronic liver disease or immunosuppression. Previously, HEV was thought to predominantly circulate in developing countries, causing an estimated 3.4 million symptomatic cases and 70,000 deaths annually (Rein et al., 2012), with cases in industrialized countries linked to travel in HEV endemic regions. In the last decade, mounting evidence has demonstrated that zoonotic HEV is endemic in many developed countries including France, England, Bulgaria and Japan, and it is estimated that there are as many as two million cases of autochthonous HEV in Europe annually (European Association for the Study of the Liver, 2018).

HEV is a small, non-enveloped, RNA virus with a positive-sense 7.2kb genome containing a 5′ untranslated region (UTR), three open-reading frames (ORF) and a polyadenylated 3′ UTR. ORF1 encodes a non-structural protein, ORF2 encodes a capsid protein of 660 amino acids and ORF3 encodes a small phosphoprotein, which enables release of virions from hepatocytes (Nagashima et al., 2011). As a member of the *Hepeviridae* family (Lu et al., 2006), HEV is classified as a species of *Orthohepevirus A* in the *Orthohepevirus* genus. Recent revisions have proposed eight genotypes and 36 subtypes that infect mammals, with genotypes 1-4 being a major cause of disease in humans (Smith et al., 2020).

The epidemiology, clinical presentation and reservoir of HEV can vary depending on the viral genotype. HEV-1 and -2 exclusively infect humans and are transmitted by the faecal-oral route via contaminated water in low-middle-income countries within Africa and Asia, typically causing symptomatic infection in patients aged 15–40 years (Kamar et al., 2014). Large outbreaks of HEV-1 infection are most common in South and East Asia and, whilst HEV-2 is less prevalent, it has been linked to outbreaks in Mexico, Nigeria and Namibia (Maila et al., 2004; Lu et al., 2006).

In contrast, HEV-3 and -4 are highly diverse, zoonotic viruses that have been isolated worldwide in humans and animals including wild boar, pigs and deer. Transmission in humans is primarily associated with the consumption of contaminated, undercooked pork or game (wild-caught) meat but has also been linked to shellfish consumption. In addition, transfusion-transmission has been reported for HEV-3 infection (Hewitt et al., 2014; Forgan-Smith and Macdonald, 2019). Whilst HEV-3 is thought to be asymptomatic in up to 98% of cases, it is associated with symptomatic disease in older men or people with liver damage (Kamar et al., 2014) and can cause chronic infection in immunosuppressed patients, which can lead to cirrhosis (Kamar et al., 2008a,b). HEV-3 has been reported in multiple countries on six continents (excluding Antarctica). HEV-4 has been found in human and swine populations in mostly Asian countries (Lu et al., 2006; Zhu et al., 2014) but has also been detected in France (Colson et al., 2012) and Switzerland (Fraga et al., 2018).

There is limited data about the genotype distribution of HEV in Australia or the incidence of autochthonous HEV. Hepatitis E has been a notifiable disease in Australia since 1999 and, over the last 10 years, approximately 30–60 cases of Hepatitis E infection have been reported to the Commonwealth Department of Health annually, with 250 total cases reported between 2015 and 2020 (Australian Government Department of Health, 2020). The HEV seroprevalence in Australian blood donors has been reported as 5.9% (Shrestha et al., 2014), which is comparable to New Zealand (9%) (Hewitt et al., 2018) and Scotland (6.1%) (Thom et al., 2018) and lower than that seen in South Western France (52%) and The Netherlands (30%). In Australia, most cases of Hepatitis E infection have previously been ascribed to travel to a HEV endemic region (Gunaratnam et al., 2014). However, locally acquired infection with HEV-3 does occur, with a previous report of HEV transmission by blood transfusion (Speers et al., 2015) and a small outbreak in 2013/4 linked to consumption of undercooked pork at a New South Wales (NSW) restaurant (Yapa et al., 2016). Phylogenetic analysis was used to genotype a single HEV-3 isolate in a 2016 study investigating the prevalence of HEV RNA in Australian blood donations (Shrestha et al., 2016). However, there has not been any comprehensive phylogenetic analysis of Australian HEV isolates. This study aims to characterize the isolates of HEV positive samples sent to the Victorian Infectious Disease Reference Laboratory (VIDRL) from laboratories across Australia within the last 5 years and link them to possible sources of transmission by comparing the genetic relatedness of the isolates.

MATERIALS AND METHODS

Patient Samples

VIDRL is a specialized laboratory that performs HEV molecular testing for most states and territories of Australia with the exception of Queensland (QLD). Serum samples may be sent specifically for primary or confirmatory HEV RNA testing, or referred from the VIDRL serology laboratory after a reactive IgG result is obtained. Ninety one HEV RNA positive samples tested at VIDRL between 2015 and 2020 were sequenced for genotype determination using an in-house PCR and Sanger sequencing as below. We were able to perform subsequent sequence analysis using the HEVNet primers on 66 of these and phylogenetic analysis. These 66 isolates (phyloanalysis population) were compared to demographic details and relevant clinical history of the patient. Ethics approval for this study was granted by the Melbourne Health Human Research Ethics Committee (HREC/62312/MH-2020).

Serological Testing

Anti-HEV IgG was tested using the HEV ELISA 4.0 kit, (MP Biomedicals Asia Pacific Pte Ltd, Singapore, formerly Genelabs Diagnostics Pte) as per manufacturer's instructions. The test results are expressed as the absorbance of the patient sample to the assay cut-off (s/co value). A sample with a s/co value greater than 1.0 is considered reactive and less than 1.0 is considered negative. As per VIDRL protocol, prior to 2018 serum/plasma samples with a s/co value of greater than 3 were automatically referred on for PCR testing, as these samples were considered

more likely to reflect current HEV infection (Zaaijer et al., 1993). This practice was amended in 2018 to include all samples with reactive IgG.

Molecular Testing

Serum or plasma samples for HEV PCR were stored at −70°C before assaying. RNA was extracted from 140 μL of serum with the QIAamp Viral RNA Mini Kit (QIAGEN, Hilden, Germany) as per manufacturer's instructions. Between 2015 and 2020, 909 samples were screened for HEV RNA. The qualitative RealStar HEV RT-PCR Kit (Altona Diagnostics, Hamburg, Germany) was used from 2015 to 2018 and the quantitative RealStar HEV RT-PCR Kit 2.0 from 2018 onward.

For 91 HEV RNA positive samples, genotypes were determined prospectively by sequencing a 191bp fragment of the ORF2 in an in-house PCR as described previously (Erker et al., 1999) and using the BLASTn program on the NCBI BLAST web server.

Sixty-six of the 91 HEV positive serum samples were assayed in a two-step, nested RT-PCR using the HEVnet primers shared by RIVM (Rijksinstituut voor Volksgezondheid en Milieu, Nederlands) (Mulder et al., 2019). A 493 nt amplicon to the HEVORF2 region was generated with Superscript III Reverse Transcriptase (Invitrogen, Waltham, MA, United States) and HotStarTaq DNA Polymerase (Qiagen, Hilden, Germany) as previously described (Bruni et al., 2018). Purified PCR product was sequenced using ABI-Prism 3730 Genetic Analyser (Applied Biosystems, Life Technologies, Ltd, Paisley, United Kingdom) with capillary electrophoresis performed by Micromon, Monash University. The 66 HEVNet sequences (phyloanalysis population) were also analyzed phylogenetically.

Genotype and Phylogenetic Analysis

Using the sequence analysis program SeqScape version 2.1.1 (Applied Biosystems, Life Technologies, Ltd, Paisley, United Kingdom), a consensus sequence was generated for all isolates by aligning the forward and reverse sequence strands of the amplicon of sample to the stored reference HEV genome sequences NC_001434 and AB248521 from Genbank. For phylogenetic analyses, HEVnet sequences obtained from the 66 patient samples (phyloanalysis group) were multi-sequence aligned with full-length HEV sequences downloaded from GenBank (Benson et al., 2014) using the MAFFT program (Katoh et al., 2002) implemented in the bioinformatics software platform Geneious v7.1.9[1]. The alignments were then manually trimmed to the same length of the amplicon sequences. Subtype confirmation and assessment of phylogenetic relationships between the HEV sequences were performed using Maximum Likelihood (ML) phylogenetic trees constructed in IQtree (Nguyen et al., 2015). The best-fit substitution model used for each sequence data set was determined using ModelFinder (Kalyaanamoorthy et al., 2017) based on Bayesian Information Criterion. Clade support was assessed using 1,000 pseudo-replicates generated with the UFBoot non-parametric bootstrap procedure (Hoang et al., 2018) implemented in IQ-TREE. The ML tree used to confirm HEV genotype of patient isolates determined by screening primers, was constructed using the GTR + F + R10 model by assessing their phylogenetic relationships with 594 reference HEV genome sequences from GenBank, including the proposed set of reference HEV sequences for each subtype (Smith et al., 2020).

Once the genotype was confirmed, a ML phylogenetic tree was generated separately for HEV genotypes 1, 3, and 4 with the sample isolates and four to five representative sequences for each subtype, using the TNe + I + G4 model for HEV-1, the TIM3e + I + G4 model for HEV-3 and the TIM2e + G4 model for HEV-4 as recommended by the ModelFinder (Kalyaanamoorthy et al., 2017) option on IQtree, with bootstrap replicates of 1000 using UFBoot (Hoang et al., 2018).

Sample isolates that were unable to be subtyped were uploaded to the HEVnet typing tool[2] for further analysis.

The nucleotide sequences of this study were deposited in GenBank (Accession number MW355681 – MW355746).

RESULTS

Laboratory Data

Between 2015 and 2020, a total of 909 serum or plasma samples were tested at VIDRL for HEV RNA with samples sent from laboratories across all states and territories except for QLD. Of these, 91 samples had detectable HEV RNA with 55 isolates identified as HEV-1 (60.4%); 34 (37.4%) as HEV-3, and 2 (2.2%) as HEV-4 based on the 191nt in-house PCR amplicon of ORF2.

Serology Data

Anti-HEV IgG serological data was also available for 34/91 HEV RNA positive samples (data not shown), with 24/34 (71%) having a s/co value greater than 3. Eight samples had a s/co value less than 3 including four with a s/co value below 2. Two additional samples did not have detectable anti-HEV IgG at the time that HEV RNA was detected including one patient who was immunosuppressed.

Genotyping and Phylogenetic Analyses

Sixty-six of the 91 HEV RNA positive samples were available for testing using the HEVnet protocol while 25 samples had insufficient serum for testing. **Table 1** shows the demographic and clinical characteristics of these 66 cases (phyloanalysis population) including average patient age, travel history and genotype distribution with 42 (63.6%) isolates identified as HEV-1, 22 (33.3%) as HEV-3 and 2 as HEV-4 (3.0%). This is consistent with the genotype distribution of the 91 HEV RNA positive samples tested with the in-house protocol at VIDRL.

HEV Genotype 1

Figure 1 shows the result of the phylogenetic analysis of 42 HEV-1 VIDRL isolates and the HEV-1 subtype references. This tree shows that the reference sequences are consistently grouped into the correct subtypes despite the short amplicon length.

[1]https://www.geneious.com

[2]https://www.rivm.nl/mpf/typingtool/hev/

TABLE 1 | Demographic and clinical history of HEV RNA positive cases, divided by genotype.

	HEV-1	HEV-3	HEV-4	Total
Number of isolates (% of total number)	42 (64%)	22 (33%)	2 (3%)	66
Median Age (range), years	30.6 (17-66.9)	61.7 (35-82.1)	63.5 (63.5)	
Gender				
Male	22 (52.4%)	11 (50.0%)	1 (50%)	34
Female	14 (33.3%)	8 (36.4%)	1 (50%)	23
Unknown	6 (14.3%)	3 (13.6%)	0	9
State/Territory of Origin				
Victoria	12 (28.6%)	3 (13.6%)	1 (50%)	16
NSW	15 (35.7%)	14 (63.6%)	1 (50%)	30
Western Australia	9 (21.4%)	4 (18.2%)	0	13
South Australia	4 (9.5%)	0	0	4
Tasmania	1 (2.4%)	0	0	1
ACT	0	1 (4.5%)	0	1
Northern Territory	1 (2.4%)	0	0	1
Recent travel				
India	20 (47.6%)	0	0	20
Pakistan	3 (7.1%)	0	0	3
China	0	0	1 (50%)	1
Japan	1 (2.4%)	0	0	1
Portugal	0	1 (4.5%)	0	1
England	0	1 (4.5%)		1
Singapore	0	1 (4.5%)		1
Undisclosed travel	2 (4.8%)	0	0	2
Unknown	16 (38.1%)	13 (59.1%)	1 (50%)	29
Locally acquired	0	6 (27.3%)	0	6

Phylogenetic analysis showed that six isolates clustered within the 1f subtype with bootstrap support of 99% but the majority of VIDRL isolates (n = 36/42, 86%) cluster with the subtype 1g with 97% bootstrap support. This includes a cluster of five sequences (100% bootstrap value) from unrelated individuals who had all traveled to Gujarat, India within a similar time period. However, most cases that are genetically related were identified from different years. More than 60% (n = 26/42) of patients diagnosed with HEV-1 reported a history of travel including 23/42 (54%) patients who had traveled to India or Pakistan.

HEV Genotype 3

Figure 2 shows the phylogenetic relationship of 22 HEV-3 isolates with the reference HEV-3 subtypes. Despite the short amplicon length, the reference sequences clustered appropriately into subtype groups. All 22 study isolates clustered within HEV-3 group 2 (HEV-3 abchij) but only five isolates could be subtyped. Three of the five isolates clustered to subtypes that have been reported in European countries. Of these, two isolates were identified as subtype 3c, one of whom had a history of travel to the United Kingdom during the incubation period. The third isolate was identified as subtype 3m and the patient had reported recent travel to Portugal. The two remaining isolates that could be subtyped had clustered with 3a sequences previously reported in Singapore. One patient was confirmed to have traveled to Singapore in the incubation period.

Twelve isolates collected between 2015 and 2020 from patients in Western Australia (WA), Victoria (VIC), NSW and Australian Capital Territory (ACT), formed a unique clade in the HEV-3 group 2 (HEV-3 abchij) monophyletic group and could not be subtyped. This clade included the index case from the 2013/4 NSW HEV-3 outbreak that was linked to consumption of undercooked pork (Yapa et al., 2016). Two distinct clusters were observed in this clade. The first cluster had seven isolates (>98% bootstrap support) that shared a common ancestor with the NSW outbreak index strain and had pairwise sequence homology that ranged between 97.2 and 99.2%. Two of these patients were suspected to have locally acquired infection. A related second cluster of 3 isolates from NSW and ACT (100 bootstrap support) shared a common ancestor with the 2013/4 NSW HEV-3 outbreak sequence with 93.8–94.6% sequence identity. Two of these patients were suspected to have locally acquired infection.

Additionally, five isolates collected from NSW patients in 2019 and 2020 formed a second unique clade with a bootstrap value of 99% within the HEV-3abchij group. These isolates formed a sister clade with the European subtype 3h sequences, which has been reported in humans, pigs and wild boar (Adlhoch et al., 2009; Smith et al., 2020). Three of these isolates were collected within a 6-week period in NSW and shared 100% sequence homology. Whilst no clinical information was available for these cases, a fourth isolate suspected of being locally acquired, shared 97.4% homology with these sequences. A fifth isolate, collected from an NSW patient in 2020 and believed to be locally acquired, also clustered with this group with 90% sequence identity.

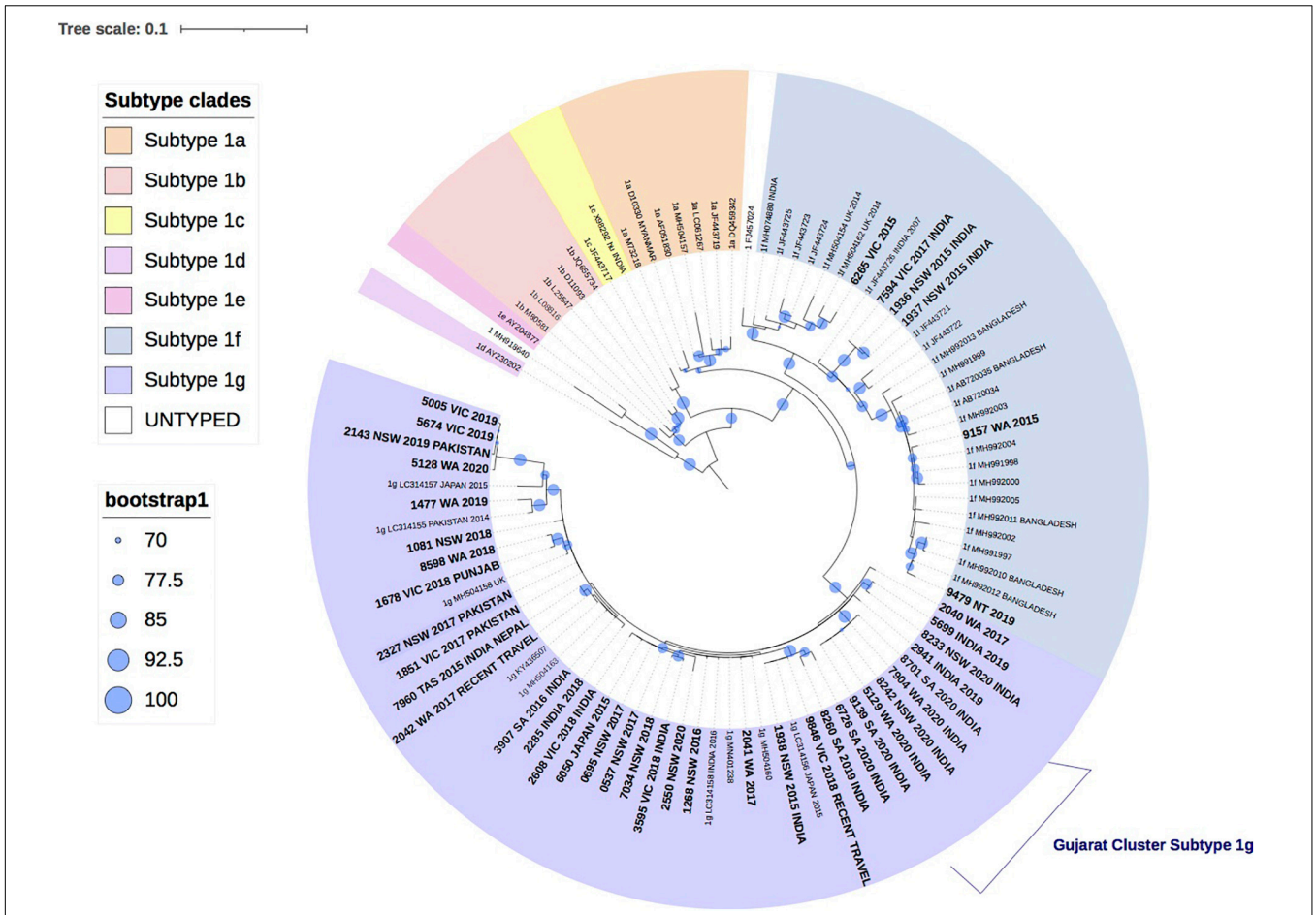

FIGURE 1 | A maximum likelihood phylogenetic tree of the hepatitis E virus (HEV) ORF2 region (493nt) showing the inferred relationship between the HEV-1 sequences isolated from this study, and HEV sequences of known genotype 1 retrieved from Genbank. The tree was constructed using the model TNe + I + G4, and branches are annotated with UF bootstrap support (1,000 replicates). Significant bootstrap values (>70) are reported. Accession numbers and country of origin of the sequences downloaded from Genbank are indicated on the tree. Subtypes are indicated by color. The HEV sequences from this study are depicted in bold, with known country of origin or recent travel indicated.

HEV Genotype 4

Figure 3 shows the two HEV-4 isolates cluster with subtypes 4b and 4d, respectively, with strong bootstrap support.

DISCUSSION

This is the first study to characterize the genotype distribution and perform phylogenetic analysis of HEV in Australia. The phyloanalysis demonstrates that the majority of HEV infections in Australia are associated with overseas travel (n = 48, 73%), however, autochthonous infections continue to occur annually. As expected, the majority of HEV infections tested at VIDRL were genotype 1 (n = 55, 60.4% of all isolates, 64% of the phyloanalysis group), which is prevalent in South Asian countries including India and Pakistan, where more than 54% (n = 23/42) of the phyloanalysis population with HEV-1 had reported recent travel. Whilst travel history was not provided for 16/42 (38%) of this group with HEV-1, the phylogenetic analysis indicates that these isolates are related to 1g or 1f sequences previously reported in India, Pakistan, and Japan (Nishizawa et al., 2017; Pathak and Barde, 2017). This is consistent with previous publications in which 88% of people diagnosed with hepatitis E in NSW had reportedly traveled to South Asia (Gunaratnam et al., 2014).

Phylogenetic analysis further revealed a cluster of five HEV-1 sequences, classified as subtype 1g, which shared 99.9% sequence identity and were collected from unrelated patients in South Australia (SA), WA and NSW, who had all traveled to Gujarat, India in 2019 and 2020. Outbreaks of hepatitis E infection are common in India and have been linked to faecal contaminated water, with subtypes 1a, 1c, and 1f most commonly identified (Gurley et al., 2014; Tripathy et al., 2019; Malhotra et al., 2020). HEV subtype 1g was first proposed as a new subtype in 2017 and has retrospectively been identified as the cause of a hepatitis E outbreak in Nepal and Mongolia in 2014 and 2015, respectively (Nishizawa et al., 2017).

Patients with HEV-1 were more likely to be younger with an average age of 30.6 years compared to 61.7 years for patients with

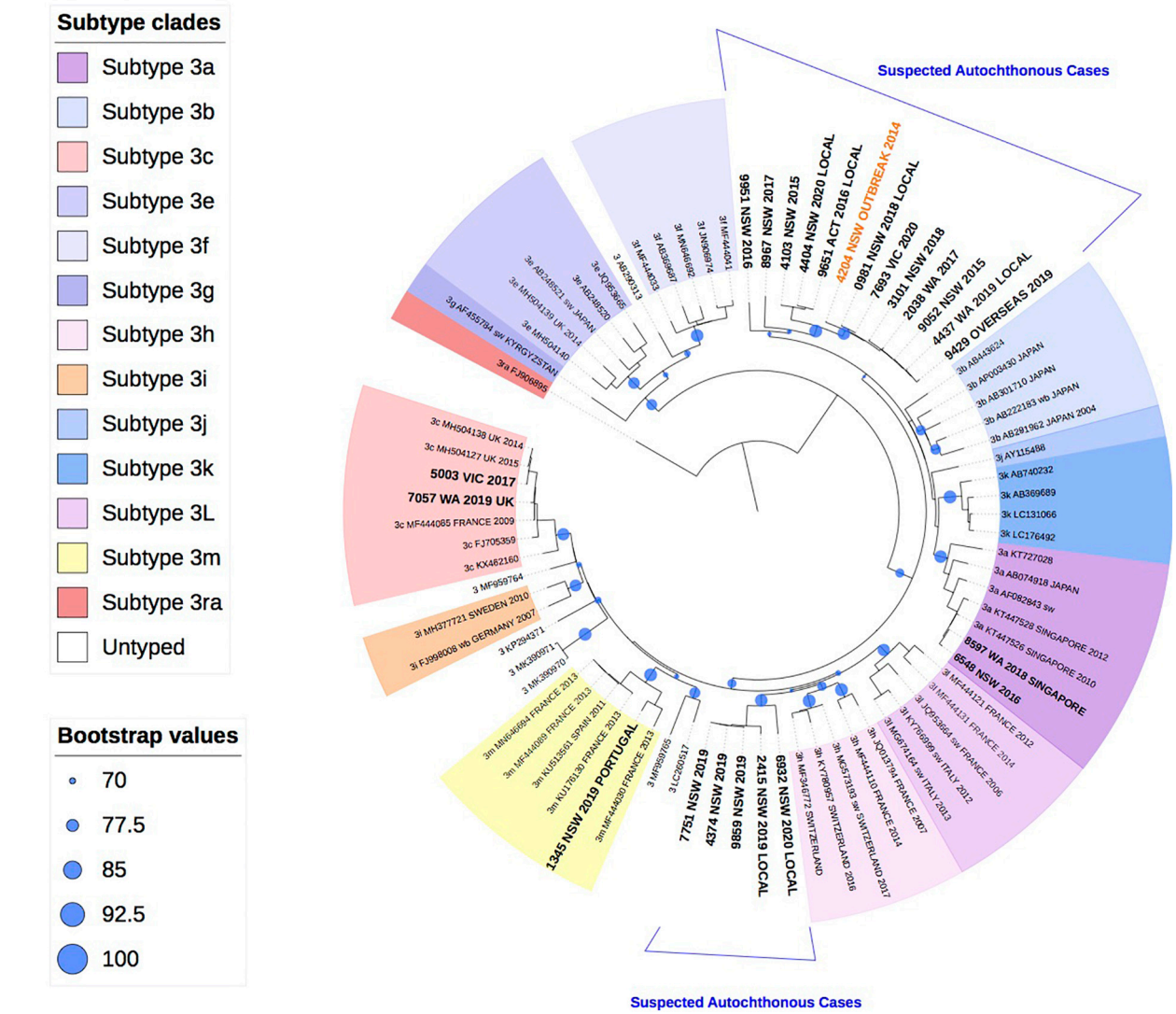

FIGURE 2 | A maximum likelihood phylogenetic tree of the hepatitis E virus (HEV) ORF2 region (493nt) showing the inferred relationship between the HEV-3 sequences isolated from this study, and HEV sequences of known genotype 3 retrieved from Genbank. The tree was constructed using the model TIM3e + I + G4, and branches are annotated with UF bootstrap support (1,000 replicates). Significant bootstrap values (>70) are reported. Accession numbers and country of origin of the sequences downloaded from Genbank are indicated on the tree. Subtypes are indicated by color. The HEV sequences from this study are depicted in bold, with known country of origin indicated.

HEV-3 as has previously been described (Kamar et al., 2017). Where gender was known, males were more likely to be infected, representing 61.1% of HEV-1 cases and 57.9% of HEV-3 cases. This is consistent with previous studies where symptomatic HEV-3 infections affect males in 75% of reported cases (Dalton et al., 2008). It is important to note that two of the female patients were receiving immunosuppressive therapy and another two women were >70 years of age. In our analysis, ten patients had only weakly reactive anti-HEV IgG, including two patients who were immunocompromised, so if HEV is clinically suspected, it is important to consider molecular testing even if the serological results are not indicative of current infection. IgG is used at VIDRL in favor of IgM, which is not recommended in settings where HEV infection is uncommon (Lin et al., 2000). However, it may have been useful in this case and its implementation may be considered. Nonetheless follow-up samples are routinely requested when the s/co is <3 to detect potential sero-converters.

In this report 34 (37.4%) (or 22, 33% phyloanalysis group) strains over the 6-year study period in Australia were classified as genotype 3. This is lower than that reported in other industrialized nations where HEV-3 is the dominant cause of hepatitis E infections including Belgium (92%), England (85%)

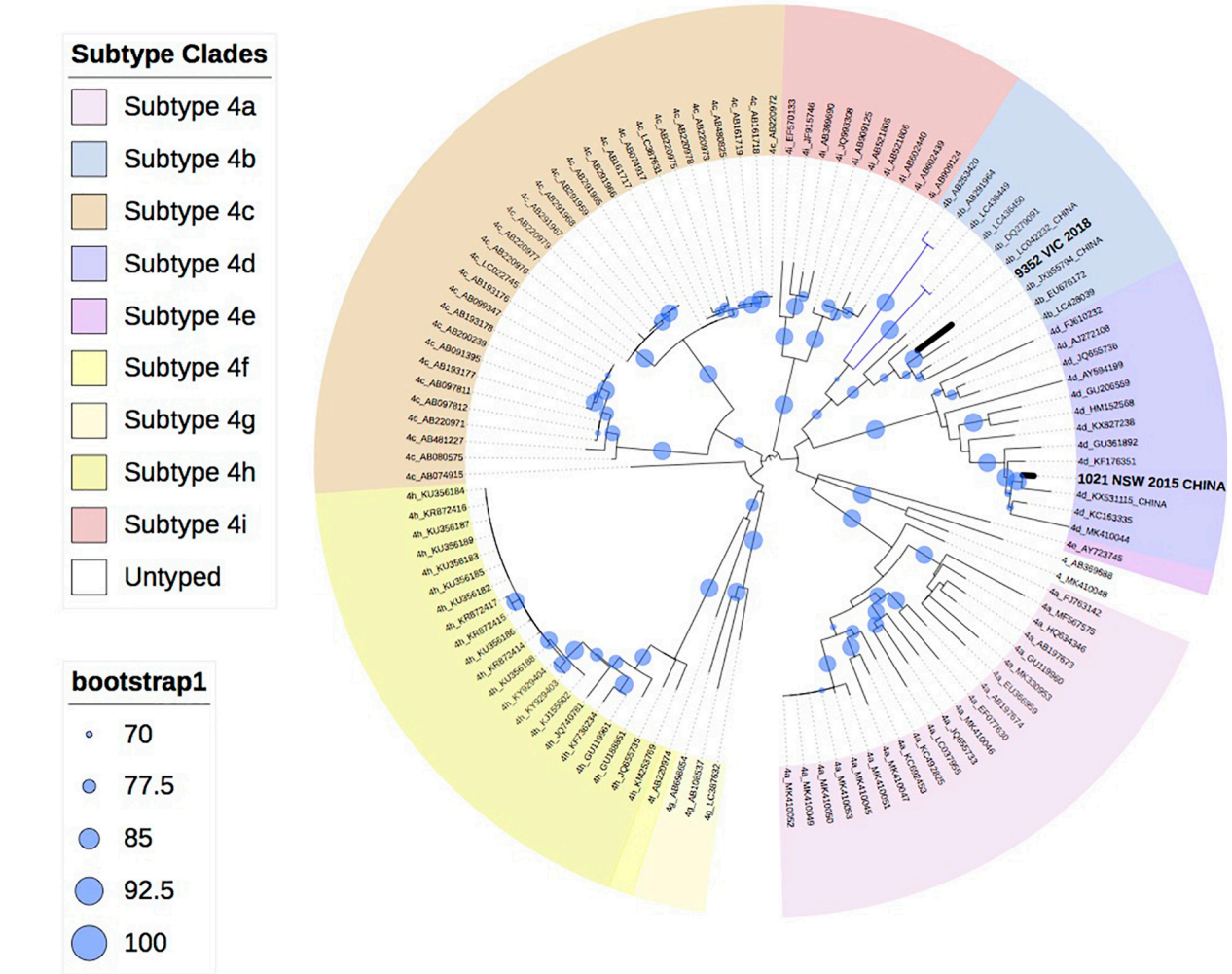

FIGURE 3 | A maximum likelihood phylogenetic tree of the hepatitis E virus (HEV) ORF2 region (493nt) showing the inferred relationship between the HEV-4 sequences isolated from this study, and HEV sequences of known genotype 4 retrieved from Genbank. The tree was constructed using the TIM2e + G4 model and branches are annotated with UF bootstrap support (1,000 replicates). Significant bootstrap values (>70) are reported. Accession numbers and country of origin of the sequences downloaded from Genbank are indicated on the tree. Subtypes are indicated by color. The HEV sequences from this study are depicted in bold, with known country of origin indicated.

and Bulgaria (98%) (Bruni et al., 2018; Oeser et al., 2019; Suin et al., 2019). Indeed, a 2016 investigation of HEV RNA prevalence in Australian blood donors was estimated to be 1 in 74,131, which was lower than the HEV RNA prevalence estimates reported by other developed countries (Hoad et al., 2017). Prior to 2013, there had only been one published case of locally acquired HEV-3 in Australia in 1995 (Heath et al., 1995), with a further two unpublished reports in 2005. In this analysis, there were potentially 17 cases of HEV-3 that were locally acquired, **Figure 2**, with at least six of these having reported a clinical suspicion of local acquisition. The phylogenetic analysis shows that these isolates formed two distinct clusters in the GT3 group 2 clade. These isolates could not be subtyped based on the reference sequences or by the HEVnet typing tool and may actually represent new "Australian" subtype/s. This certainly warrants further investigation.

As HEV-3 is considerably diverse the subtype classification continues to be updated and recent revisions have proposed 13 HEV-3 subtypes and six unclassified isolates (Smith et al., 2020). For a new subtype to be designated, at least three full-length HEV sequences are required from sources that are not epidemiologically linked. Whilst 17/22 (77%) of HEV-3 isolates in this study could not be subtyped, full genome analysis will be attempted in future analyses to ascertain if these isolates constitute new Australian subtype/s. Nonetheless, results here suggest that there are at least two genetically distinct HEV strains circulating in Australia. The first group of cases consisted of five HEV-3 isolates that were all collected in NSW within a 12-month

period. Whilst three of these cases did not have travel history provided, they shared a common ancestor with two cases with suspected local acquisition. For the second group, ten of the HEV-3 study isolates shared a common ancestor with the index case of the 2013/4 NSW HEV outbreak, sharing 93.8–99.2% sequence identity. This supports our findings that these cases are likely to be autochthonous infections. There were another two isolates that were genetically related to this group, sharing 89.6–90.1% sequence identity with the outbreak sequence inferring that these cases may also have been locally acquired.

The 2013/4 HEV outbreak in NSW was linked to consumption of undercooked pork liver sourced in Australia (Yapa et al., 2016). Indeed, all fresh pork consumed in Australia is locally sourced and imported pork products must be cooked to inactivate any pig diseases (Australian Pig Farmers, 2020). Previous studies, in China and the Nordic region of Europe, have shown high similarities between HEV strains isolated from patients and pigs in the same geographical region (Norder et al., 2009; Fu et al., 2010). In our study, 13/17 (76%) of suspected locally acquired HEV-3 infections occurred in patients from NSW and ACT, which was a surprising observation given that NSW produces only 16% of the total number of Australian pigs annually (Australian Pork Limited, 2019). It is not known if NSW is over-represented in autochthonous HEV-3 cases due to a higher incidence of HEV circulating in NSW pigs or perhaps due to enhanced surveillance for locally acquired hepatitis E following the 2013/4 outbreak. Interestingly, in a 1999 seroprevalence study of anti-HEV IgG in Australian pigs, four of six farms included were in NSW, including two where 90–95% of pigs were seropositive for HEV antibodies by 16 weeks of age (Chandler et al., 1999). A research pig farm from VIC was used as a HEV negative control in this study and wild pigs from the Northern Territory (NT) were also sampled but no pig farms from other states or territories were included. As such, further investigation is needed to ascertain the current prevalence of HEV in pigs across Australia as well as characterizing the circulating HEV strains.

In this analysis, only 5/22 (23% phyloanalysis group) HEV-3 cases were believed to have been acquired overseas based on travel history and phylogenetic analysis. Two isolates, that were collected 2 years apart and from different states, were subtyped as 3a based on phylogenetic analysis and clustered with sequences reported in Singapore. This concurs with the travel history of one of these patients who had traveled to Singapore within the incubation period. HEV genotype 3a strains have been isolated from patients and pigs in Singapore, where HEV-3 is the dominant genotype (Wong et al., 2019). Similarly, two isolates were identified as subtype 3c, which has been reported in European countries including France and England and was consistent with the travel history reported for one of the patients. Interestingly, the second patient was chronically infected with HEV and had previously received a liver transplant. No travel history was available for this patient and we cannot rule out that the infection was acquired through transplantation or transfusion, which has previously been described (Speers et al., 2015). Alternatively, HEV infection may have contributed to the patient's liver failure prior to transplantation.

Finally, this analysis showed that infection with HEV-4 is rare in Australia, with two cases in the last 5 years. This is consistent with industrialized countries in Europe, where HEV-4 has been infrequently reported, such as Belgium (<2%), England (<1%) and Switzerland (3.6%) (Fraga et al., 2018; Oeser et al., 2019; Suin et al., 2019). In this study, phylogenetic analysis indicates that both isolates clustered with sequences identified in China, which is consistent with the travel history of one of the patients. HEV-4 is largely confined to Asia and a study in Eastern China found that 93.5% of hepatitis E infections were caused by the zoonotic HEV-4 (Lu et al., 2006; Zhu et al., 2014).

A limitation of this study is that it only represents HEV positive samples that were tested at VIDRL. No samples from QLD were analyzed and NSW laboratories were most likely to send samples that were suspected of being locally acquired as recommended by the Department of Health. QLD only reported the first published case of locally acquired HEV-3 in 2019 (Taylor et al., 2020). As such, the incidence of HEV-3 may be over-represented in this analysis. Nonetheless, this study has shown that small numbers of locally acquired infections continue to occur annually and it is important to screen for HEV infections in patients presenting with unexplained hepatitis with and without travel history.

CONCLUSION

This study characterizes the HEV isolates detected in Australia thereby contributing to the global understanding of autochthonous HEV-3 and confirming the acquisition of HEV-1 during travel to endemic areas. Results also suggest the presence of a new "Australian" subtype/s of HEV-3 which requires further investigation. This provides valuable information to clinicians and scientists with respect to diagnosis of HEV infection and may prepare us to implement public health measures should an outbreak occur. The morbidity and mortality associated with HEV infection, particularly in immunosuppressed individuals and pregnant women, reaffirms the need to screen for HEV both locally and in travelers.

DATA AVAILABILITY STATEMENT

The nucleotide sequences of this study were deposited in GenBank (Accession numbers MW355681–MW355746).

AUTHOR CONTRIBUTIONS

JO'K: data curation, formal analysis, investigation, methodology, and roles/writing – original draft. LT: methodology. LY: formal analysis and writing – review and editing. SB and XL: methodology. BC: investigation and writing – review and editing. SN: methodology and writing – review and editing. KJ: conceptualization, project administration, supervision, and writing – review and editing. All authors contributed to the article and approved the submitted version.

FUNDING

This study was funded by VIDRL.

ACKNOWLEDGMENTS

The authors wish to acknowledge Professor Scott Bowden.

REFERENCES

Adlhoch, C., Wolf, A., Meisel, H., Kaiser, M., Ellerbrok, H., and Pauli, G. (2009). High HEV presence in four different wild boar populations in East and West Germany. *Vet. Microbiol.* 139, 270–278. doi: 10.1016/j.vetmic.2009.06.032

Australian Government Department of Health (2020). *Number of Notifications of Hepatitis E, Received From State and Territory Health Authorities in the Period of 1991 to 2019 and Year to Date Notifications for 2020. [Online]*. Available online at: http://www9.health.gov.au/cda/source/rpt_4.cfm (accessed October 10, 2020).

Australian Pig Farmers (2020). *Buying Pork [Online]*. Available online at: https://aussiepigfarmers.com.au/pork/our-plate/buying-pork/ (accessed October 20, 2020).

Australian Pork Limited (2019). *Import, Export and Domestic production Report (May-19). (ACT)*. East Victoria Park, WA: Australian Pork Limited.

Benson, D. A., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., and Sayers, E. W. (2014). GenBank. *Nucleic Acids Res.* 42, D32–D37. doi: 10.1093/nar/gkt1030

Bruni, R., Villano, U., Equestre, M., Chionne, P., Madonna, E., Trandeva-Bankova, D., et al. (2018). Hepatitis E virus genotypes and subgenotypes causing acute hepatitis, Bulgaria, 2013-2015. *PLoS One* 13:e0198045. doi: 10.1371/journal.pone.0198045

Chandler, J. D., Riddell, M. A., Li, F., Love, R. J., and Anderson, D. A. (1999). Serological evidence for swine hepatitis E virus infection in Australian pig herds. *Vet. Microbiol.* 68, 95–105.

Colson, P., Romanet, P., Moal, V., Borentain, P., Purgus, R., Benezech, A., et al. (2012). Autochthonous infections with hepatitis E virus genotype 4. *France Emerg. Infect. Dis.* 18, 1361–1364. doi: 10.3201/eid1808.111827

Dalton, H. R., Stableforth, W., Thurairajah, P., Hazeldine, S., Remnarace, R., Usama, W., et al. (2008). Autochthonous hepatitis E in Southwest England: natural history, complications and seasonal variation, and hepatitis E virus IgG seroprevalence in blood donors, the elderly and patients with chronic liver disease. *Eur. J. Gastroenterol. Hepatol.* 20, 784–790.

Erker, J. C., Desai, S. M., and Mushahwar, I. K. (1999). Rapid detection of Hepatitis E virus RNA by reverse transcription-polymerase chain reaction using universal oligonucleotide primers. *J. Virol. Methods* 81, 109–113.

European Association for the Study of the Liver (2018). EASL clinical practice guidelines on hepatitis E virus infection. *J. Hepatol.* 68, 1256–1271.

Forgan-Smith, K. R., and Macdonald, G. (2019). Chronic hepatitis E infection in an immunosuppressed, solid organ transplant patient. *Intern Med. J.* 49, 1335–1336.

Fraga, M., Doerig, C., Moulin, H., Bihl, F., Brunner, F., Müllhaupt, B., et al. (2018). Hepatitis E virus as a cause of acute hepatitis acquired in Switzerland. *Liver Int.* 38, 619–626.

Fu, H., Li, L., Zhu, Y., Wang, L., Geng, J., Chang, Y., et al. (2010). Hepatitis E virus infection among animals and humans in Xinjiang, China: possibility of swine to human transmission of sporadic hepatitis E in an endemic area. *Am. J. Trop. Med. Hyg.* 82, 961–966.

Gunaratnam, P., Tobin, S., Seale, H., and Mcanulty, J. M. (2014). Infectious diseases in returned travellers, NSW, 2010-2011. *N. S. W. Public Health Bull.* 24, 171–175.

Gurley, E. S., Hossain, M. J., Paul, R. C., Sazzad, H. M., Islam, M. S., Parveen, S., et al. (2014). Outbreak of hepatitis E in urban Bangladesh resulting in maternal and perinatal mortality. *Clin. Infect. Dis.* 59, 658–665.

Heath, T. C., Burrow, J. N., Currie, B. J., Bowden, F. J., Fisher, D. A., Demediuk, B. H., et al. (1995). Locally acquired hepatitis E in the Northern Territory of Australia. *Med. J. Aust.* 162, 318–319.

Hewitt, J., Harte, D., Sutherland, M., Croucher, D., Fouche, L., Flanagan, P., et al. (2018). Prevalence of hepatitis E virus antibodies and infection in New Zealand blood donors. *N. Z. Med. J.* 131, 38–43.

Hewitt, P. E., Ijaz, S., Brailsford, S. R., Brett, R., Dicks, S., Haywood, B., et al. (2014). Hepatitis E virus in blood components: a prevalence and transmission study in southeast England. *Lancet* 384, 1766–1773.

Hoad, V. C., Seed, C. R., Fryk, J. J., Harley, R., Flower, R. L. P., Hogema, B. M., et al. (2017). Hepatitis E virus RNA in Australian blood donors: prevalence and risk assessment. *Vox. Sang* 112, 614–621.

Hoang, D. T., Chernomor, O., Von Haeseler, A., Minh, B. Q., and Vinh, L. S. (2018). UFBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* 35, 518–522. doi: 10.1093/molbev/msx281

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., Von Haeseler, A., and Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14, 587–589.

Kamar, N., Dalton, H. R., Abravanel, F., and Izopet, J. (2014). Hepatitis E virus infection. *Clin. Microbiol. Rev.* 27, 116–138.

Kamar, N., Izopet, J., Pavio, N., Aggarwal, R., Labrique, A., Wedemeyer, H., et al. (2017). Hepatitis E virus infection. *Nat. Rev. Dis. Primers* 3:17086.

Kamar, N., Mansuy, J. M., Cointault, O., Selves, J., Abravanel, F., Danjoux, M., et al. (2008a). Hepatitis E virus-related cirrhosis in kidney- and kidney-pancreas-transplant recipients. *Am. J. Transplant* 8, 1744–1748.

Kamar, N., Selves, J., Mansuy, J. M., Ouezzani, L., Péron, J. M., Guitard, J., et al. (2008b). Hepatitis E virus and chronic hepatitis in organ-transplant recipients. *N. Engl. J. Med.* 358, 811–817.

Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30, 3059–3066. doi: 10.1093/nar/gkf436

Lin, C. C., Wu, J. C., Chang, T. T., Chang, W. Y., Yu, M. L., Tam, A. W., et al. (2000). Diagnostic value of immunoglobulin G (IgG) and IgM anti-hepatitis E virus (HEV) tests based on HEV RNA in an area where hepatitis E is not endemic. *J. Clin. Microbiol.* 38, 3915–3918.

Lu, L., Li, C., and Hagedorn, C. H. (2006). Phylogenetic analysis of global hepatitis E virus sequences: genetic diversity, subtypes and zoonosis. *Rev. Med. Virol.* 16, 5–36. doi: 10.1002/rmv.482

Maila, H. T., Bowyer, S. M., and Swanepoel, R. (2004). Identification of a new strain of hepatitis E virus from an outbreak in Namibia in 1995. *J. Gen. Virol.* 85, 89–95. doi: 10.1099/vir.0.19587-0

Malhotra, B., Deba, F., Sharma, P., Trivedi, K., Tiwari, J., Chouhan, A., et al. (2020). Hepatitis E outbreak in Jaipur due to Genotype IA. *Indian J. Med. Microbiol.* 38, 46–51.

Mulder, A. C., Kroneman, A., Franz, E., Vennema, H., Tulen, A. D., Takkinen, J., et al. (2019). HEVnet: a One Health, collaborative, interdisciplinary network and sequence data repository for enhanced hepatitis E virus molecular typing, characterisation and epidemiological investigations. *Euro Surveill.* 24:1800407. doi: 10.2807/1560-7917.ES.2019.24.10.1800407

Nagashima, S., Takahashi, M., Jirintai, Tanaka, T., Yamada, K., Nishizawa, T., et al. (2011). A PSAP motif in the ORF3 protein of hepatitis E virus is necessary for virion release from infected cells. *J. Gen. Virol.* 92, 269–278. doi: 10.1099/vir.0.025791-0

Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300

Nishizawa, T., Primadharsini, P. P., Namikawa, M., Yamazaki, Y., Uraki, S., Okano, H., et al. (2017). Full-length genomic sequences of new subtype 1g hepatitis E virus strains obtained from four patients with imported or autochthonous acute hepatitis E in Japan. *Infect. Genet. Evol.* 55, 343–349. doi: 10.1016/j.meegid.2017.10.007

Norder, H., Sundqvist, L., Magnusson, L., Østergaard Breum, S., Löfdahl, M., Larsen, L. E., et al. (2009). Endemic hepatitis E in two Nordic countries. *Euro Surveill.* 14:19211.

Oeser, C., Vaughan, A., Said, B., Ijaz, S., Tedder, R., Haywood, B., et al. (2019). Epidemiology of Hepatitis E in England and Wales: a 10-Year Retrospective Surveillance Study, 2008-2017. *J. Infect. Dis.* 220, 802–810.

Pathak, R., and Barde, P. V. (2017). Detection of genotype 1a and 1f of hepatitis E Virus in patients treated at tertiary care hospitals in central India. *Intervirology* 60, 201–206. doi: 10.1159/000487052

Rein, D. B., Stevens, G. A., Theaker, J., Wittenborn, J. S., and Wiersma, S. T. (2012). The global burden of hepatitis E virus genotypes 1 and 2 in 2005. *Hepatology* 55, 988–997.

Shrestha, A. C., Flower, R. L., Seed, C. R., Keller, A. J., Harley, R., Chan, H. T., et al. (2016). Hepatitis E virus RNA in Australian blood donations. *Transfusion* 56, 3086–3093.

Shrestha, A. C., Seed, C. R., Flower, R. L., Rooks, K. M., Keller, A. J., Harley, R. J., et al. (2014). Hepatitis E virus and implications for blood supply safety, Australia. *Emerg. Infect. Dis.* 20, 1940–1942.

Smith, D. B., Izopet, J., Nicot, F., Simmonds, P., Jameel, S., Meng, X. J., et al. (2020). Update: proposed reference sequences for subtypes of hepatitis E virus (species *Orthohepevirus* A). *J. Gen. Virol.* 101, 692–698.

Speers, D. J., Ma, M. X., Faddy, H. M., Nash, M., Bowden, S., De Boer, B., et al. (2015). Domestically acquired hepatitis E successfully treated with ribavirin in an Australian liver transplant recipient. *Med. J. Aust.* 202, 209–211.

Suin, V., Klamer, S. E., Hutse, V., Wautier, M., Jacques, M., Abady, M., et al. (2019). Epidemiology and genotype 3 subtype dynamics of hepatitis E virus in Belgium, 2010 to 2017. *Euro Surveill.* 24:1800141.

Taylor, A., Ebert, E., Mcmahon, J., Britton, L., Siebert, D., Slinko, V., et al. (2020). Locally acquired, human hepatitis E infection in Queensland. *Intern Med. J.* 50, 643–644.

Thom, K., Gilhooly, P., Mcgowan, K., Malloy, K., Jarvis, L. M., Crossan, C., et al. (2018). Hepatitis E virus (HEV) in Scotland: evidence of recent increase in viral circulation in humans. *Euro Surveill.* 23:17-00174.

Tripathy, A. S., Sharma, M., Deoshatwar, A. R., Babar, P., Bharadwaj, R., and Bharti, O. K. (2019). Study of a hepatitis E virus outbreak involving drinking water and sewage contamination in Shimla, India, 2015-2016. *Trans. R. Soc. Trop. Med. Hyg.* 113, 789–796.

Wong, C. C., Thean, S. M., Ng, Y., Kang, J. S. L., Ng, T. Y., Chau, M. L., et al. (2019). Seroepidemiology and genotyping of hepatitis E virus in Singapore reveal rise in number of cases and similarity of human strains to those detected in pig livers. *Zoonoses Public Health* 66, 773–782.

Yapa, C. M., Furlong, C., Rosewell, A., Ward, K. A., Adamson, S., Shadbolt, C., et al. (2016). First reported outbreak of locally acquired hepatitis E virus infection in Australia. *Med. J. Aust.* 204:274.

Zaaijer, H. L., Kok, M., Lelie, P. N., Timmerman, R. J., Chau, K., and Van Der Pal, H. J. (1993). Hepatitis E in The Netherlands: imported and endemic. *Lancet* 341:826.

Zhu, F. C., Huang, S. J., Wu, T., Zhang, X. F., Wang, Z. Z., Ai, X., et al. (2014). Epidemiology of zoonotic hepatitis E: a community-based surveillance study in a rural population in China. *PLoS One* 9:e87154.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 03 March 2021
doi: 10.3389/fmicb.2021.640889

Hepatitis B Virus (HBV) Genotype Mixtures, Viral Load, and Liver Damage in HBV Patients Co-infected With Human Immunodeficiency Virus

*Alexis Jose-Abrego[1,2], Sonia Roman[1,2], João Renato Rebello Pinho[3,4], Vanessa Fusco Duarte de Castro[4] and Arturo Panduro[1,2]**

[1] Department of Molecular Biology in Medicine, Civil Hospital of Guadalajara, "Fray Antonio Alcalde," Guadalajara, Mexico, [2] Health Sciences Center, University of Guadalajara, Guadalajara, Mexico, [3] LIM-07, Department of Gastroenterology, School of Medicine, Institute of Tropical Medicine, University of São Paulo, São Paulo, Brazil, [4] Hospital Israelita Albert Einstein, São Paulo, Brazil

OPEN ACCESS

Edited by:
Carla Osiowy,
National Microbiology Laboratory,
Public Health Agency of Canada
(PHAC), Canada

Reviewed by:
Guido van Marle,
University of Calgary, Canada
Masahiko Ito,
Hamamatsu University School
of Medicine, Japan

***Correspondence:**
Arturo Panduro
apanduro@prodigy.net.mx;
biomomed@cencar.udg.mx

Specialty section:
This article was submitted to
Virology,
a section of the journal
Frontiers in Microbiology

Received: 12 December 2020
Accepted: 10 February 2021
Published: 03 March 2021

Citation:
Jose-Abrego A, Roman S,
Rebello Pinho JR, de Castro VFD and
Panduro A (2021) Hepatitis B Virus
(HBV) Genotype Mixtures, Viral Load,
and Liver Damage in HBV Patients
Co-infected With Human
Immunodeficiency Virus.
Front. Microbiol. 12:640889.
doi: 10.3389/fmicb.2021.640889

Hepatitis B virus (HBV) co-infection is possible in patients who are positive for human immunodeficiency virus (HIV) since both share similar transmission routes. Furthermore, through the continuous risk of exposure, they potentially can be infected by mixtures of distinct HBV genotypes which can result in the presence of two or more genotypes in a single patient. This study aimed to specify the frequency of mixtures of HBV genotypes and their potential clinic importance in HIV-infected Mexican patients. HBV infection was assessed by serological testing and molecular diagnostics. HBV mixtures were detected by multiplex PCR and DNA sequencing. Liver fibrosis was evaluated using transitional elastography, the Aspartate aminotransferase to Platelets Ratio Index score, and Fibrosis-4 score. Among 228 HIV-infected patients, 67 were positive for HBsAg. In 25 HBV/HIV co-infected patients, 44 HBV genotypes were found: H (50.0%, 22/44), G (22.7%, 10/44), D (15.9%, 6/44), A (9.1%, 4/44), and F (2.3%, 1/44). Among these, 44.0% (11/25) were single genotype, 36.0% (9/25) were dual and 20.0% (5/25) were triple genotype. The most frequent dual combination was G/H (44.4%, 4/9), while triple-mixtures were H/G/D (60.0%, 3/5). The increase in the number of genotypes correlated positively with age (Spearman's Rho = 0.53, p = 0.0069) and negatively with platelet levels (Spearman's Rho = − 0.416, p = 0.039). HBV viral load was higher in triply-infected than dually infected (31623.0 IU/mL vs. 1479.0 IU/mL, p = 0.029) patients. Triple-mixed infection was associated with significant liver fibrosis (OR = 15.0 95%CI = 1.29 – 174.38, p = 0.027). In conclusion, infection with mixtures of HBV genotypes is frequent in HIV patients causing significant hepatic fibrosis related to high viral load, especially in triple genotype mixtures.

Keywords: liver fibrosis, hepatitis B virus, co-infection, HBV genotype, HBV genotype mixtures, human immunodeficiency virus

INTRODUCTION

One severe clinical issue in patients with human immunodeficiency virus (HIV) is the high frequency of liver diseases, particularly those caused by chronic HBV infection (Sherman et al., 2017). Intravenous drug use and sexual behavior are some of the important risk factors associated with HBV/HIV co-infections. In hospitalizations related to liver complications, HBV/HIV co-infection is associated with poor clinical outcomes (Rajbhandari et al., 2016). It has also been observed that the progression of chronic HBV to cirrhosis or hepatocellular carcinoma is more rapid in HBV/HIV co-infected patients than in those who are HBV mono-infected (Thio et al., 2002). It is probably linked to an immunosuppression state by reducing the frequency of spontaneous recovery and increasing the risk of chronicity (Seetharam et al., 2014). Also, the clinical management in HBV/HIV co-infected patients can be difficult because they present a more insufficient response to interferon-alpha, have a high risk of HBV post-therapy reactivation, or develop HBV resistance mutations (Di Martino et al., 2002; Mendes-Correa et al., 2010; Tengan et al., 2017).

On the other hand, people with HIV who lack protective immunity present many risk factors exposing them to different HBV genotypes, leading to a mixed infection (Thio et al., 2002). Besides, the common factors mentioned above are social determinants, such as unemployment or living in poverty affect infection risk (Jose-Abrego et al., 2017). These can lead individuals engaging in high-risk behaviors for infection, such as prostitution and intravenous drug use, which may favor infection by different HBV genotypes. Together, with the immunological status of the HIV-infected individual, a mixture of HBV genotypes could develop. Mixtures of genotypes are when more than one HBV genotype is detected in the blood resulting in a simultaneous infection or superinfection (i.e., when a second genotype is detected after a single HBV genotype infection) (Cunningham et al., 2015). HBV genotype mixtures have been reported in some groups, such as patients treated with interferon and injection drug users (Hannoun et al., 2002; Chen et al., 2004). Toan et al. (2006) showed the clinical importance of HBV genotypes mixtures in people negative to hepatitis C virus and HIV. Data on the frequency of mixtures of HBV genotypes among HBV/HIV co-infected patients remains understudied.

Approximately 220,000 persons live with HIV in Mexico. Between 24 and 29% have evidence of HBV chronic infection (Torres-Baranda et al., 2006; Jose-Abrego et al., 2017; The Joint United Nations Programme on HIV/AIDS, 2019). Furthermore, previous studies have reported a high prevalence of occult hepatitis B (OBI) in HIV cohorts. Torres-Baranda et al. (2006) reported an OBI prevalence of 18.4% in West Mexico. In central Mexico, the prevalence of OBI varied between 36% and 49% (Alvarez-Munoz et al., 2014; Enriquez-Navarro et al., 2020). Although the endemic genotype H is common (64%) among HBV/HIV co-infected patients, other genotypes such as G (16%), A2(12%), F1b (4%), and D4(4%) have been reported (Jose-Abrego et al., 2017). This finding suggests that new introductions have occurred, requiring molecular epidemiology studies to decipher if mixed genotypes in HBV infections prevail. Furthermore, in our country, people who inject drugs and are MSM have a high risk of HBV/HIV co-infection (Jose-Abrego et al., 2017). These factors suggest that some people with HIV might acquire infections with different HBV genotypes. Therefore, in this study, we aimed to detect HBV genotypes' mixtures and their impact on viral load and liver damage in Mexican patients with HIV.

MATERIALS AND METHODS

Study Design and Socio-Demographic Data

This cross-sectional study included 228 patients with a serological or molecular diagnosis of HIV. All patients were attended at the Department of Molecular Biology in Medicine and the HIV Clinic of the Civil Hospital of Guadalajara "Fray Antonio Alcalde" from January 2014 to December 2016 (Jose-Abrego et al., 2017). Written informed consent was obtained from all participants. Demographic and clinical information, such as gender, age, HIV viral load, and the combination antiretroviral therapy (cART) received during recruitment, were obtained from the patients' medical records. The levels of liver enzymes, platelets, albumin, CD4 cells, and CD8 cells were obtained from our hospital's general laboratory. Also, hepatitis C virus (HCV) infection was assessed by the presence of anti-HCV antibodies (AxSYM®, Abbott Laboratories, Abbott Park, IL, United States). The methodological summary of this study is shown in **Figure 1**.

HBV Serological Markers and Viral Load

All serum samples were tested for hepatitis B surface antigen (HBsAg) using an enzyme-linked immunosorbent assay (MONOLISA^{TM} HBsAg Ultra, Bio-Rad, Poincaré, MC, France) according to the manufacturer's instructions. In samples positive to HBsAg, HBV viral load was quantified by real-time polymerase chain reaction (RT PCR) (COBAS® Ampliprep/COBAS® TaqMan® HBV test (Roche, V2.0), which has a detection limit between 20 and 170 million IU/mL (1IU corresponding to 5.82 genomic copies). Levels $\geq$ 2,000 IU/mL was considered as high viral load and <2,000 IU/mL as low viral load.

Extraction of HBV DNA

In the samples with detectable HBV viral load, total viral DNA was extracted using the QIAamp DNA Mini Kit (Qiagen Science, Hilden, Germany) according to the manufacturer's instructions. Additionally, a carrier DNA (1 μg/μL) (Qiagen Science, Hilden, Germany) and an elution volume of 50 μL were used to improve the extraction yield.

Qualitative Detection of HBV DNA

The yield of DNA extracts was tested by nested PCR. The amplification reaction was prepared with 5 μL of DNA, 1 μL of each primer (10 μM) (**Supplementary Table 1**), 0.5 μL of dNTPs mix (10 mM each), 1 μL of $MgCl_2$ (50 mM), 2.5 μL of PCR buffer (10×), 13.8 μL of water and 0.2 μL of Taq DNA polymerase (5U/μL) (Invitrogen, San Diego, CA, United States) in a final

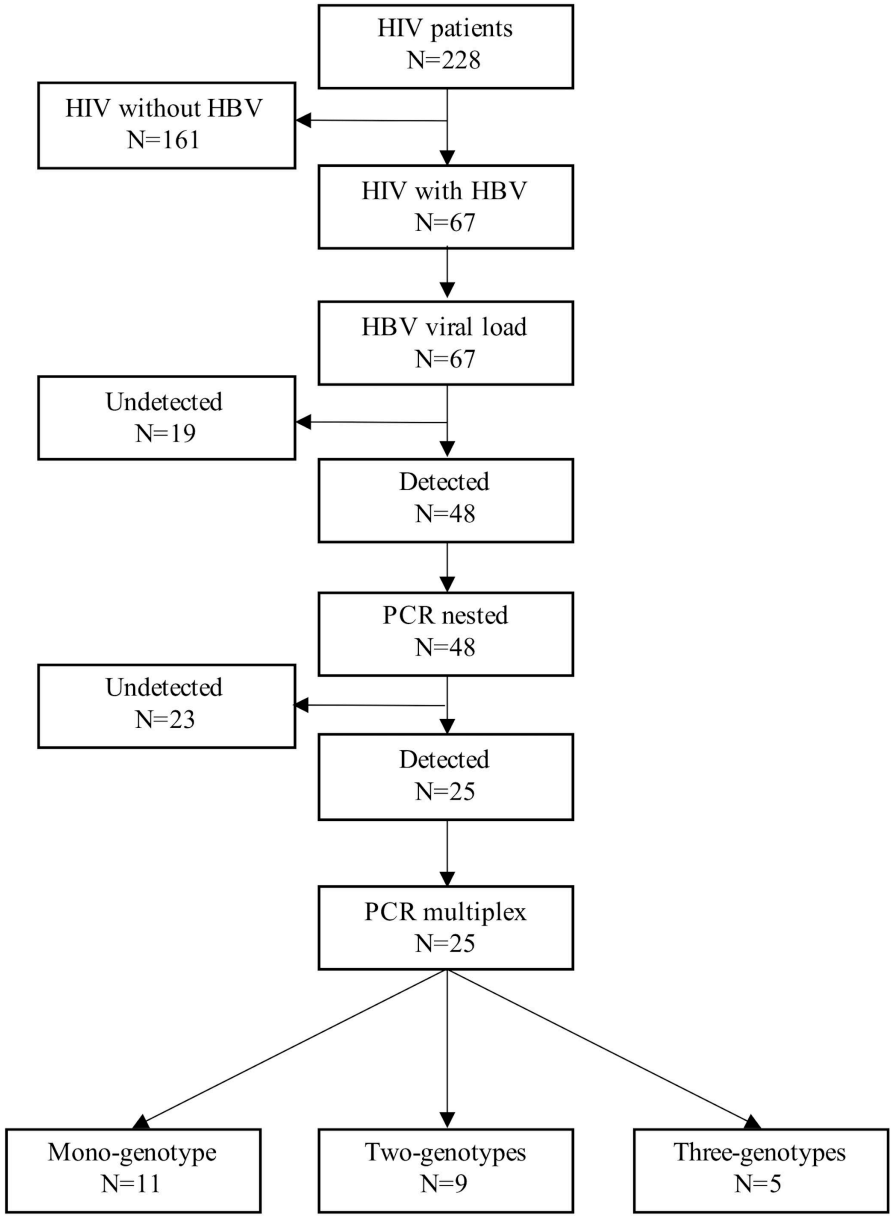

FIGURE 1 | Workflow used to detect mixtures of HBV genotypes in patients with HIV. First, serum samples were collected from all patients. These samples were analyzed by ELISA to evaluate the presence of HBsAg. Next, in the positive samples from the previous step, the viral load levels were measured by real-time PCR. A nested PCR was used to confirm HBV infection in the samples with detectable viral load. After that, a multiplex PCR was used to detect the predominant HBV genotypes in Mexico. Finally, the products were sequenced and genotyped by Sanger's method and phylogenetic analysis, respectively.

volume of 25 μL. The first round amplified a 418-bp fragment of the surface gene under the following conditions: 94°C for 3 min followed by 40 cycles of 94°C for 30 s, 55°C for 30 s, 72°C for 30 s, with a final extension step of 72°C for 5 min. The second round amplified a 232-bp product under the same conditions.

Identification of HBV Genotype Mixtures

DNA extracts were genotyped by multiplex PCR to identify the most predominant HBV genotypes in Mexico (H, G, A, and D) (Roman et al., 2014). These PCRs amplified different parts of the polymerase gene, a 370 bp fragment for genotype A, 147 bp for genotype D, and 279 bp for H, while a 584 bp fragment of the core gene was amplified for genotype G (**Supplementary Table 1**). The reaction was prepared with 2 μL of DNA extract, 1 μL of each genotype-specific primer set (10 μM), 8.5 μL of water, and 12.5 of 2× Hot StarTaq Master Mix (Qiagen Science, GmbH, Hilden, Germany). Thermal cycling conditions were as followed: preheating at 95°C, 15 min and 35 amplification cycles at 94°C for 1 min, 60°C for 1 min, 72°C for 2 min, and a final extension at 72°C for 10 m. DNA amplicons were run on a 2% TBE-agarose

gel stained with SYBR Safe DNA gel (Invitrogen, Carlsbad, CA, United States) and evaluated under UV light. The sizes of PCR products were estimated based on the migration pattern of a 50-bp DNA ladder (Invitrogen, Carlsbad, CA, United States). Quality control procedures included adding positive and negative controls in each PCR assay, and standard precautions were taken to avoid cross-contamination. The primer sequences designed for genotypes H and G were previously published (Panduro et al., 2013; Roman et al., 2014), whereas primers for genotypes A and D were used as published by Kirschberg et al. (2004) (**Supplementary Table 1**). Mixtures of HBV genotypes were defined as finding two or more products in agarose gel, and the genotype was determined according to the band's size of the amplified DNA.

HBV-DNA Sequencing and Phylogenetic Analysis

Hepatitis B virus genotypes identified by multiplex PCR were confirmed by DNA Sanger sequencing. The bands were cut and purified with Illustra GFX PCR DNA and Gel Band Purification Kit (GE Healthcare, Pittsburgh, PA, United States). The bidirectional sequencing reaction was prepared with 1 μL of each purified band, 2 μL of each genotype-specific primer (1 μM), 1 μL of sequencing buffer (5×), 4 μL of water, and 2 μL of BigDye® Terminator v3.1 (Applied Biosystems, Foster City, CA, United States). Thermal cycling conditions were as followed: preheating at 96°C, 1 min and 35 amplification cycles at 96°C for 10 s, 50°C for 5 s, 60°C for 4 min, and a final extension at 60°C for 1 m. Finally, the products were analyzed on an ABI 3130 DNA Sequencer (Applied Biosystems, Foster City, CA, United States). HBV DNA sequences were genotyped by phylogenetic analysis using the Molecular Evolutionary Genetic Analysis software, MEGA 10.0[1]. The reference sequences used in this study were: X02763, X51970, and AF090842 for genotype A; AF100309 and AB033554 for genotype B; AB014381, X04615, and AY123041 for genotype C; M32138, X65259, and X85254 for genotype D; X75657 and AB032431 for genotype E; X69798, AF223965, and AB036910 for genotype F; AB064310 and AB625342 for genotype G; and AY090460, AB516395, and AB516394 for genotype H. ClustalW method was used for multiple sequence alignment, while Maximum Likelihood method was used to construct the phylogenetic tree with a bootstrap test of 1000 replicates. The substitution model was General Time Reversible with Gamma Distribution and Invariants Sites (GTR + G + I). Accession numbers submitted to GenBank[2] of the HBV samples identified in this study appear as MH780886-MH780888, MF150683 for genotype A2; MH780889-MH780897, MF150677 for genotype G; MH780898-MH780909, MF150649-MF150650, MF150658-MF150659, MF150666-MF150669, MF150672-MF150673 for genotype H; MF150685 for genotype D4 and MF150688 for genotype F1b. An alternative sequencing strategy was required for the F1b and D4 sub-genotypes (Mendes-Correa et al., 2010). This strategy was necessary as the D multiplex products were untypeable by direct sequencing. While, for the F1b sub-genotype, we did not expect to find genotype F in Mexico.

[1]http://www.megasoftware.net/

[2]http://www.ncbi.nlm.nih.gov

Assessment of Liver Damage

Liver fibrosis was assessed through three non-invasive methods: transitional elastography (TE), Aspartate aminotransferase (AST) to Platelets Ratio Index score (APRI), and Fibrosis-4 score (FIB-4). TE was carried out using a FibroScan® instrument (Echosens, Paris, France). Liver stiffness measurement (LSM) expressed in kiloPascals (kPa) was performed by the same qualified physician. LSM was considered reliable if ten successful measurements were obtained, and the success rate was >80%. APRI was calculated considering the upper limit of normal (ULN) serum AST for men equal to 40 IU/in the equation AST *IU/L*/(ULN AST/(Platelets count 10^9 *L*) ×100 (Sandrin et al., 2003; Borsoi Viana et al., 2009). FIB-4 was calculated with the equation (Age-*years*) (AST level *IU/L*)/(Platelets count 10^9*L*) (ALT)$^{1/2}$ *IU/L*) (Kim et al., 2010). APRI cut-off value ≥ 0.7 was considered significant hepatic fibrosis, while FIB-4 cut-off value ≥ 3.25 was interpreted as advanced fibrosis (Sterling et al., 2006; Lin et al., 2011). Also, LSM was classified as F1 (<7.1 kPa), F2 (≥7.1 kPa), F3 (≥9.3 kPa), and F4 or cirrhosis (≥12.3 kPa). F1-F2 was taken as mild liver fibrosis and F3-F4 as advanced liver fibrosis (Hawkins et al., 2013). Levels of liver enzymes, platelets, HIV viral load, CD4, and CD8 lymphocytes were attained by routine analysis from the hospital's Central Laboratory.

Statistical Analysis

Continuous variables were expressed as the median and interquartile range, while categorical variables were expressed in frequency or percentage. Categorical variables were compared using the $\chi 2$ test or Fisher's exact test where appropriate. Data normality was calculated using the Shapiro–Wilk test. Non-parametric continuous variables were tested by Mann–Whitney or Kruskal–Wallis test. All correlations were tested with Spearman's Rho, and the cut-off of quantitative values was calculated with Receiver Operating Characteristic (ROC) curve analysis. All statistical tests were performed with the Statistical Program for Social Sciences software (SPSS 22.0, IBM, Inc., Armonk, NY, United States). For all statistical tests, a *P*-value of <0.05 was considered significant.

Ethics

The Ethics Committee of the University Center for Health Sciences of the University of Guadalajara approved this protocol (#CI-07218). Furthermore, this study was carried out by the principles of the Helsinki Declaration.

RESULTS

General Characteristics of the Study Population

A total of 228 patients with HIV were analyzed; most of them were men (86.4%, 197/228) between 32 and 46 years of age (**Table 1**). At the time of this study, 82.0% (187/228)

TABLE 1 | Demographic data of patients with HIV and HIV/HBV.

	Total (*N* = 228)	HIV alone (*N* = 161)	HIV/HBV (*N* = 67)	*p*-value
Gender				
Female	31 (13.6%)	30 (18.6%)	1 (1.5%)	<0.001[1]
Male	197 (86.4%)	131 (81.4%)	66 (98.5%)	
Age (years)				
Mean (SD)	39.9 (10.9)	39.7 (11.7)	39.2 (9.7)	0.758[2]
Age group (years)				
<15	1 (0.4%)	1 (0.6%)	0 (0.0%)	1.000[3]
15–20	4 (1.8%)	4 (2.5%)	0 (0.0%)	0.322[3]
21–25	10 (4.5%)	7 (4.4%)	3 (4.5%)	1.000[3]
26–30	32 (14.3%)	23 (14.6%)	9 (13.6%)	1.000[3]
31–35	41 (18.3%)	29 (18.4%)	12 (18.2%)	1.000[3]
36–40	47 (21.0%)	29 (18.4%)	18 (27.3%)	0.151[3]
41–45	29 (12.9%)	21 (13.3%)	8 (12.1%)	1.000[3]
46–50	24 (10.7%)	17 (10.8%)	7 (10.6%)	1.000[3]
51–55	17 (7.6%)	11 (7.0%)	6 (9.1%)	0.789[3]
56–60	9 (4.0%)	8 (5.1%)	1 (1.5%)	0.288[3]
61–65	5 (2.2%)	4 (2.5%)	1 (1.5%)	1.000[3]
>66	5 (2.2%)	4 (2.5%)	1 (1.5%)	1.000[3]

[1]*Pearson's Chi-squared test.*
[2]*T-student test.*
[3]*Fisher Exact test.*
Age had a normal distribution. SD, standard deviation; HBV, hepatitis B virus; HIV, human immunodeficiency virus.

of the participants were on antiretroviral therapy, with a median HIV viral load of 40.0 IU/mL [Interquartile range (IQR): 39.0–326.0 IU/mL]. Among the 187 treated patients, the medical records revealed that cART was tailored individually in which none were monotherapy. All received at least one drug that is known to affect the HBV's life cycle, such as tenofovir, lamivudine, or emtricitabine. However, the most frequent combinations were tenofovir + emtricitabine + efavirenz in 39.6% (74/187), followed by tenofovir + emtricitabine + lopinavir + ritonavir in 11.2% (21/187) and 49.2% (92/187) received diverse unique combinations with these and other antiviral drugs. Also, 62 individuals out of 132 had evidence of hepatitis C virus (HCV) infection (**Table 2**). Overall, liver injury was detected in 24.6% (31/126), of which 18.3% (24/131) had significant liver fibrosis, 6.9% (9/130) had advanced fibrosis, and with liver stiffness, 22.5% (11/71) had advanced liver fibrosis (F3-F4) (**Table 3**).

HBV Infection Status

Of the 228 HIV individuals, 67 were positive for HBsAg; these were considered HBV/HIV co-infected patients (**Table 2**). HIV/HBV co-infection cases were detected between 21 and 25 years, while the maximum age was reached in persons between 36 and 40 years. Overall, the HBV viral load was detected in 71.6% (48/67) of the cases, with a median HBV viral load of 145.0 IU/mL (IQR: 20.0–166616.0 IU/mL). A low HBV viral load (<2,000 IU/mL) was detected in 66.6% (32/48) of HBV/HIV patients, while 33.3% (16/48) had a high HBV viral load (≥2,000 IU/mL). In this study, 89.1% (57/67) of these HBsAg positive cases were under antiretroviral treatment. Among these, 43.9% (25/57) were treated with tenofovir + emtricitabine + efavirenz, followed by 10.5% (6/57) receiving tenofovir + emtricitabine + lopinavir + ritonavir, and 45.6% (26/57) received as mentioned above several unique combinations with other drugs. The liver enzyme profile was similar between HIV alone (without HBV or HCV) and HBV/HIV patients. However, ALT levels were higher in HBV/HIV co-infection than HIV alone (30 IU/L (IQR: 22 – 52.5 IU/L) vs. 24.5 (IQR: 17.0–36.2), p = 0.006). Among the HBV/HIV cases, the frequency of liver inflammation was 18.2% (10/55), 15.3% (9/59) had significant liver fibrosis, 5.2% (3/58) had advanced fibrosis, and with LSM, 19.1% (9/47) had advanced liver fibrosis (F3-F4) (**Table 3**).

Distribution of HBV Genotypes and Mixtures

Among the 48 patients with detectable HBV viral load, 25 cases were positive to the nested PCR assay in which 44 HBV genotypes were detected by multiplex PCR (**Figures 2A, B**). According to our methodology, the minimum viral load for HBV genotyping was 82.0 IU/mL [Area Under the Curve, (AUC) = 0.93, $p = 5 \times 10^{-5}$] (**Supplementary Figure 1A**). The predominant HBV genotype was H (50.0%, 22/44), followed by G (22.7%, 10/44), D (15.9%, 6/44), A (9.1%, 4/44), and F (2.3%, 1/44). **Figure 2C** illustrates the phylogenetic tree built for each genotype of the 38 strains confirmed by direct DNA sequencing. All genotypes A were sub-genotyped as A2, one genotype F was classified as F1b, and one D was D4. Among typeable samples, 44.0% (11/25) had a single genotype (10 were H and one G), whereas 36.0% (9/25) had dual-mixtures (four were G/H), two were D/H, two were A2/H, and one was A2/G) Moreover, 20.0% (5/25) were triple-mixtures (three were H/G/D, one was A2/D/G, and one was H/D/F1b). G/H (44.4%, 4/9) and H/G/D (60.0%, 3/5) were the most frequent genotype mixtures. Nineteen out of 25 genotyped cases were under cART in which five mono-infected, three dual-mixed, and one triple-mixed were treated with tenofovir + emtricitabine + efavirenz. Two triple-mixed cases were treated with tenofovir + emtricitabine + lopinavir + ritonavir whereas three mono-infected, four dually infected and one triple-mixed case received diverse combinations. Interestingly, ten patients had high HBV viral loads (≥2,000 IU/mL) despite receiving a tenofovir-based regimen; five were mono-infected, (four H and one G), one dual-mixed (A/G), and four were triple-mixed (two D/G/H, one A/D/G, and one D/H/F1b).

Relationship Between HBV Mixtures and Clinical Characteristics

We compared different clinical parameters as dependent variables relative to the number of HBV genotypes (**Figures 3A–C** and **Supplementary Table 2**), finding a positive correlation between increasing age and mixtures (Spearman's Rho = 0.53, p = 0.0069) (**Figure 3A**). Interestingly, the median age of patients with two HBV genotypes was 37.0 years, while the age-associated with triple-mixed cases was ≥42.0 years (AUC = 0.805, p = 0.038)

TABLE 2 | Baseline clinical characteristics of patients with HIV and HIV/HBV.

	Total (*N* = 228)	HIV alone (*N* = 161)	HIV/HBV (*N* = 67)	*p*-value
cART				
Without cART	32 (14.0%)	25 (15.5%)	7 (10.4%)	0.371[1]
With cART	196 (86.0%)	136 (84.5%)	60 (89.6%)	
HBV viral load (IU/mL)				
Median (Q1, Q3)	145.0 (20.0, 18312.5)	NA	145.0 (20.0, 18312.5)	
HIV viral load (IU/mL)				
Median (Q1, Q3)	40.0 (39.5, 303.5)	40.0 (40.0, 388.0)	40.0 (34.8, 137.5)	0.520[2]
HCV infection				
Negative	132 (68.0%)	73 (55.7%)	59 (93.7%)	<0.001[1]
Positive	62 (32.0%)	58 (44.3%)	4 (6.3%)	
ALT IU/L				
Median (Q1, Q3)	27.0 (20.0, 44.0)	24.5 (17.0, 36.2)	30.0 (22.0, 52.5)	0.006[2]
AST IU/L				
Median (Q1, Q3)	30.0 (24.0, 43.0)	30.5 (24.0, 48.5)	29.0 (24.0, 40.0)	0.734[2]
GGT IU/L				
Median (Q1, Q3)	41.0 (24.2, 61.8)	44.0 (29.0, 82.0)	34.0 (24.0, 48.0)	0.126[2]
ALP IU/L				
Median (Q1, Q3)	97.0 (80.2, 126.0)	95.0 (73.0, 135.0)	100.0 (85.0, 119.5)	0.996[2]
Platelets cells/μL				
Median (Q1, Q3)	226.0 (181.0, 293.0)	232.0 (179.0, 309.0)	224.5 (187.5, 281.0)	0.697[2]
Albumin g/dL				
Median (Q1, Q3)	3.8 (3.2, 4.1)	3.6 (2.6, 3.9)	3.9 (3.7, 4.1)	<0.001[2]
CD4 cells/mm3				
Median (Q1, Q3)	299.0 (179.2, 547.8)	273.0 (92.2, 447.2)	339.0 (236.5, 607.5)	0.007[2]
CD8 cells/mm3				
Mean (SD)	954.7 (553.4)	908.3 (555.2)	1039.0 (565.2)	0.129[3]

[1] *Pearson's Chi-squared test.*
[2] *Wilcoxon-Mann-Whitney test.*
[3] *T-student test.*
Patients with HCV were excluded for the comparison of laboratory data. CD8 levels had a normal distribution. NA, not available; SD, standard deviation; cART, combination antiretroviral therapy; HBV, hepatitis B virus; HCV, hepatitis C virus; HIV, human immunodeficiency virus. ALT, alanine aminotransferase; AST, aspartate aminotransferase; GGT, gamma-glutamyl transferase; ALP, alkaline phosphatase; CD4, lymphocyte CD4 count; CD8, lymphocyte CD8 count; Q1, quartile 1; Q3, quartile 3.

(**Supplementary Figure 1B**). Moreover, platelet levels were negatively correlated with the increase in the number of HBV genotypes (Spearman's Rho = − 0.416, $p = 0.039$) (**Figure 3B**). HBV viral load (31623.0 IU/mL vs. 1479.0 IU/mL, $p = 0.029$) was higher in triple-mixtures than dual-mixtures (**Figure 3C**). Regarding liver damage, the degree of fibrosis was similar between single and dual genotype mixtures (LSM, $p = 0.34$, FIB-4, $p = 0.97$), while the liver fibrosis grade was significantly higher in triple-mixtures than dual-mixtures (APRI, $p = 0.029$; FIB-4, $p = 0.007$) (**Figures 3D–F**). Based on APRI, patients with HIV and three HBV genotypes had a high risk of presenting significant liver fibrosis (OR = 15.0 95%CI = 1.29 – 174.38, $p = 0.027$) compared with dual-mixtures or single genotype infection (**Table 4**).

DISCUSSION

Previously, we reported the frequency of HBV infection and its risk factors associated with co-infection in this HIV cohort (Jose-Abrego et al., 2017). However, the frequency of liver fibrosis and the proportion of patients with HBV genotype mixtures remained unknown. In this study, the assessment of LSM with transitional elastography detected more cases of liver fibrosis than APRI or FIB-4. Overall, the frequency of advanced liver fibrosis in Mexican patients with HIV was 22.5%, while among HBV/HIV co-infected patients were 19.1%. This proportion was slightly lower than that found in two HBV/HIV co-infected Nigerian studies with frequencies of 22.1 and 22.6% each. Using liver biopsy, Sterling and colleges reported 22% of advanced fibrosis in populations from the United States and Canada (Hawkins et al., 2013; Grant et al., 2017; Sterling et al., 2019). The high frequency of advanced liver fibrosis may be due to inadequate evaluation and diagnostics of hepatitis B in our population. Furthermore, the high frequency of asymptomatic and OBI (Torres-Baranda et al., 2006; Enriquez-Navarro et al., 2020) may facilitate the non-detection of co-infections with other hepatotropic viruses. Alcohol consumption, drug abuse, and potential hepatotoxicity of cART drugs can also be involved (Neff et al., 2006). In regards to liver enzymes, significantly higher levels of ALT but not AST or GGT were found in the HIV group compared to the HIV/HBV co-infected patients. This finding may be due to an insufficient

TABLE 3 | Baseline non-invasive markers of liver damage in HIV and HBV/HIV patients.

	Total (*N* = 228)	HIV alone (*N* = 161)	HIV/HBV (*N* = 67)	*p*-value
APRI value				
Median (Q1, Q3)	0.3 (0.2, 0.5)	0.3 (0.2, 0.6)	0.3 (0.2, 0.5)	0.921[2]
FIB-4 value				
Median (Q1, Q3)	0.9 (0.7, 1.4)	1.1 (0.7, 1.6)	0.9 (0.7, 1.3)	0.407[2]
Liver stiffness (Kpa)				
Median (Q1, Q3)	6.1 (4.8, 8.6)	6.1 (4.5, 9.5)	6.8 (5.0, 8.2)	0.766[2]
Liver inflammation				
ALT or AST or GGT $<$ 40 IU/mL	95 (75.4%)	50 (70.4%)	45 (81.8%)	0.141[1]
ALT or AST or GGT $\geq$ 40 IU/mL	31 (24.6%)	21 (29.6%)	10 (18.2%)	
APRI class				
Without significant liver fibrosis (<0.7)	107 (81.7%)	57 (79.2%)	50 (84.7%)	0.411[1]
With significant liver fibrosis (≥0.7)	24 (18.3%)	15 (20.8%)	9 (15.3%)	
FIB-4 class				
Without advanced fibrosis (<3.25)	121 (93.1%)	66 (91.7%)	55 (94.8%)	0.480[1]
With advanced fibrosis (≥3.25)	9 (6.9%)	6 (8.3%)	3 (5.2%)	
Liver stiffness class				
F1	44 (62.0%)	16 (66.7%)	28 (59.6%)	0.614[3]
F2	11 (15.5%)	1 (4.2%)	10 (21.3%)	0.262[3]
F3	5 (7.0%)	4 (16.7%)	1 (2.1%)	0.042[3]
F4	11 (15.5%)	3 (12.5%)	8 (17.0%)	0.739[3]

[1] *Pearson's Chi-squared test.*
[2] *Wilcoxon-Mann-Whitney test.*
[3] *Fisher Exact test.*
Patients with HCV were excluded from the analyses. HBV, hepatitis B virus; HIV, human immunodeficiency virus; ALT, alanine aminotransferase; AST, aspartate aminotransferase; GGT, gamma-glutamyl transferase; APRI, AST to Platelet Ratio Index; FIB-4, fibrosis-4 score; Q1, quartile; Q3, quartile 3.

sample size per group to achieve statistical differences. Another reason from a clinical viewpoint, it is known that altered ALT levels or transaminitis is a marker of liver damage than AST whereas abnormal GGT is seen in cases of alcoholic liver disease. In the HIV mono-infected group, high ALT levels could be due to several factors such as medication-related hepatotoxicity, alcohol abuse, drug use, non-alcoholic steatohepatitis or HIV itself.

Our results provide evidence for reconstructing a part of HBV infection's natural history in patients with HIV. It is known that HIV patients are infected near the age of 15 (Wilson et al., 2010). Later, due to injection drug use or high-risk sexual behavior including men who have sex with men (MSM) (Jose-Abrego et al., 2017), they are infected with hepatitis B between 21 and 25 years. In our cohort, no patient was positive to HBsAg before the age of 20, the first cases of HBV/HIV co-infection (4.5%) occurred around 21, then the maximum frequency of HBV/HIV co-infection was reached at 36–40 years (27.3%). These data suggest that HIV infection occurs before HBV infection. Future longitudinal studies are necessary to demonstrate or refute this observation. Nonetheless, if patients co-infected with HBV/HIV do not change their risk factors, they could acquire HBV genotypes mixtures. Furthermore, our findings show that HIV patients could be co-infected with two HBV genotypes near the age of 37 and up to three HBV genotypes at older ages (42 years). Using PCR multiplex and sequencing, we found that 56% of HIV patients with typeable HBV were infected with more than one HBV genotype. This frequency is higher than those reported in Swiss people and the Netherlands, with 14 and 37%, respectively (Van Der Kuyl et al., 2013; Bihl et al., 2015). This difference is mainly due to our study population's characteristics, who are people with very low-income. Many of them are sex workers, and most do not have access to social security (Jose-Abrego et al., 2017). Another difference with those studies is the type of mixture. A/G mixtures predominate in European HBV/HIV co-infected people, whereas H/G mixtures were predominant in our results. The reason is that genotype H is endemic to Mexico, and genotype G is strongly related to the MSM population as well (Panduro et al., 2013; Roman and Panduro, 2013). Also, we identified five HIV patients who were infected with three HBV genotypes. This evidence indicates that HBV genotype mixtures may be more common than expected in populations with HIV from Mexico and Europe.

In patients without HIV, the mixture of HBV genotypes has been associated with different clinical outcomes, especially high viral load levels (Toan et al., 2006). In this HBV/HIV cohort, high viral load levels predominate in patients with three genotypes than those with dual or single genotype infection. We did not find a clear association with viral load in patients with two genotypes, probably due to tenofovir's or lamivudine's antiviral effect. The molecular mechanism explaining the high HBV viral load in patients with mixtures has been documented in *in vitro* models showing that different genotypes interact to increase their replication. For example, genotype G uses the core promoter and core protein from sub-genotype A2 to replicate efficiently. Also, genotype G is abundantly expressed when co-transfected with a construct

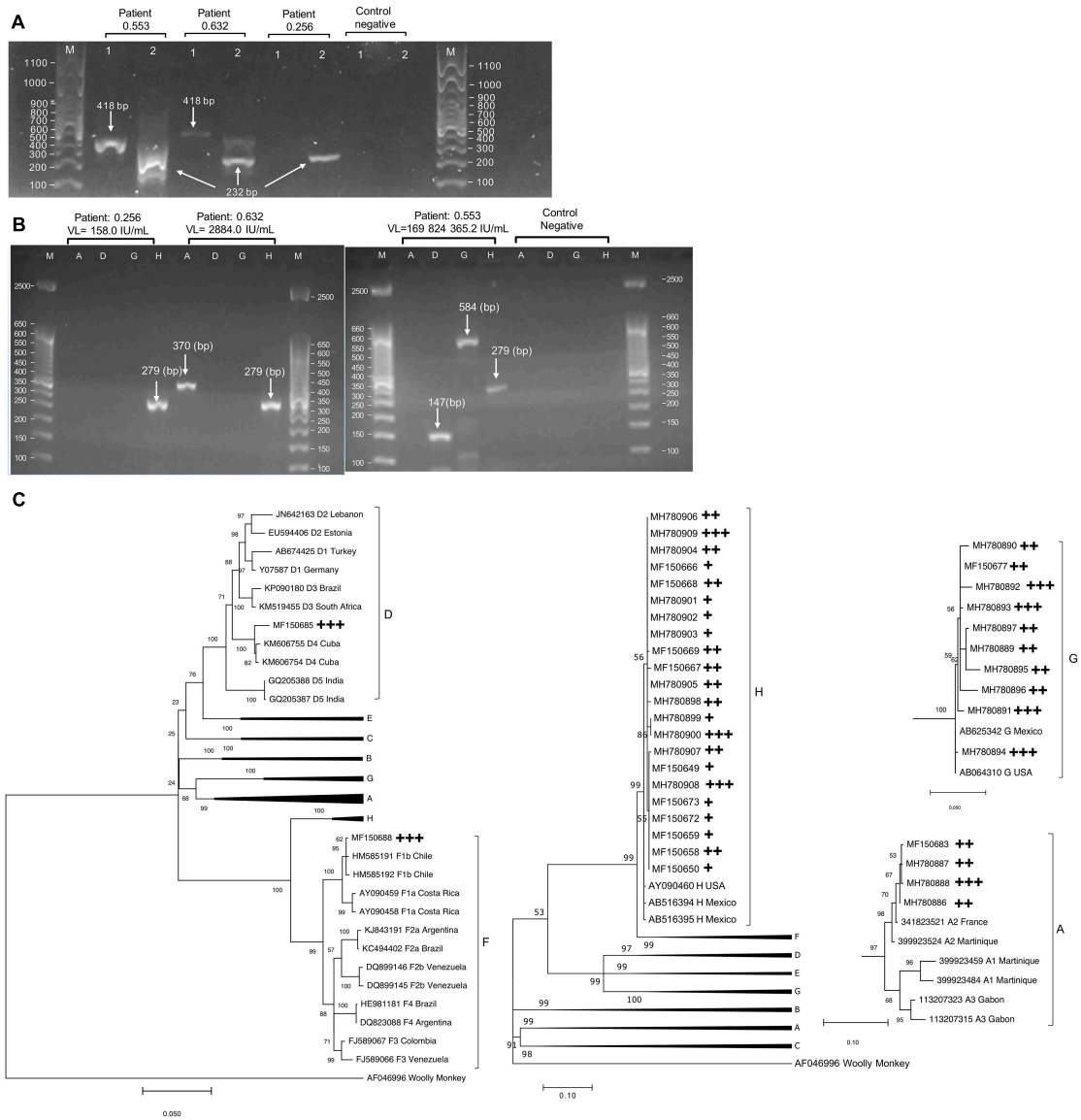

FIGURE 2 | (A) Different size bands were expected for the nested PCR. In patients with high viral loads, both bands' presence is evident (418 bp and 232 bp), but only the second band is visualized in patients with relatively low viral load. **(B)** Multiplex PCR results, after the molecular weight marker, the first four lanes show a patient's results with one HBV genotype (patient 0.256), while the next four lines show a mixture of two HBV genotypes (patient 0.632). Finally, the lanes of patient 0.553 show a typical result of a person infected with three HBV genotypes. Based on the band intensity, we noted that each HBV genotype has a different viremia level, and one genotype is usually dominant over others. **(C)** Phylogenetic trees are used for genotyping and sub-genotyping of sequenced multiplex PCR products. The numbers in the trees represent the bootstrap values of each node. Since different genomic locations classified each genotype, four phylogenetic trees were built. The cross symbol indicates the type of infection in which the genotype was detected: +, single-infected sample; ++, dual-infection sample; +++, triple-infection sample.

that expresses the genotype D core protein. D. This occurs because the core promoter of genotype G cannot generate enough core protein for its replication and it may also be defective, affecting the synthesis of virions (Sakamoto et al., 2013). It is plausible that genotype H uses the same mechanism since the replication of genotype G in chimeric mice is enhanced when co-infected with genotype H (Tanaka et al., 2008). Interestingly, our study found patients with G/H, G/A2, and H/G/D mixtures suggesting that in some HIV patients, the high HBV viral load may be due to HBV genotypes mixtures.

This study found a negative correlation between platelet levels and the increase of HBV genotypes; this is understandable as thrombocytopenia is one of the most frequent complications in patients with advanced liver disease (Nusrat et al., 2014). Also, we associated the presence of three HBV genotypes with significant liver fibrosis. The liver damage linked to HBV mixtures could be explained from an immunological viewpoint.

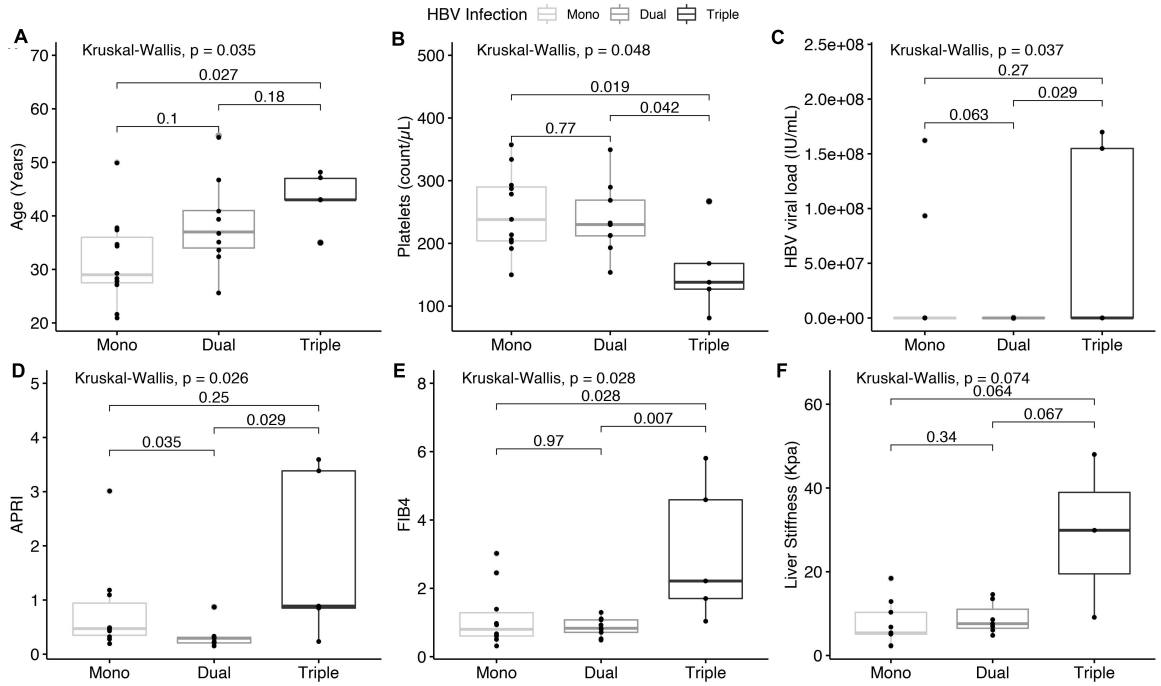

FIGURE 3 | (A) Shows a significant relationship between increasing age and the presence of a mixture of HBV genotypes. **(B)**, Platelet levels decreased while increasing the number of HBV genotypes. **(C)** Shows viral load levels by type of mixes. **(D–F)** Shows the comparison of the degree of fibrosis in patients with one, two, and three HBV genotypes, evaluated through the APRI, FIB-4, and LSM, respectively. The boxes' middle lines represent the median, while the outer lines represent 25 and 75 quartiles. Black dots indicate each patient studied. LSM, liver stiffness measurement.

TABLE 4 | Association between the number of HBV genotypes and liver fibrosis in patients with HIV/HBV co-infection.

	APRI			FIB-4			LSM		
	Without significant liver fibrosis (<0.7)	**With significant liver fibrosis (≥0.7)**	**Odds ratio (95% CI) *p*-value[1]**	**Without advanced fibrosis (<3.25)**	**With advanced fibrosis (≥3.25)**	**Odds ratio (95% CI) *p*-value[1]**	**Mild liver fibrosis (F1-F2)**	**Advanced liver fibrosis (F3-F4)**	**Odds ratio (95% CI) *p*-value[1]**
HBV genotypes									
Single *n* (%)	7 (43.8)	3 (37.5)	0.771 (0.136–4.391) $p = 0.561$	10 (45.5)	0 (0.0)	0.857 (0.692–1.062) $p = 0.330$	6 (50.0)	3 (42.9)	0.750 (0.115–4.898) $p = 0.570$
Dual *n* (%)	8 (50.0)	1 (12.5)	0.143 (0.014–1.444) $p = 0.087$	9 (40.9)	0 (0.0)	0.867 (0.711–1.057) $p = 0.380$)	5 (41.7)	2 (28.6)	0.560 (0.076–4.144) $p = 0.474$
Triple *n* (%)	1 (6.3)	4 (50)	15.0 (1.290–174.386) $p = 0.027$	3 (13.6)	2 (100.0)	1.667 (0.815–3.409) $p = 0.036$)	1 (8.3)	2 (28.6)	4.4 (0.319–60.614) $p = 0.296$)

[1] *Fisher Exact test.*

APRI, AST to Platelet Ratio Index; FIB-4, fibrosis-4 score; LSM, liver stiffness measurement.

First, the increase of viral load due to interaction among HBV genotypes can activate resident liver macrophages. Then, these cells release cytokines, which recruit more macrophages and induce hepatic stellate cell activation. Finally, these signs stimulate the extracellular matrix's proliferation and deposition, triggering fibrosis development (Koyama and Brenner, 2017). If the immune response or antiviral therapy does not control the infection, people with a mixture of HBV genotypes could eventually develop cirrhosis or liver cancer. In this context, our results indicate that HIV patients may be living with two or more genotypes that could have been acquired simultaneously or through successive contacts. This data is relevant because HBV mono-infected individuals (without HIV), such as heterosexual individuals with multiple partners, people born in HBV endemic areas, or patients under hemodialysis (Trepo et al., 2014), could also present a mixture of HBV genotypes. Perhaps, these groups may be at risk of developing liver damage related to mixtures, as in our HIV cohort. This highlights the importance of assessing the presence of a mixture of HBV genotypes in other risk groups.

Another clinical importance of HBV genotype mixtures is that it can impact treatment outcomes. During treatment, lamivudine, emtricitabine, or tenofovir may only suppress sensitive HBV genotypes, favoring viral persistence due to antiviral resistance mutations (Mendes-Correa et al., 2010; Singh et al., 2017). In

this study, four triply-infected patients and one dually infected case had high HBV viral load (≥2,000 IU/mL) despite receiving tenofovir combined with lamivudine or emtricitabine. This finding suggests that in some HBV/HIV co-infected patients, virological failure may be caused by the presence of mixtures of HBV genotypes. However, in this study, the frequency of potential resistance mutations was not evaluated. Thus, further research will be necessary to increase the number of mixed-genotype cases and evaluate the presence of antiviral resistance mutations.

In addition to the risk of fibrosis, high viral loads, and the emergence of resistance mutations, HBV genotype mixtures may play a key role in the origin of new viral strains. This event can occur when a cell is infected with different HBV genotypes that exchange genetic information, resulting in a recombinant genotype (Simmonds and Midgley, 2005). Whole-genome studies have reported HBV recombinant genotype infection in China (Liu et al., 2017), South Africa (Matlou et al., 2019), Japan (Kojima et al., 2015), Argentina, and Brazil (Araujo et al., 2013), particularly in patients with HIV. Currently, a total of 24 phylogenetic recombinant forms are recognized (Simmonds and Midgley, 2005). If recombination confers replicative and spreading advantages, this new virus could fixate in a population and follow its evolutionary path independent of its precursor genotypes, giving rise to new HBV genotypes.

On the other hand, HBV genotypes found in mixtures could accelerate the appearance of mutations (Lok et al., 2000). This event could be due to antivirals that rapidly change the intracellular environment and inhibit viral replication. In response to this selective pressure, HBV genotypes accumulate mutations in the reverse transcriptase domain of the polymerase gene that allow them to change their protein structure and escape antivirals (Locarnini and Yuen, 2010). These features suggest that HBV evolution can occur very quickly when natural selection is robust. Under this approach, the presence of mixtures of HBV genotypes would have biological importance, functioning as a mechanism of HBV diversity.

CONCLUSION

In conclusion, a high frequency of advanced liver fibrosis was detected in Mexican patients with HIV. Among the HBV/HIV co-infected, we confirmed the presence of a mixture of HBV genotypes. This infection type was associated with significant liver fibrosis and high viral load levels, especially when three HBV genotypes were detected.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/genbank/, MH780886-MH780897 https://www.ncbi.nlm.nih.gov/genbank/, MF150683 https://www.ncbi.nlm.nih.gov/genbank/, MF150677 https://www.ncbi.nlm.nih.gov/genbank/, MH780898-MH780909 https://www.ncbi.nlm.nih.gov/genbank, MF150649-MF150650 https://www.ncbi.nlm.nih.gov/genbank/, MF150658-MF150659 https://www.ncbi.nlm.nih.gov/genbank/, MF150666-MF150669 https://www.ncbi.nlm.nih.gov/genbank/, MF150672-MF150673 https://www.ncbi.nlm.nih.gov/genbank/, MF150685-MF150688.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Commitee of the University Center for Health Sciences of the University of Guadalajara (#CI-07218). Written informed consent to participate in this study was provided by the participants or the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

AP and SR: conceptualization. AP, SR, JR, and VFDC: methodology. VFDC: validation. AJ-A, SR, and AP: formal analysis and investigation. AJ-A: data curation and writing-original draft preparation. SR, AP, and JR: writing-review and editing. SR: supervision. AP: funding acquisition. All authors contributed to the article and approved the submitted version.

FUNDING

This research was partly funded by the Consejo Nacional de Ciencia y Tecnología (CONACYT), Grant Number CONACYT PN-2017-01-5254 to AP and Programa de Apoyo a la Incorporación de NPTC No. UDG-PTC-1439 to AJ-A. This work was part of AJ-A thesis to obtain the Doctor in Sciences degree in Molecular Biology in Medicine, University of Guadalajara.

ACKNOWLEDGMENTS

The authors would like to thank the HIV Clinic of Civil Hospital of Guadalajara, Fray Antonio Alcalde for their kind support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.640889/full#supplementary-material

Supplementary Figure 1 | Analysis to calculate the optimal viral load for HBV genotyping **(A)** and cut-off age for detecting triple-mixed infection **(B)** in patients with HIV. SP, specificity; SE, sensitivity.

Supplementary Table 1 | Sequence of primers used in nested PCR, multiplex PCR and DNA sequencing.

Supplementary Table 2 | Characteristics of patients with and without HBV genotype mixtures.

REFERENCES

Alvarez-Munoz, M. T., Maldonado-Rodriguez, A., Rojas-Montes, O., Torres-Ibarra, R., Gutierrez-Escolano, F., Vazquez-Rosales, G., et al. (2014). Occult hepatitis B virus infection among Mexican human immunodeficiency virus-1-infected patients. *World J. Gastroenterol.* 20, 13530–13537. doi: 10.3748/wjg.v20.i37.13530

Araujo, N. M., Araujo, O. C., Silva, E. M., Villela-Nogueira, C. A., Nabuco, L. C., Parana, R., et al. (2013). Identification of novel recombinants of hepatitis B virus genotypes F and G in human immunodeficiency virus-positive patients from Argentina and Brazil. *J. Gen. Virol.* 94, 150–158. doi: 10.1099/vir.0.047324-0

Bihl, F., Martinetti, G., Wandeler, G., Weber, R., Ledergeber, B., Calmy, A., et al. (2015). HBV genotypes and response to tenofovir disoproxil fumarate in HIV/HBV-coinfected persons. *BMC Gastroenterol.* 15:79. doi: 10.1186/s12876-015-0308-0

Borsoi Viana, M. S., Takei, K., Collarile Yamaguti, D. C., Guz, B., and Strauss, E. (2009). Use of AST platelet ratio index (APRI Score) as an alternative to liver biopsy for treatment indication in chronic hepatitis C. *Ann. Hepatol.* 8, 26–31.

Chen, B. F., Chen, P. J., Jow, G. M., Sablon, E., Liu, C. J., Chen, D. S., et al. (2004). High prevalence of mixed genotype infections in hepatitis B virus infected intravenous drug users. *J. Med. Virol.* 74, 536–542. doi: 10.1002/jmv.20211

Cunningham, E. B., Applegate, T. L., Lloyd, A. R., Dore, G. J., and Grebely, J. (2015). Mixed HCV infection and reinfection in people who inject drugs–impact on therapy. *Nat. Rev. Gastroenterol. Hepatol.* 12, 218–230. doi: 10.1038/nrgastro.2015.36

Di Martino, V., Thevenot, T., Colin, J. F., Boyer, N., Martinot, M., Degos, F., et al. (2002). Influence of HIV infection on the response to interferon therapy and the long-term outcome of chronic hepatitis B. *Gastroenterology* 123, 1812–1822. doi: 10.1053/gast.2002.37061

Enriquez-Navarro, K., Maldonado-Rodriguez, A., Rojas-Montes, O., Torres-Ibarra, R., Bucio-Ortiz, L., De La Cruz, M. A., et al. (2020). Identification of mutations in the S gene of hepatitis B virus in HIV positive Mexican patients with occult hepatitis B virus infection. *Ann. Hepatol.* 19, 507–515. doi: 10.1016/j.aohep.2020.06.002

Grant, J., Agbaji, O., Kramvis, A., Yousif, M., Auwal, M., Penugonda, S., et al. (2017). Hepatitis B virus sequencing and liver fibrosis evaluation in HIV/HBV co-infected Nigerians. *Trop. Med. Int. Health* 22, 744–754. doi: 10.1111/tmi.12873

Hannoun, C., Krogsgaard, K., Horal, P., Lindh, M., and Interpred Trial, G. (2002). Genotype mixtures of hepatitis B virus in patients treated with interferon. *J. Infect. Dis.* 186, 752–759. doi: 10.1086/342599

Hawkins, C., Agbaji, O., Ugoagwu, P., Thio, C. L., Auwal, M. M., Ani, C., et al. (2013). Assessment of liver fibrosis by transient elastography in patients with HIV and hepatitis B virus co-infection in Nigeria. *Clin. Infect. Dis.* 57, e189–e192. doi: 10.1093/cid/cit564

Jose-Abrego, A., Panduro, A., Fierro, N. A., and Roman, S. (2017). High prevalence of HBV infection, detection of subgenotypes F1b, A2, and D4, and differential risk factors among Mexican risk populations with low socioeconomic status. *J. Med. Virol.* 89, 2149–2157. doi: 10.1002/jmv.24913

Kim, B. K., Kim, D. Y., Park, J. Y., Ahn, S. H., Chon, C. Y., Kim, J. K., et al. (2010). Validation of FIB-4 and comparison with other simple noninvasive indices for predicting liver fibrosis and cirrhosis in hepatitis B virus-infected patients. *Liver Int.* 30, 546–553. doi: 10.1111/j.1478-3231.2009.02192.x

Kirschberg, O., Schuttler, C., Repp, R., and Schaefer, S. (2004). A multiplex-PCR to identify hepatitis B virus–enotypes A-F. *J. Clin. Virol.* 29, 39–43. doi: 10.1016/s1386-6532(03)00084-2

Kojima, Y., Kawahata, T., Mori, H., Furubayashi, K., Taniguchi, Y., Itoda, I., et al. (2015). Identification of novel recombinant forms of hepatitis B virus generated from genotypes Ae and G in HIV-1-positive japanese men who have sex with men. *AIDS Res. Hum. Retroviruses* 31, 760–767. doi: 10.1089/aid.2014.0281

Koyama, Y., and Brenner, D. A. (2017). Liver inflammation and fibrosis. *J. Clin. Invest.* 127, 55–64. doi: 10.1172/JCI88881

Lin, Z. H., Xin, Y. N., Dong, Q. J., Wang, Q., Jiang, X. J., Zhan, S. H., et al. (2011). Performance of the aspartate aminotransferase-to-platelet ratio index for the staging of hepatitis C-related fibrosis: an updated meta-analysis. *Hepatology* 53, 726–736. doi: 10.1002/hep.24105

Liu, T., Wang, F., Zhang, S., Wang, F., Meng, Q., Zhang, G., et al. (2017). Whole-gene analysis of two groups of hepatitis B virus C/D inter-genotype recombinant strains isolated in Tibet. China. *PLoS One* 12, e0179846. doi: 10.1371/journal.pone.0179846

Locarnini, S. A., and Yuen, L. (2010). Molecular genesis of drug-resistant and vaccine-escape HBV mutants. *Antivir. Ther.* 15, 451–461. doi: 10.3851/IMP1499

Lok, A. S., Hussain, M., Cursano, C., Margotti, M., Gramenzi, A., Grazi, G. L., et al. (2000). Evolution of hepatitis B virus polymerase gene mutations in hepatitis B e antigen-negative patients receiving lamivudine therapy. *Hepatology* 32, 1145–1153. doi: 10.1053/jhep.2000.19622

Matlou, M. K., Gaelejwe, L. R., Musyoki, A. M., Rakgole, J. N., Selabe, S. G., and Amponsah-Dacosta, E. (2019). A novel hepatitis B virus recombinant genotype D4/E identified in a South African population. *Heliyon* 5:e01477. doi: 10.1016/j.heliyon.2019.e01477

Mendes-Correa, M. C., Pinho, J. R., Locarnini, S., Yuen, L., Sitnik, R., Santana, R. A., et al. (2010). High frequency of lamivudine resistance mutations in Brazilian patients co-infected with HIV and hepatitis B. *J. Med. Virol.* 82, 1481–1488. doi: 10.1002/jmv.21845

Neff, G. W., Jayaweera, D., and Sherman, K. E. (2006). Drug-induced liver injury in HIV patients. *Gastroenterol. Hepatol. (N. Y.)* 2, 430–437.

Nusrat, S., Khan, M. S., Fazili, J., and Madhoun, M. F. (2014). Cirrhosis and its complications: evidence based treatment. *World J. Gastroenterol.* 20, 5442–5460. doi: 10.3748/wjg.v20.i18.5442

Panduro, A., Maldonado-Gonzalez, M., Fierro, N. A., and Roman, S. (2013). Distribution of HBV genotypes F and H in Mexico and Central America. *Antivir. Ther.* 18, 475–484. doi: 10.3851/IMP2605

Rajbhandari, R., Jun, T., Khalili, H., Chung, R. T., and Ananthakrishnan, A. N. (2016). HBV/HIV co-infection is associated with poorer outcomes in hospitalized patients with HBV or HIV. *J. Viral. Hepat.* 23, 820–829. doi: 10.1111/jvh.12555

Roman, S., Jose-Abrego, A., Fierro, N. A., Escobedo-Melendez, G., Ojeda-Granados, C., Martinez-Lopez, E., et al. (2014). Hepatitis B virus infection in Latin America: a genomic medicine approach. *World J. Gastroenterol.* 20, 7181–7196. doi: 10.3748/wjg.v20.i23.7181

Roman, S., and Panduro, A. (2013). HBV endemicity in Mexico is associated with HBV genotypes H and G. *World J. Gastroenterol.* 19, 5446–5453. doi: 10.3748/wjg.v19.i33.5446

Sakamoto, T., Tanaka, Y., Watanabe, T., Iijima, S., Kani, S., Sugiyama, M., et al. (2013). Mechanism of the dependence of hepatitis B virus genotype G on co-infection with other genotypes for viral replication. *J. Viral. Hepat.* 20, e27–e36. doi: 10.1111/jvh.12022

Sandrin, L., Fourquet, B., Hasquenoph, J. M., Yon, S., Fournier, C., Mal, F., et al. (2003). Transient elastography: a new noninvasive method for assessment of hepatic fibrosis. *Ultrasound Med. Biol.* 29, 1705–1713. doi: 10.1016/j.ultrasmedbio.2003.07.001

Seetharam, A., Perrillo, R., and Gish, R. (2014). Immunosuppression in patients with chronic hepatitis B. *Curr. Hepatol. Rep.* 13, 235–244. doi: 10.1007/s11901-014-0238-2

Sherman, K. E., Peters, M. G., and Thomas, D. (2017). Human immunodeficiency virus and liver disease: a comprehensive update. *Hepatol. Commun.* 1, 987–1001. doi: 10.1002/hep4.1112

Simmonds, P., and Midgley, S. (2005). Recombination in the genesis and evolution of hepatitis B virus genotypes. *J. Virol.* 79, 15467–15476. doi: 10.1128/JVI.79.24.15467-15476.2005

Singh, K. P., Crane, M., Audsley, J., Avihingsanon, A., Sasadeusz, J., and Lewin, S. R. (2017). HIV-hepatitis B virus co-infection: epidemiology, pathogenesis, and treatment. *AIDS* 31, 2035–2052. doi: 10.1097/QAD.0000000000001574

Sterling, R. K., King, W. C., Wahed, A. S., Kleiner, D. E., Khalili, M., Sulkowski, M., et al. (2019). Evaluating noninvasive markers to identify advanced fibrosis by liver biopsy in HBV/HIV co-infected adults. *Hepatology* 71, 411–421. doi: 10.1002/hep.30825

Sterling, R. K., Lissen, E., Clumeck, N., Sola, R., Correa, M. C., Montaner, J., et al. (2006). Development of a simple noninvasive index to predict significant fibrosis in patients with HIV/HCV co-infection. *Hepatology* 43, 1317–1325. doi: 10.1002/hep.21178

Tanaka, T., Sanchez, L. V., Sugiyama, M., Sakamoto, T., Kurbanov, F., Tatematsu, K., et al. (2008). Characteristics of hepatitis B virus genotype G co-infected with genotype H in chimeric mice carrying human hepatocytes. *Virology* 376, 408–415. doi: 10.1016/j.virol.2008.04.001

Tengan, F. M., Abdala, E., Nascimento, M., Bernardo, W. M., and Barone, A. A. (2017). Prevalence of hepatitis B in people living with HIV/AIDS in Latin America and the Caribbean: a systematic review and meta-analysis. *BMC Infect. Dis.* 17:587. doi: 10.1186/s12879-017-2695-z

The Joint United Nations Programme on HIV/AIDS (2019). *AIDS in Mexico*. Available online at: https://www.unaids.org/en (accessed November 11, 2019).

Thio, C. L., Seaberg, E. C., Skolasky, R. Jr., Phair, J., Visscher, B., Munoz, A., et al. (2002). HIV-1, hepatitis B virus, and risk of liver-related mortality in the multicenter cohort study (MACS). *Lancet* 360, 1921–1926. doi: 10.1016/s0140-6736(02)11913-1

Toan, N. L., Song Le, H., Kremsner, P. G., Duy, D. N., Binh, V. Q., Koeberlein, B., et al. (2006). Impact of the hepatitis B virus genotype and genotype mixtures on the course of liver disease in Vietnam. *Hepatology* 43, 1375–1384. doi: 10.1002/hep.21188

Torres-Baranda, R., Bastidas-Ramirez, B. E., Maldonado-Gonzalez, M., Sanchez-Orozco, L. V., Vazquez-Vals, E., Rodriguez-Noriega, E., et al. (2006). Occult hepatitis B in Mexican patients with HIV, an analysis using nested polymerase chain reaction. *Ann. Hepatol.* 5, 34–40.

Trepo, C., Chan, H. L. Y., and Lok, A. (2014). Hepatitis B virus infection. *Lancet* 384, 2053–2063. doi: 10.1016/S0140-6736(14)60220-8

Van Der Kuyl, A. C., Zorgdrager, F., Hogema, B., Bakker, M., Jurriaans, S., Back, N. K., et al. (2013). High prevalence of hepatitis B virus dual infection with genotypes A and G in HIV-1 infected men in Amsterdam, the Netherlands, during 2000-2011. *BMC Infect. Dis.* 13:540. doi: 10.1186/1471-2334-13-540

Wilson, C. M., Wright, P. F., Safrit, J. T., and Rudy, B. (2010). Epidemiology of HIV infection and risk in adolescents and youth. *J. Acquir. Immune Defic. Syndr.* 54(Suppl. 1), S5–S6. doi: 10.1097/QAI.0b013e3181e243a1

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 14 April 2021
doi: 10.3389/fmicb.2021.645020

Analyses of Clinical and Biological Data for French and Belgian Immunocompetent Patients Infected With Hepatitis E Virus Genotypes 4 and 3

*Florence Micas[1], Vanessa Suin[2], Jean-Marie Péron[3], Caroline Scholtes[4,5,6], Edouard Tuaillon[7], Thomas Vanwolleghem[8,9], Laurence Bocket[10], Sébastien Lhomme[1,11], Chloé Dimeglio[1,11], Jacques Izopet[1,11] and Florence Abravanel[1,11]**

[1] Virology Laboratory, National Reference Centre of Hepatitis E Viruses, Federal Institute of Biology, University Hospital Center, Toulouse, France, [2] National Reference Centre of Hepatitis Viruses, Sciensano, Brussels, Belgium, [3] Department of Gastroenterology, Rangueil University Hospital, Toulouse, France, [4] INSERM U1052-Cancer Research Center of Lyon (CRCL), Lyon, France, [5] University of Lyon, University Claude Bernard Lyon 1 (UCBL1), Lyon, France, [6] Department of Virology, Croix Rousse Hospital, Hospices Civils de Lyon, Lyon, France, [7] Pathogenesis and Control of Chronic Infections, INSERM, University of Montpellier, Etablissement Français du Sang, CHU Montpellier, Montpellier, France, [8] Department of Gastroenterology and Hepatology, Antwerp University Hospital, Edegem, Belgium, [9] Laboratory of Experimental Medicine and Pediatrics, University of Antwerp, Antwerp, Belgium, [10] Virology Laboratory EA3610, Faculty of Medicine, University Hospital Center, Lille, France, [11] UMR Inserm U1043, UMR CNRS, U5282, Centre de Physiopathologie de Toulouse Purpan, Toulouse, France

OPEN ACCESS

Edited by:
Lilly Yuen,
Victorian Infectious Diseases Reference Laboratory, Australia

Reviewed by:
Sven Pischke,
University Medical Center Hamburg-Eppendorf, Germany
Andrea Caprioli,
Experimental Zooprophylactic Institute of the Lazio and Tuscany Regions (IZSLT), Italy

***Correspondence:**
Florence Abravanel
abravanel.f@chu-toulouse.fr

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: 22 December 2020
Accepted: 22 March 2021
Published: 14 April 2021

Citation:
Micas F, Suin V, Péron J-M, Scholtes C, Tuaillon E, Vanwolleghem T, Bocket L, Lhomme S, Dimeglio C, Izopet J and Abravanel F (2021) Analyses of Clinical and Biological Data for French and Belgian Immunocompetent Patients Infected With Hepatitis E Virus Genotypes 4 and 3. Front. Microbiol. 12:645020. doi: 10.3389/fmicb.2021.645020

Hepatitis E virus (HEV) genotypes 3 and 4 are the major causes of acute hepatitis in industrialized countries. Genotype 3 is mainly found in Europe and America, while genotype 4 is predominant in Asia. Several Japanese studies have suggested that genotype 4 is more virulent than genotype 3. We investigated this aspect by analyzing the clinical and biological data for 27 French and Belgian immunocompetent patients infected with HEV genotype 4. Their infections were probably acquired locally, since none of these patients reported traveling outside France or Belgium during the 2–8 weeks before symptoms onset. Each patient was matched for age (±5 years) and gender with two patients infected with HEV genotype 3. Bivariate analysis indicated that the HEV genotype 4-infected patients had significantly higher alanine aminotransferase (ALT) (2067 IU/L) and aspartate aminotransferase (AST) (1581 IU/L) activities and total bilirubin concentrations (92.4 μmol/L) than did those infected with HEV genotype 3 (1566 IU/L, $p = 0.016$; 657 IU/L, $p = 0.003$ and 47 μmol/L, $p = 0.046$) at diagnosis. In contrast, more patients infected with HEV genotype 3 reported dark urine (71% vs. 39%, $p = 0.02$) and experienced asthenia (89% vs. 58%, $p < 0.01$) than did those infected with HEV genotype 4. Two HEV genotype 4-infected patients died of multi-organ failure, while none of the genotype 3-infected patients died ($p = 0.035$). Finally, stepwise regression analysis retained only a greater increase in ALT (odds-ratio: 1.0005, 95% confidence interval: 1.00012–1.00084) and less frequent fever (odds-ratio = 0.1244; 95% confidence interval: 0.01887–0.82020) for patients infected with HEV genotype 4.

We conclude that HEV-4 infections are likely to be associated with higher ALT activity than HEV-3 infections. Additional immunological and virological studies are required to confirm these findings and better understand the influence, if any, of genotype on HEV pathophysiology.

Keywords: hepatitis E virus, genotype 4, pathogenicity, immunocompetent, Europe, acute hepatitis

INTRODUCTION

Hepatitis E is the main cause of viral hepatitis worldwide. The 20 million cases that occur each year include more than 3 million symptomatic cases and 60,000 fatalities (Kamar et al., 2017). The virus is responsible for disease in both developing countries with suboptimal sanitary conditions and in industrialized countries where its transmission is mainly zoonotic (Kamar et al., 2017). The eight hepatitis E virus (HEV) genotypes (HEV-1 to HEV-8) that infect mammals include four major ones (HEV-1–4) each with distinct characteristics (Kamar et al., 2017). Smith et al. recently proposed reference sequences for HEV subtypes (HEV-1 to HEV-7), including six subtypes (1a–1f) in HEV-1, two (2a and 2b) in HEV-2, eleven (3a–3j and 3ra) in HEV-3 and nine (4a–4i) in HEV-4 (Smith et al., 2020). HEV-1 and HEV-2 infect only humans, mainly in resource-limited areas. HEV-3 and HEV-4 include zoonotic strains that infect both animals and humans (Doceul et al., 2016; Kamar et al., 2017).

The major transmission routes in industrialized countries are zoonotic. Eating contaminated food like undercooked pork products or game meat is probably the major mode of transmission (Kamar et al., 2017). Veterinarians and farmers working with pigs are more likely to have anti-HEV antibodies due to exposure to infected animals than are blood donors (Meng et al., 2002; Christensen et al., 2008). However, a contaminated environment could be another transmission route as the HEV genome and infectious virus have been found in sewage water and rivers (Clemente-Casares et al., 2003; Rutjes et al., 2009). Lastly, HEV has also been transmitted via blood and organ donations (Kamar et al., 2017).

HEV is a major cause of human disease in European Union countries. For the *European Food Safety Authority,* between 2007 and 2017 more than 21,000 HEV acute clinical cases with 28 fatalities were notified in European Union (Efsa Panel on Biological Hazards (Biohaz) et al., 2017). HEV-3 is the prevalent genotype and subtypes 3f and 3c the main infectious agents in the EU (Efsa Panel on Biological Hazards (Biohaz) et al., 2017; Izopet et al., 2019), while HEV genotype 4 is predominant in Asia, particularly China and Japan (Doceul et al., 2016; Kamar et al., 2017; Primadharsini et al., 2019). Several autochthonous cases of HEV-4 have been reported in Europe, in both humans (Wichmann et al., 2008; Colson et al., 2012a,b; Tesse et al., 2012; Garbuglia et al., 2013; Jeblaoui et al., 2013; Bouamra et al., 2014; Midgley et al., 2014) and pigs (Hakze-van der Honing et al., 2011), over the past 10 years or so. The first HEV-4 reports came from a single autochthonous case in Germany (HEV-4f) and from surveys of swine in Belgium (HEV-4b) (Wichmann et al., 2008; Hakze-van der Honing et al., 2011). In 2011, an outbreak in Italy affecting five people living in the same area with no history of traveling to areas where HEV-4 is endemic was identified as due to HEV-4d, a strain that is similar to Chinese swine isolates (Garbuglia et al., 2013). The authors suggest that autochthonous HEV-4 infections are emerging in Europe and have been transmitted by at least two distinct sources.

Hepatitis E produces a benign infection in most immunocompetent people. Classic forms can have a wide range of symptoms, including fever, nausea, anorexia, asthenia, vomiting, abdominal pain, and icterus (Kamar et al., 2017). Liver function tests are often abnormal, with signs of hepatic cytolysis (increased transaminases), cholestasis, and sometimes even liver failure. However, severe acute or fulminant hepatitis is rare, usually occurring in patients with underlying chronic liver disease and pregnant women with an HEV-1 or HEV-2 infection (Wu et al., 2020). The virus can also produce extra-hepatic symptoms like neurological disorders, renal failure, pancreatitis, and hematological disorders (Kamar et al., 2017). The biological and clinical manifestations are usually less severe in immunocompromised patients, but in few cases HEV infection can become chronic and lead to cirrhosis (Kamar et al., 2017).

A few Japanese studies (Mizuo et al., 2005; Ohnishi et al., 2006; Murata et al., 2019) and a single European one (Jeblaoui et al., 2013) have investigated the differences in the pathogenicities of genotypes 4 and 3. Some studies conducted in Japan demonstrated higher peak aminotransferase activities (Mizuo et al., 2005; Ohnishi et al., 2006; Murata et al., 2019), shorter prothrombin times (Mizuo et al., 2005; Ohnishi et al., 2006; Murata et al., 2019) and higher total bilirubin concentrations (Mizuo et al., 2005; Murata et al., 2019) in HEV-4 infected patients. A study conducted in France reported higher ALT activities and more icterus among HEV-4 infected patients (Jeblaoui et al., 2013). However, all the above-mentioned studies compared only a limited number of HEV-3/HEV-4 infected patients. This study assesses the differences in the severity of disease caused by HEV-3 and HEV-4 infections by analyzing the clinical and biological data for two groups of immunocompetent patients diagnosed with acute hepatitis E in France and Belgium.

MATERIALS AND METHODS

Patients

Plasma samples and data from HEV IgM positive patients from French or Belgian hospitals or laboratories were sent to our reference centeres for HEV RNA detection and genotyping. We included all the immunocompetent

Abbreviations: ALT, alanine aminotransferase; AST, aspartate aminotransferase; HEV, hepatitis E virus; IFN- λ3, interferon lambda 3; RNA, ribonucleic acid.

HEV-4-infected patients reported to the Belgian and French reference centeres between January 2017 and August 2019. We collected the biological, virological and clinical data for these 27 patients, 3 Belgian and 24 French, none of whom had travelled outside France or Belgium during the 2–8 weeks before symptoms onset. Immunocompromised patients were excluded.

Each HEV-4-infected patient was matched for age (±5 years) and gender with 2 HEV-3-infected patients selected randomly from the HEV-3 immunocompetent cases reported to the French national center. We collected the clinical symptoms (asthenia, anorexia, fever—defined as a temperature ≥38°C, abdominal pain, nausea, vomiting, icterus, dark urine, discolored stools, itching) for all the patients by telephone or from their medical reports. We recorded both extrahepatic manifestations (renal failure, pancytopenia, pancreatitis) and neurological manifestations (neuralgic amyotrophy, neuropathic pain, facial paralysis). Hepatic encephalopathy was not classified as a neurological manifestation *per se* as it is linked to hepatic failure. We recorded several biological parameters, including plasma virus load, transaminase activities [alanine aminotransferase (ALT) and aspartate aminotransferase (AST)], prothrombin time and total bilirubin) at the time of diagnosis. We also looked for evidence of existing underlying liver disease (cirrhosis or steatosis). Lastly, we recorded the outcome of the hepatitis E infection: hospitalization or the patient's death.

Virological Methods

The frozen plasma samples collected at the Belgian and French reference centers from suspected HEV infected patients were analyzed in Toulouse or Brussels. HEV RNA was extracted from 500 μl plasma using the MagnaPure 96 instrument (France, elution volume 50 μl), or from 140 μl plasma with the EasyMag instrument (Brussels, elution volume 60 μl). The HEV RNA concentration in 25 μl aliquots of extract was measured using the RealStar HEV RNA 2.0 kit from Altona (Abravanel et al., 2013). The HEV genotype and subtype were determined by population sequencing of open reading frame 2 (ORF2) (Legrand-Abravanel et al., 2009; Suin et al., 2019) and by phylogenetic analysis based on the reference sequences proposed by Smith et al. (2020). A 345 nt fragment of ORF2 (nucleotides 5,994–6,294 of the M73218 genome) was amplified with a hemi-nested PCR protocol (Legrand-Abravanel et al., 2009; Suin et al., 2019). For phylogenetic analysis, nested RT-PCR was performed, targeting the ORF2 region (structural proteins), using primers HEV5944S (5′-GTGGCYGAGGAGGAGGCKAC-3′) and HEV-6484AS (5′-CCCTTRTCCTGCTGNGCATTCTCGACAGA-3′) in the first round and HEV-5930S (5′-CCCTTRTCCTGCTGNGCATTCTCGACAGA-3′) and HEV-6320AS (5′-TGYTGGTTRTCATAATCCTG-3′) in the second round. The PCR products of these amplifications were purified and sequenced using the fluorescent dye terminator method for Big DyeTerminator cycle sequencing (Applied Biosystems, Paris, France) with the same primers as those used for amplification on an Applied Biosystems ABI 3130 XL analyzer. The nucleotide alignments were performed using ClustalX v2.0[1]. The tree was constructed using the MEGA v5.0 software[2] with the neighbor-joining method. Reference sequences of avian HEV were included as an outgroup to root the tree. The final tree was prepared using the bootstrap method (bootstrapped with 1,000 replicates). The sequences have been submitted to Genbank (accession numbers: MW655496 to MW655522).

Statistical Analysis

The anonymized data were analyzed using Stata version 14 (StataCorp LP, College Station, TX, United States). The χ^2-test was used for categorical variables and the Mann-Whitney *U*-test or Wilcoxon rank-sum test were used to compare continuous variables. A *P*-value of < 0.05 was considered to be statistically significant. We used stepwise logistic regression analysis to identify variables independently associated with HEV-4 infections. The model takes into account all the χ^2 or Mann-Whitney test parameters with *p*-values of < 0.20.

RESULTS

All 27 immunocompetent HEV-4-infected patients were infected with subtype 4b (**Figure 1**). They were matched with 42 patients infected with HEV-3f, 9 with HEV-3c, and 3 with HEV-3l.

The median age of the 27 HEV-4-infected patients and 54 HEV-3-infected patients was 60 years (range 31–84 years). Their median ALT and AST activities were 1,652 (range 37–8,900 IU/L) and 1022 IU/L (range 22–11,401 IU/L); their total median bilirubin, prothrombin, and HEV RNA plasma concentrations were 68.9 μmol/L (range 3.8–446 μmol/L), 91% (range 19–100%) and 636 000 UI/mL (range 1,110–146.000.000 UI/mL), respectively.

More HEV-3-infected subjects appeared to suffer from asthenia than did those infected with genotype 4 ($p = 0.002$) and more of them also reported dark urine than did the HEV-4 patients ($p = 0.017$). There were no other significant differences in the clinical manifestations of the two groups; particularly no difference in their neurological manifestations or other extra-hepatic disorders. Two HEV-4 patients with underlying cirrhosis died of multi-organ failure, while none of the genotype 3 patients died ($p = 0.035$).

Bivariate analysis found that the HEV-4 patients had significantly higher ALT (2,067 IU/L) and AST (1,581 IU/L) activities and total bilirubin (92.4 μmol/L) concentrations than did the HEV-3 patients (ALT: 1,566 IU/L; $p = 0.016$; AST: 657 IU/L; $p = 0.003$; bilirubin: 47 μmol/L; $p = 0.046$) at diagnosis (**Table 1**). We found no genotype-dependent differences in existing chronic disease, prothrombin time, or virus load.

Stepwise logistic regression indicated that genotype 4 infections were associated with higher ALT concentrations (OR = 1.0005; $p = 0.01$) and fewer fevers (OR = 0.1244; $p = 0.03$) than were HEV-3 infections (**Table 1**).

[1]http://www.clustal.org/download/current/

[2]http://www.megasoftware.net/

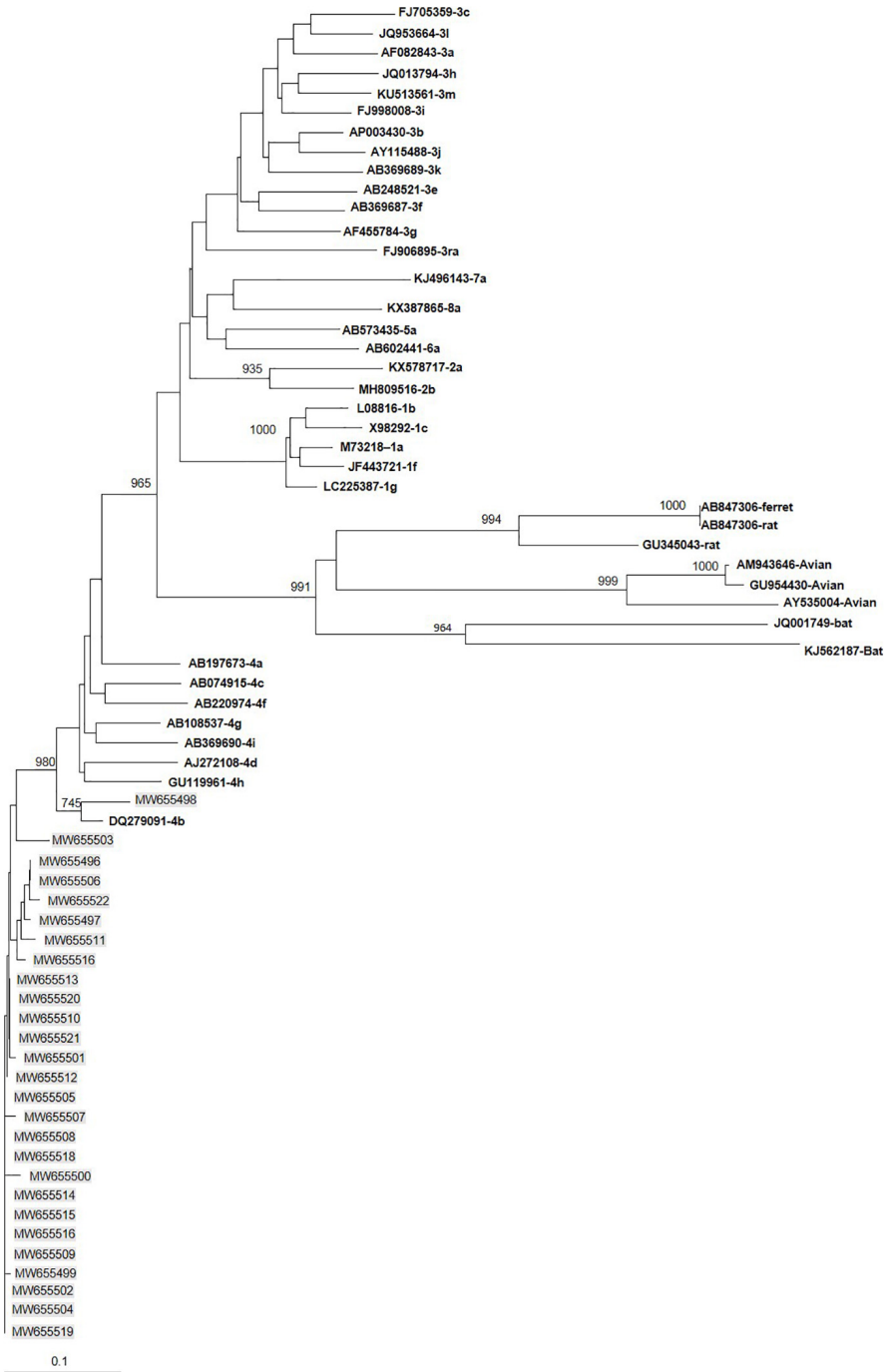

FIGURE 1 | Phylogenetic tree constructed using 345-nt-long partial sequences within ORF2. Genetic distances were calculated using the Kimura two-parameter method, phylogenetic trees were plotted by the neighbor-joining method. Bootstrap values acquired after 1,000 replications are shown (branch lengths measured in the number of substitutions per site). Patients' sequences, with the accession number highlighted in gray, were compared to the reference sequences of subtypes 3 proposed by Smith et al. (2020). Accession numbers, genotype and subtypes of the reference sequences are in bold.

DISCUSSION

The evaluation of clinical and biological data of 27 cases of HEV-4 infections matched for age and gender with 54 cases of HEV-3 infections detected in Belgium and France demonstrated, based on the logistic regression analysis, that the HEV-4 infected patients had significant higher ALT activities.

HEV-4 is predominant in Asia, but this genotype has also been found in human populations in several European countries

TABLE 1 | Patient characteristics according to HEV genotype.

	HEV-4 infected patients *N* = 27	HEV-3 infected patients *N* = 54	*P* value	Odds-ratio	Standard error	*P* > *z*	95% Confidence interval
Gender							
men	18	36	1.0				
women	9	18					
Median age in year (standard deviation)	60 (12.5)	60 (13.1)	0.9				
Existing chronic liver disease	5 (18.5%)	6 (11.1%)	0.36				
Hospitalization	22 (81.5%)	38 (71.7%)	0.34				
Asthenia	14 (58.3%)	47 (88.7%)	0.002				
Anorexia	9 (37.5%)	30 (55.6%)	0.14				
Fever> 38°C	5 (20.8%)	21 (38.9%)	0.12	0.1244	0.12	0.03	0.018-0.82
Nausea	6 (25%)	21 (38.9%)	0.23				
Vomiting	4 (16.7%)	21 (38.9%)	0.05				
Abdominal pain	13 (54.2%)	17 (31.5%)	0.06				
Mucosal icterus	16 (66.7%	27 (50%)	0.17				
Colored urine	7 (38.9%)	34 (70.8%)	0.02				
Discolored stool	6 (37.5%)	22 (45.8%)	0.561				
Pruritus	4 (16.7%	17 (31.5%)	0.17				
Extrahepatic manifestations (except neurological)	4 (15.4%)	5 (12.5%)	0.74				
Neurological manifestations (except hepatic encephalopathy)	2 (8.3%)	11 (20.4%)	0.19				
Median Virus load (UI/mL)	996 000	453 000	0.63				
Median ALT on entry (UI/L)	2067	1566	0.016	1.0005	0.0002	0.01	1.00012–1.00084
Median AST on entry (UI/L)	1581	657	0.003				
Median total bilirubin (μ mol/L) on entry	92.4	47	0.04				

(Wichmann et al., 2008; Colson et al., 2012a,b; Tesse et al., 2012; Garbuglia et al., 2013; Jeblaoui et al., 2013; Bouamra et al., 2014; Midgley et al., 2014). Japanese studies suggest that HEV-4 infections are associated with more severe hepatitis, higher peak aminotransferase activities (Mizuo et al., 2005; Ohnishi et al., 2006; Murata et al., 2019), shorter prothrombin times (Mizuo et al., 2005; Ohnishi et al., 2006; Murata et al., 2019) and higher total bilirubin concentrations (Mizuo et al., 2005; Murata et al., 2019). A previous European study also found higher ALT activities and more icterus among HEV-4 infected patients (Jeblaoui et al., 2013). However, these studies of the two zoonotic genotypes often involved limited, unbalanced patient groups. The Japanese studies included only a few ($\leq$ 11) HEV-3-infected patients (Mizuo et al., 2005; Ohnishi et al., 2006; Murata et al., 2019), while the French study compared 9 cases of HEV-4 with 208 cases of HEV-3 (Jeblaoui et al., 2013). The results of our bivariate and multivariate analyses agree with those of previous studies (Mizuo et al., 2005; Ohnishi et al., 2006; Jeblaoui et al., 2013; Murata et al., 2019): HEV-4 infections were associated with higher ALT activities. These patients also suffered fewer episodes of fever than did patients with HEV-3 infections; this clinical manifestation was not investigated in previous Japanese and French studies. Bivariate analysis also indicated that HEV-4 infected-patients had higher AST activities and greater total bilirubin concentrations. Surprisingly, more of the HEV-3-infected patients reported producing dark urine and suffered from asthenia than did the HEV-4-infected patients. The numbers of severe hepatitis cases (defined as a prothrombin level $<$ 50%) in the two groups were similar. However, two deaths were recorded among HEV-4 infected patients while none among HEV-3 infected patients. This observation, although statistically significant, was not confirmed by the multivariate analysis, probably because of the small number of deaths recorded. It is always possible that the deaths of our genotype-4-infected patients were linked to other comorbidities that were not recorded by our questionnaire. In this regard, no difference in the mortality rate was reported in a study performing a systematic review and pooled analysis of acute liver failure caused by HEV-3 and HEV-4 (Haffar et al., 2018).

Further studies are needed to determine whether genotype can influence the mortality rate in HEV infected patients in industrialized countries.

We previously reported that different HEV-3 subtypes could influence the clinical and biological features of an HEV infection (Subissi et al., 2019; Abravanel et al., 2020). In this study, immunocompetent patients infected with HEV subtype 3f suffered more from fever, had higher virus loads and required hospitalization than did those with subtype 3c infections (Abravanel et al., 2020). All the genotype 4-infected patients in the present study were infected with subtype 4b. Therefore, the influence of the HEV-4 subtype was not investigated.

Interestingly, we found that HEV-3-infected patients were more likely to have a fever than HEV-4 infected patients. This observation, although possibly influenced by many other factors, might be due to genotype-linked differences in the innate and inflammatory responses. Indeed, several studies reported HEV genotype-dependent stimulation of immune mediators (Syedbasha and Egli, 2017; Gouilly et al., 2018; Murata et al., 2019), possibly leading to differences in pathogenicity.

The main limitations of our study concern data that were not collected, such as the influence of patient comorbidities. The severity of an infection could be altered by conditions like obesity or diabetes and we were unable to assess the histological severity of the liver disease, as liver biopsies were not taken. Another limitation is a lack of information on the possible sources of infection.

CONCLUSION

We conclude that HEV-4 infections are likely to be associated with higher ALT activity than HEV-3 infections. Additional immunological and virological studies are now required to fully describe the pathophysiology of the main HEV genotypes.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

Ethical approval was not provided for this study on human participants because This was a non-interventional study. Biological material and clinical data were obtained only via standard viral diagnostics following a physician's order (no supplemental or modified sampling). Data were analyzed anonymously. The French Public Health law (CSP Art L 1121-1.1) does not require written informed consent from the patients for such a protocol. Written informed consent was not provided because this was a non-interventional study. Biological material and clinical data were obtained only via standard viral diagnostics following a physician's order (no supplemental or modified sampling). Data were analyzed anonymously. The French Public Health law (CSP Art L 1121-1.1) does not require written informed consent from the patients for such a protocol.

AUTHOR CONTRIBUTIONS

JI, FM, and FA: drafting and refining the manuscript. VS, SL, and TV: critical reading of the manuscript. CD statistical analyses. FM, VS, J-MP, CS, ET, TV, LB, and SL: collected the data. All authors have read and approved the manuscript, contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

We thank Owen Parkes for editing the English text.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.645020/full#supplementary-material

REFERENCES

Abravanel, F., Chapuy-Regaud, S., Lhomme, S., Dubois, M., Peron, J. M., Alric, L., et al. (2013). Performance of two commercial assays for detecting hepatitis E virus RNA in acute or chronic infections. *J. Clin. Microbiol.* 51, 1913–1916. doi: 10.1128/jcm.00661-13

Abravanel, F., Dimeglio, C., Castanier, M., Peron, J. M., Kamar, N., Lhomme, S., et al. (2020). Does HEV-3 subtype play a role in the severity of acute hepatitis E? *Liver Int*. 40, 333–337. doi: 10.1111/liv.14329

Bouamra, Y., Gerolami, R., Arzouni, J. P., Grimaud, J. C., Lafforgue, P., Nelli, M., et al. (2014). Emergence of autochthonous infections with hepatitis E virus of genotype 4 in Europe. *Intervirology* 57, 43–48. doi: 10.1159/000354801

Christensen, P. B., Engle, R. E., Hjort, C., Homburg, K. M., Vach, W., Georgsen, J., et al. (2008). Time trend of the prevalence of hepatitis E antibodies among farmers and blood donors: a potential zoonosis in Denmark. *Clin. Infect. Dis.* 47, 1026–1031. doi: 10.1086/591970

Clemente-Casares, P., Pina, S., Buti, M., Jardi, R., MartIn, M., Bofill-Mas, S., et al. (2003). Hepatitis E virus epidemiology in industrialized countries. *Emerg. Infect. Dis.* 9, 448–454. doi: 10.3201/eid0904.020351

Colson, P., Romanet, P., Moal, V., Borentain, P., Purgus, R., Benezech, A., et al. (2012a). Autochthonous infections with hepatitis E virus genotype 4. France. *Emerg. Infect. Dis.* 18, 1361–1364. doi: 10.3201/eid1808.111827

Colson, P., Swiader, L., Motte, A., Ferretti, A., Borentain, P., and Gerolami, R. (2012b). Circulation of almost genetically identical hepatitis E virus of genotype 4 in France. *J. Clin. Virol.* 55, 181–183. doi: 10.1016/j.jcv.2012.06.014

Doceul, V., Bagdassarian, E., Demange, A., and Pavio, N. (2016). Zoonotic hepatitis E virus: classi fi cation, animal reservoirs and transmission routes. *Viruses* 8:270. doi: 10.3390/v8100270

Efsa Panel on Biological Hazards (Biohaz), Ricci, A., Allende, A., Bolton, D., Chemaly, M., Davies, R., et al. (2017). Public health risks associated with hepatitis E virus (HEV) as a food-borne pathogen. *EFSA J.* 15:e04886.

Garbuglia, A. R., Scognamiglio, P., Petrosillo, N., Mastroianni, C. M., Sordillo, P., Gentile, D., et al. (2013). Hepatitis E virus genotype 4 outbreak, Italy, 2011. *Emerg. Infect. Dis*. 19, 110–114. doi: 10.3201/eid1901.120983

Gouilly, J., Chen, Q., Siewiera, J., Cartron, G., Levy, C., Dubois, M., et al. (2018). Genotype specific pathogenicity of hepatitis E virus at the human maternal-fetal interface. *Nat. Commun*. 9:4748.

Haffar, S., Shalimar, Kaur, R. J., Wang, Z., Prokop, L. J., Murad, M. H., et al. (2018). Acute liver failure caused by hepatitis E virus genotype 3 and 4: a systematic review and pooled analysis. *Liver Int*. 38, 1965–1973. doi: 10.1111/liv.13861

Hakze-van der Honing, R. W., van Coillie, E., Antonis, A. F., and van der Poel, W. H. (2011). First isolation of hepatitis E virus genotype 4 in Europe through swine surveillance in the Netherlands and Belgium. *PLoS One* 6:e22673. doi: 10.1371/journal.pone.0022673

Izopet, J., Tremeaux, P., Marion, O., Migueres, M., Capelli, N., Chapuy-Regaud, S., et al. (2019). Hepatitis E virus infections in Europe. *J. Clin. Virol*. 120, 20–26.

Jeblaoui, A., Haim-Boukobza, S., Marchadier, E., Mokhtari, C., and Roque-Afonso, A. M. (2013). Genotype 4 hepatitis e virus in france: an autochthonous infection with a more severe presentation. *Clin. Infect. Dis*. 57, e122–e126.

Kamar, N., Izopet, J., Pavio, N., Aggarwal, R., Labrique, A., Wedemeyer, H., et al. (2017). Hepatitis E virus infection. *Nat. Rev. Dis. Primers* 3:17086.

Legrand-Abravanel, F., Mansuy, J. M., Dubois, M., Kamar, N., Peron, J. M., Rostaing, L., et al. (2009). Hepatitis E virus genotype 3 diversity, France. *Emerg. Infect. Dis*. 15, 110–114. doi: 10.3201/eid1501.080296

Meng, X. J., Wiseman, B., Elvinger, F., Guenette, D. K., Toth, T. E., Engle, R. E., et al. (2002). Prevalence of antibodies to hepatitis E virus in veterinarians working with swine and in normal blood donors in the United States and other countries. *J. Clin. Microbiol*. 40, 117–122. doi: 10.1128/jcm.40.1.117-122.2002

Midgley, S., Vestergaard, H. T., Dalgaard, C., Enggaard, L., and Fischer, T. K. (2014). Hepatitis E virus genotype 4, Denmark, 2012. *Emerg. Infect. Dis*. 20, 156–157. doi: 10.3201/eid2001.130600

Mizuo, H., Yazaki, Y., Sugawara, K., Tsuda, F., Takahashi, M., Nishizawa, T., et al. (2005). Possible risk factors for the transmission of hepatitis E virus and for the severe form of hepatitis E acquired locally in Hokkaido, Japan. *J. Med. Virol*. 76, 341–349. doi: 10.1002/jmv.20364

Murata, K., Kang, J. H., Nagashima, S., Matsui, T., Karino, Y., Yamamoto, Y., et al. (2019). IFN-lambda3 as a host immune response in acute hepatitis E virus infection. *Cytokine* 125:154816. doi: 10.1016/j.cyto.2019.154816

Ohnishi, S., Kang, J. H., Maekubo, H., Arakawa, T., Karino, Y., Toyota, J., et al. (2006). Comparison of clinical features of acute hepatitis caused by hepatitis E virus (HEV) genotypes 3 and 4 in Sapporo, Japan. *Hepatol. Res*. 36, 301–307. doi: 10.1016/j.hepres.2006.08.002

Primadharsini, P. P., Nagashima, S., and Okamoto, H. (2019). Genetic variability and evolution of hepatitis E Virus. *Viruses* 11, 456. doi: 10.3390/v11050456

Rutjes, S. A., Lodder, W. J., Lodder-Verschoor, F., van den Berg, H. H., Vennema, H., Duizer, E., et al. (2009). Sources of hepatitis E virus genotype 3 in the Netherlands. *Emerg. Infect. Dis*. 15, 381–387.

Smith, D. B., Izopet, J., Nicot, F., Simmonds, P., Jameel, S., Meng, X. J., et al. (2020). Update: proposed reference sequences for subtypes of hepatitis E virus (species *Orthohepevirus A*). *J. Gen. Virol*. 101, 692–698. doi: 10.1099/jgv.0.001435

Subissi, L., Peeters, M., Lamoral, S., Klamer, S., Suin, V., and Van Gucht, S. (2019). Subtype-specific differences in the risk of hospitalisation among patients infected with hepatitis E virus genotype 3 in Belgium, 2010-2018. *Epidemiol. Infect*. 147:e224.

Suin, V., Klamer, S. E., Hutse, V., Wautier, M., Jacques, M., Abady, M., et al. (2019). Epidemiology and genotype 3 subtype dynamics of hepatitis E virus in Belgium, 2010 to 2017. *Euro Surveill*. 24:1800141.

Syedbasha, M., and Egli, A. (2017). Interferon lambda: modulating immunity in infectious diseases. *Front. Immunol*. 8:119. doi: 10.3389/fimmu.2017.00119

Tesse, S., Lioure, B., Fornecker, L., Wendling, M. J., Stoll-Keller, F., Bigaillon, C., et al. (2012). Circulation of genotype 4 hepatitis E virus in Europe: first autochthonous hepatitis E infection in France. *J. Clin. Virol*. 54, 197–200. doi: 10.1016/j.jcv.2012.02.007

Wichmann, O., Schimanski, S., Koch, J., Kohler, M., Rothe, C., Plentz, A., et al. (2008). Phylogenetic and case-control study on hepatitis E virus infection in Germany. *J. Infect. Dis*. 198, 1732–1741. doi: 10.1086/593211

Wu, C., Wu, X., and Xia, J. (2020). Hepatitis E virus infection during pregnancy. *Virol. J*. 17:73.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

REVIEW
published: 24 May 2021
doi: 10.3389/fmicb.2021.653684

Origins and Evolution of the Primate Hepatitis B Virus

*Stephen A. Locarnini, Margaret Littlejohn and Lilly K. W. Yuen**

Victorian Infectious Diseases Reference Laboratory, Royal Melbourne Hospital, The Peter Doherty Institute for Infection and Immunity, Melbourne, VIC, Australia

Recent interest in the origins and subsequent evolution of the hepatitis B virus (HBV) has strengthened with the discovery of ancient HBV sequences in fossilized remains of humans dating back to the Neolithic period around 7,000 years ago. Metagenomic analysis identified a number of African non-human primate HBV sequences in the oldest samples collected, indicating that human HBV may have at some stage, evolved in Africa following zoonotic transmissions from higher primates. Ancestral genotype A and D isolates were also discovered from the Bronze Age, not in Africa but rather Eurasia, implying a more complex evolutionary and migratory history for HBV than previously recognized. Most full-length ancient HBV sequences exhibited features of inter genotypic recombination, confirming the importance of recombination and the mutation rate of the error-prone viral replicase as drivers for successful HBV evolution. A model for the origin and evolution of HBV is proposed, which includes multiple cross-species transmissions and favors subsequent recombination events that result in a pathogen and can successfully transmit and cause persistent infection in the primate host.

Keywords: hepatitis B virus, genotype, evolution, human migration, ancient DNA

OPEN ACCESS

Edited by:
Anna Kramvis,
University of the Witwatersrand, South Africa

Reviewed by:
Dieter Glebe,
University of Giessen, Germany
Pakorn Aiewsakun,
Mahidol University, Thailand
C.-Thomas Bock,
Robert Koch Institute (RKI), Germany

****Correspondence:***
Lilly K. W. Yuen
lilly.yuen@vidrl.org.au

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *15 January 2021*
Accepted: *21 April 2021*
Published: *24 May 2021*

Citation:
Locarnini SA, Littlejohn M and Yuen LKW (2021) Origins and Evolution of the Primate Hepatitis B Virus. Front. Microbiol. 12:653684. doi: 10.3389/fmicb.2021.653684

INTRODUCTION

Infection of the human host with the hepatitis B virus (HBV) can result in a diverse spectrum of clinical outcomes ranging from asymptomatic hepatitis through to cirrhotic liver disease and hepatocellular carcinoma (HCC). Hepatitis B remains a major public health challenge with over 257 million people worldwide presently chronically infected, of whom more than 880,000 persons will die directly each year (World Health Organisation, 2017). Chronic hepatitis B causes almost 40% of cases of HCC and is the second leading cause of cancer-related mortality globally (Stanaway et al., 2016). Long-term outcomes of chronic hepatitis B can vary widely but viral biomarkers, such as HBV genotype and signature mutation profiles in the HBV genome, and serological biomarkers, such as viral load and quantitative levels of Hepatitis B surface antigen (HBsAg) and Hepatitis B e antigen (HBeAg), can predict eventual clinical outcomes (Chen et al., 2007; Yuen et al., 2008; Kramvis, 2014). In terms of host factors, progression to chronic infection is inversely related to age at the time of infection (McMahon et al., 1985; Hyams, 1995), but the final outcome for the exposed individual is dependent on the interaction between host and virus. The more important viral factors include viral genetic variation, viral genotype, and HBeAg status.

In this article, the authors review current theories for the origins of the primate hepadnaviruses putting them in the context of several recent discoveries including advances in recovering

HBV DNA sequences from ancient human skeletal and mummified remains (Kahila Bar-Gal et al., 2012; Krause-Kyora et al., 2018; Muhlemann et al., 2018; Patterson Ross et al., 2018) as well as evolutionary processes considered to be involved in the emergence of the major modern HBV genotypes. Finally, we propose a model of HBV origins and diversity drawing on the importance of cross-species transmission and subsequent recombination events not only for the recently discovered ancient HBV (aHBV) strains but also the more contemporary isolates of the virus.

THE FAMILY HEPADNAVIRUSES

General Considerations

Human HBV is the prototype member of the family *Hepadnaviridae*, which presently includes five genera: the *Orthohepadnavirus* genus that infects mammals, the *Avihepadnavirus* genus that infects birds, the *Parahepadnavirus* and *Metahepadnavirus* genera, which infect teleost fish, and the *Herpetohepadnavirus* genus whose members infect reptiles and frogs (Magnius et al., 2020). All members have similar ultra-structural and molecular genomic features including virion size and morphology, and an enveloped nucleocapsid, which contains a relaxed circular double-stranded DNA genome of 3.0–3.4 kb. This DNA genome is replicated *via* a process of reverse transcription of the key intermediate pregenomic RNA in hepatocytes (Summers and Mason, 1982; Tiollais et al., 1985). Another important feature of hepadnaviral replication is the organization of the DNA into a minichromosome in the nucleus of infected hepatocytes (Bock et al., 1994; Newbold et al., 1995). The HBV genome is organized into four overlapping, but frame-shifted open reading frames (**Figure 1**). Hepadnaviruses infecting other hosts have recently been identified including bats frogs, lizards, fish, and the capuchin monkey (MacDonald et al., 2000; Drexler et al., 2013; Lauber et al., 2017; de Carvalho Dominguez Souza et al., 2018). The phylogenetic relatedness and relationships within the Hepadnaviridae are shown in **Figure 2**, with the emphasis of this review on the primate hepadnaviruses, human and non-human.

Mutation Rate of the Hepadnavirus Genome

Varying interpretations of phylogenetic data have led to uncertainty about the evolutionary history of HBV in humans and non-human primates (NHP; Fares and Holmes, 2002; Starkman et al., 2003) including variable estimates of mutational rates. To some extent, this has been attributed to the anatomy of the HBV genome with overlapping reading frames and the underlying RNA structures embedded within the genome (e.g., epsilon), influencing the evolution of particular regions (Mizokami et al., 1997; Simmonds and Smith, 1999). Estimates obtained for the substitution rate inferred by comparing mother-to-baby transmission infections indicated lower substitution rates in the HBeAg-positive phase (anti-HBe-negative) than in the HBeAg-negative phase (Hannoun et al., 2000; Harrison et al., 2011) by around one order of magnitude. This variance in substitution rates between HBeAg-positive and HBeAg-negative chronic hepatitis B was confirmed by Tedder et al. (2013). The "red queen" hypothesis proposed by these investigators shows that many of the mutations observed in HBV genomes do not generate variability but are reversions back to the genotype consensus (Tedder et al., 2013). The authors concluded that HBV probably behaves as a self-normalizing meme *in vivo* and most of the mutations do not lead to significant changes over time in persistent infection.

Complementing the "red queen hypothesis", there is increasing evidence that short-time scale studies can artificially inflate evolutionary rates (Li et al., 2017) and the viral genetic diversity rate may decrease over longer time scales; this is referred to as the time-dependent rate phenomenon (Ho et al., 2011; Aiewsakun and Katzourakis, 2016). This reduction rate was found to be consistent with a power-law relationship between substitution rate and observational period (Aiewsakun and Katzourakis, 2017), and all recently identified aHBV fitted remarkably well within this relationship (Simmonds et al., 2019).

Primate Hepadnaviruses: Modern Genotypes and Subgenotypes

The primate hepadnaviruses are indigenous to their hosts (Norder et al., 1996; Lanford et al., 1998; Warren et al., 1999; Grethe et al., 2000; Robertson and Margolis, 2002; Sall et al., 2005), and the phylogenetic analysis of full-length genomes from the different primate HBVs, both human and non-human, reveals

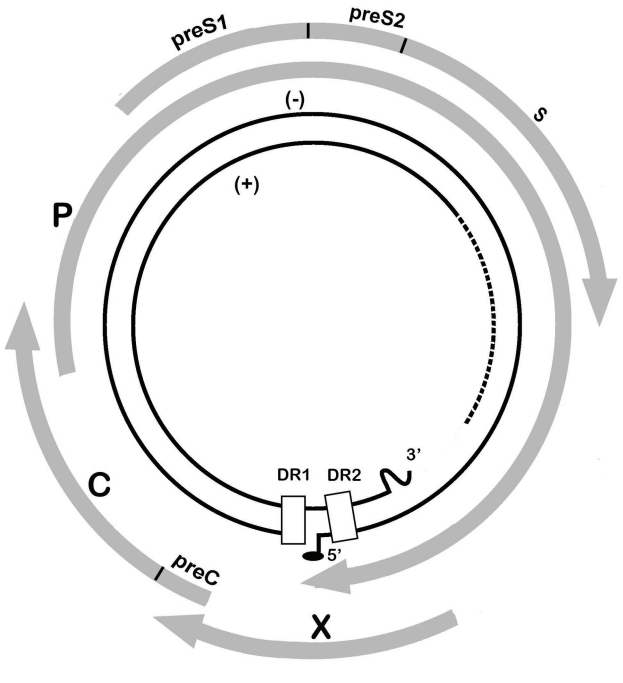

FIGURE 1 | Schematic representation of the HBV genome. The inner circles represent the partial double-stranded DNA molecule and location of the direct repeat (DR) sequences, while the solid gray arrows illustrate the four open reading frames in the HBV genome: polymerase (P), precore (preC), core (C), hepatitis b X protein (X), and envelope (preS1, preS2, and S).

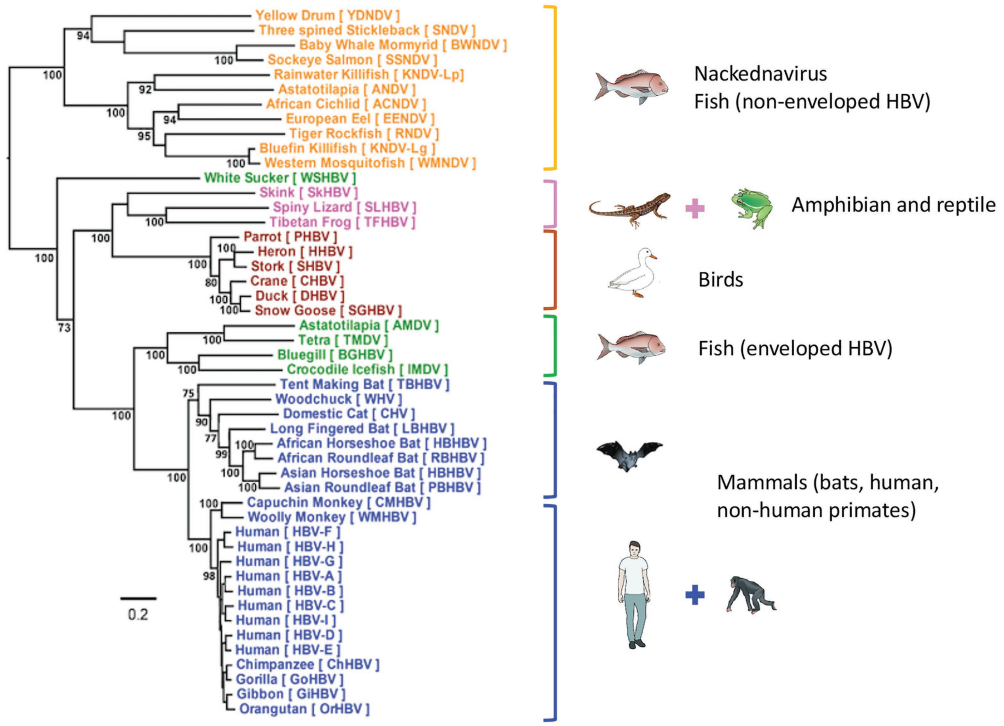

FIGURE 2 | Phylogenetic tree of *Hepadnaviridae*. Phylogenetic analysis of representative hepadnavirus full-length genome sequences from each of the host species known to harbor hepatitis viruses including the HBV genotypes from humans. In general, all HBVs from the same host species cluster together with strong bootstrap support (>80%; with the exception of WSHBV, an enveloped HBV from fish). Based on the available HBV genome sequences, the enveloped and non-enveloped HBVs had distinct evolutionary histories following divergence from their most recent common ancestor (MRCA). The enveloped HBVs are further divided into two major clusters, mammals and fish and amphibians, reptiles, and birds sharing another MRCA. Of these genome sequences, 34 were sourced from GenBank and 14 were annotated sequences from the supplementary data files of Lauber et al. (2017), representing new isolates. The maximum-likelihood (ML) phylogenetic tree was estimated using IQ-TREE v1.6.5 (Nguyen et al., 2015), which is packaged with ModelFinder (Kalyaanamoorthy et al., 2017) and UFBoot (Hoang et al., 2018). Clade support was assessed using 1,000 pseudo replicates generated with the UFBoot non-parametric bootstrap procedure. Branch lengths are scaled to the number of nucleotide substitutions per sequence site. (From (Revill et al., 2020); modified, with permission from the author).

that they essentially cluster according to the geographical locations of host habitats, particularly for NHP HBV (Starkman et al., 2003; Kramvis, 2014). For example, HBV from central chimpanzees (*Pan troglodytes*) are genetically more related to HBV collected from gorillas, with whom they share an overlapping geographical range (the region south of the Sanga River in Cameroon and west of the Oubangui River in Congo, Zaire; **Figure 3A**; Hu et al., 2001), than to HBV from other common chimpanzee subspecies in Africa (**Figure 3A**). Likewise, HBV isolated from central, eastern (*Pan troglodytes schweinfurthii*), and Nigerian-Cameroon (*Pan troglodytes ellioti* and *vellerosus*) chimpanzees, with adjacent habitats, formed sister clades. While HBV sequences from western chimpanzee (*Pan troglodytes verus*, west of the Niger River) formed the most distal clade of all African NHP HBV, and their habitat is the most distant from the other common chimpanzees (**Figure 3A**). Surprisingly, no HBV isolates have been identified from the bonobo (*Pan paniscus*) chimpanzee. The same observations are seen with the Southeast Asian NHP, where the orangutan HBV sequences cluster more closely with those from Bornean gibbons (*Hylobatidae muelleri* and *Hylobatidae albibarbis*) and Island Southeast Asian gibbon (*Hylobatidae agilis*; **Figure 3B**). The other gibbon species from mainland areas of Southeast Asia form a separate group.

The phylogenetic relationships of the NHP HBV are reflected in the nucleotide divergence data. The highest divergence is found between the African and Asian NHP HBVs, which diverge by 10–11% at the nucleotide level. Within the African groups, the divergence between the various chimpanzee and gorilla HBV species is 5–7%, while within the Asian species, orangutan, and gibbon HBV species, the divergence is slightly higher at 7–9%. The gibbon HBV sequences can be further split into mainland species, where the divergence is 7–8%, compared to the island species, where the divergence is slightly higher at 8–9%. Using the accepted definition for human HBV genotypes of >7.5% divergence, each of these NHP HBV species is regarded as individual genotypes.

Phylogenetic analysis, including both human and NHP HBV, shows both groups form unique clades, separate but interspersed (**Figure 4**; Locarnini et al., 2013). The nucleotide divergence between human and Old World NHP HBVs range 10–15%, but is substantially higher when compared to New World NHP HBVs. The level of nucleotide divergence between HBV from

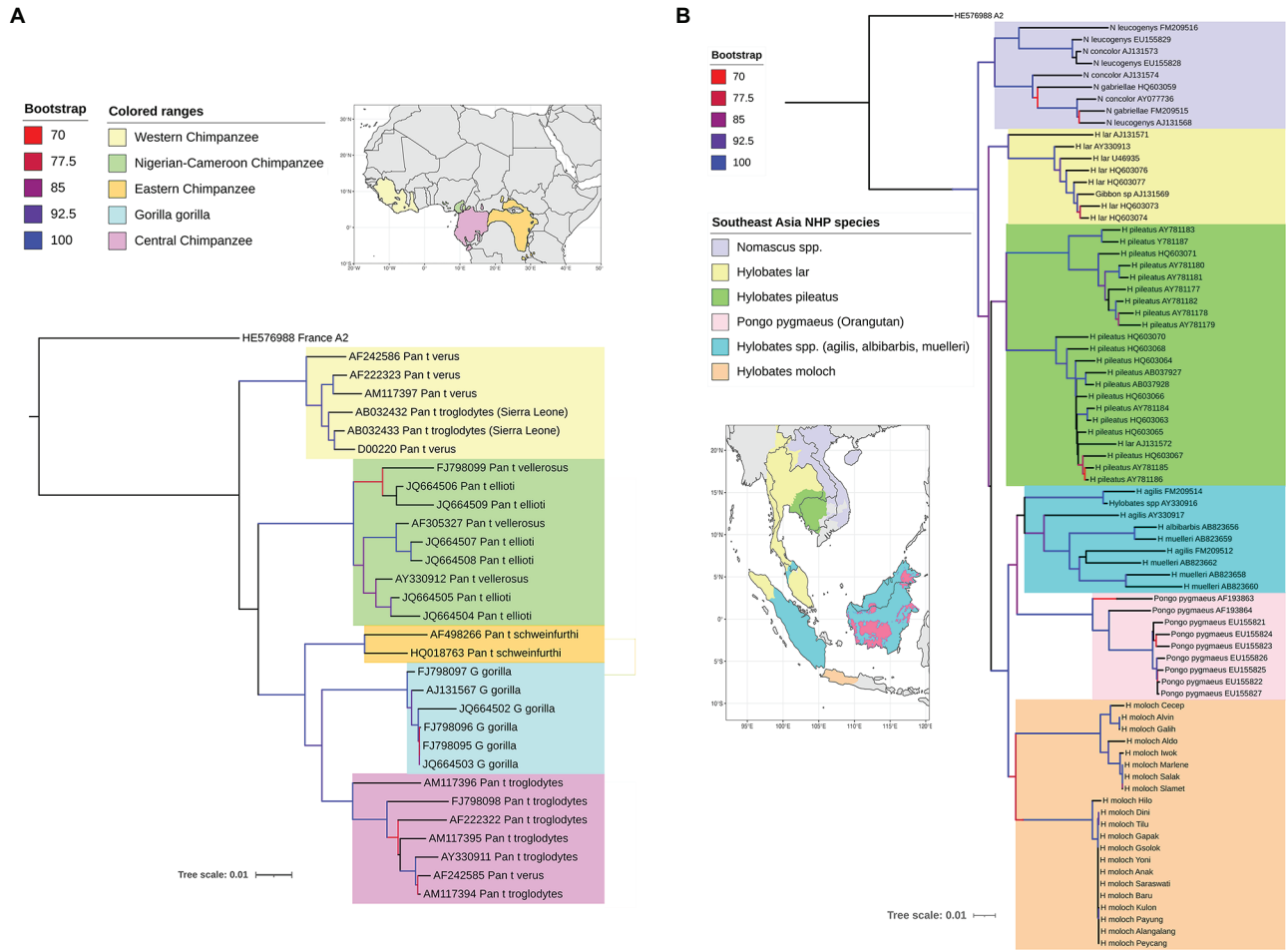

FIGURE 3 | Maximum likelihood (ML) phylogenetic tree of full-length HBV genome sequences extracted from non-human primates **(A)** in Africa and **(B)** Southeast Asia, and reference HBV subgenotype sequences obtained from GenBank. The natural habitat locations of the non-human primates are shown in the corresponding maps, with same the color code used on the ML trees for the different host species. The trees were inferred using IQ-TREE v1.6.5 (Nguyen et al., 2015). The best-fit models determined by ModelFinder (Kalyaanamoorthy et al., 2017) were GTR+F+R2 and GTR+F+R4, respectively. Clade support was assessed using 1,000 pseudo replicates generated with the UFBoot non-parametric bootstrap procedure (Hoang et al., 2018), and bootstrap values >70% are shown by a color scale on the tree branches. The ML tree was annotated using iTOL v5.7 (Letunic and Bork, 2019). Branch lengths are scaled to the number of nucleotide substitutions per sequence site.

Woolly Monkey and Capuchin Monkey, when compared to all other primate HBV sequences are 28% (Lanford et al., 1998) and 20% (de Carvalho Dominguez Souza et al., 2018), respectively.

It is worth noting that NHP HBV has several distinctive amino acid replacement features compared to a consensus sequence derived from each of the human genotypes (Robertson and Margolis, 2002), and some of these features were also detected in the NHP-like aHBV (**Table 1**). The NHP HBV distinctive amino acid replacement features include:

1. Pre-S region: a glutamic acid (E) at position 16 in chimpanzee, gorilla, and gibbon isolates.
2. S-region: three amino acid changes (L133, I/L/Y134, and A177) in the chimpanzee, gorilla, and gibbon HBV isolates.
3. Core antigen: the typical precore stop codon at position 28 (associated with HBeAg negativity) has a leucine (L) at this position in contrast to tryptophan (W) in all the human isolates. Likewise, in the core region itself, glutamine (Q) at position 113 of the chimpanzee and gorilla HBVs replaces a leucine (L) in human HBV. In addition, Takahashi et al. (2000) highlighted a proline (P) alanine (A) motif within five amino acids from the termination of the core protein (C-terminal) found in all NHP HBV sequences as well as human genotypes E and F/H and genotype G (Stuyver et al., 2000).
4. Pol region: an 11 amino acid deletion in the pol gene and the N-terminal portion of the Pre-S1, which is shared with the human genotype D strains.
5. X-gene: three amino acid changes (T11, K107, and T110) in chimpanzees and two (K107 and T110) found in gorilla sequences.

Interestingly, the Pre-S1 deletion, the precore amino acid at position 28, the core amino acid position at L113Q, and

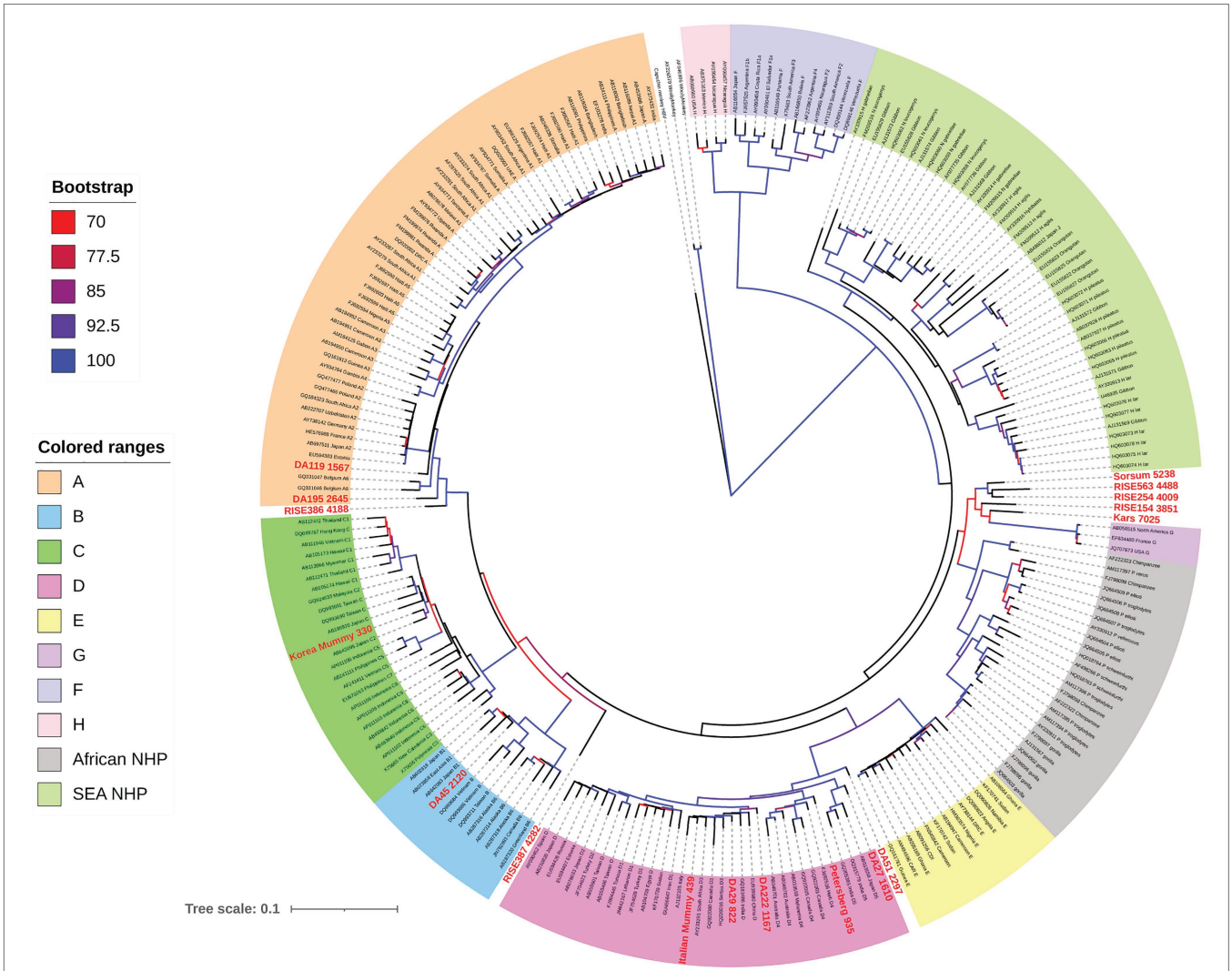

FIGURE 4 | Maximum likelihood (ML) phylogenetic tree of full-length HBV genome sequences (aHBV has shown in bold red font) extracted from human and non-human primate hosts, and reference HBV subgenotype sequences obtained from GenBank. For aHBV, the number after ID represents the approximate age of the source fossil samples. The tree was inferred using IQ-TREE v1.6.5 (Nguyen et al., 2015), which is packaged with ModelFinder (Kalyaanamoorthy et al., 2017) and UFBoot (Hoang et al., 2018). The best-fit model determined by ModelFinder was GTR+F+R5. Clade support was assessed using 1,000 pseudo replicates generated with the UFBoot non-parametric bootstrap procedure, and bootstrap values >70% are shown by a color scale on the tree branches. The ML tree was annotated using iTOL v5.7 (Letunic and Bork, 2019). Branch lengths are scaled to the number of nucleotide substitutions per sequence site.

the "PA" motif at the 3' end of core are also identified as "finger-print" differences in the NHP aHBV. The distinctive amino acid motifs in the S-region and core antigen in NHP compared to human sequences may be a result of species-specific responses to host immune pressure; both the surface and core antigens are highly immunogenic. These motifs could also be potentially a result of convergent evolution as they appear in both the African and Asian NHP HBV lineages, which are separated by human HBV genotypes. The modification at precore position 28 could be associated with a lack of disease in NHP, as a mutation at this position results in complete abrogation of expression of the HBeAg. The HBeAg negative phase in human genotypes is associated with liver disease progression, which does not generally occur in primate HBV infection and so these substitutions could be acting as compensatory changes [see section The Extinct Neolithic Genotype (5–7 kya) and **Table 1**]. Finally, the 11 amino acid deletion in the Pre-S1 gene and overlapping polymerase region could possibly have some effect on species-specific viral entry, as this is adjacent to the essential NTCP binding element on the viral surface gene.

Human HBV is currently grouped into nine genotypes (designated A–I) and one putative genotype J, based on a full genome diversity of more than 7.5% at the nucleotide (nt) level (Okamoto et al., 1988; Norder et al., 1994; Tran et al., 2008; Tatematsu et al., 2009; Kramvis, 2014), and the phylogenetic analyses of aligned full genome sequences are shown in **Figure 4** (Miyakawa and Mizokami, 2003; Norder et al., 2004; Olinger et al., 2008). HBV genotypes are

TABLE 1 | Key distinctive amino acid replacement features in non-human primate HBV sequences of aHBV compared to sequences derived from human genotypes.

Host species or Ancient isolate ID	Archeological Period	Approximate sample age (years)	Genotype	PreS1 deletion (Pol overlap)	Precore amino acid position 28	Precore amino acid position 2	Core amino acid position 113	"PA" motif at 3' end of Core
Human HBV	Modern	-	A–J	Only in Geno D	W/*	Q/*	Mostly "L"	RESQC* except Geno E, F/H, G
NHP* HBV	Modern	-	NHP	Present	Mostly L	Q	"Q" in African "L" in Southeast Asian	PASQC*
Korean Mummy	Early modern	330	C2	Undetermined	W	Q		Undetermined
Italian Mummy	Early modern	439	D3	Present	W	Q		RESQC*
DA29	Medieval	822	D3	Present	W	Q/*	L	RESQC*
Petersberg	Medieval	935	D4	Undetermined	W	Q	L	RESQC*
DA222	Medieval	1,167	D3	Present	W	Q	L	RESQC*
DA119	Medieval	1,567	A	Absent	W	Q	L	RESQC*
DA27	Iron	1,610	D5	Present	W	Q	L	RESQC*
DA45	Iron	2,120	B1	Absent	W	Q	L	RESQC*
DA51	Iron	2,297	D	Present	W	Q	L	RESQC*
DA195	Bronze	2,645	A	Absent	W	Q	L	RESQC*
RISE386	Bronze	4,188	A (ancient)	Absent	W	Q	L	RESQC*
RISE387	Bronze	4,282	A (ancient)	Absent	W/*	*	L	PASQC*
RISE154	Bronze	3,851	NHP (extinct)	Undetermined	L	Q	Q	PASQC*
RISE254	Bronze	4,009	NHP (extinct)	Present	L	Q	Q	PASQC*
RISE563	Bronze	4,488	NHP (extinct)	Present	L	*	Q	PASQC*
Sorsum	Neolithic	5,238	NHP (extinct)	Undetermined	L	Q	Q	PASQC*
Karsdorf	Neolithic	7,025	NHP (extinct)	Undetermined	L	*	Q	PASQC*

**indicates stop codon.*

geographically distributed across the globe (**Figure 5**). Generally, Generally, HBV genotypes are associated with distinct clinical outcomes ranging from mild hepatic disease (subgenotypes B1 and B5) to rapid progression to liver failure and malignancy (genotypes F and C). Also affected by genotype are HBeAg seroconversion rates and the emergence of mutation profiles in the basal core promoter and precore regions of the viral genome associated with HBeAg loss (Rodriguez-Frias et al., 1995; Chu et al., 2003; Liaw and Chu, 2009).

Phylogenetic analyses have shown that most of the genotypes have sufficient diversity to be categorized into subgenotypes differing by at least 4% (Kramvis and Kew, 2005) of their full genome sequence. Genotypes E, G, and the putative J are not subdivided into subgenotypes and this might reflect idiosyncratic or uniquely different epidemiological patterns such as a more recent origin or re-introduction following a relatively recent extinction or loss. The distribution of HBV subgenotypes is also generally geographical (**Figure 5**). For example, subgenotype A1 is found mainly in Africa, Asia, and Latin America with dispersal outside Africa suggested to have occurred as a result of the slave trade (Kramvis and Paraskevis, 2013), while A2 is in Europe and North America (Sugauchi et al., 2004); B1 is found in Japan, especially in the south, while B2, a recombinant with genotype C in the precore-core region, is found in the rest of Asia (Sugauchi et al., 2004); C1 is the dominant strain in South and Southeast Asia, while C2 is found mainly in North Asia especially Korea (Huy et al., 2004; Chan et al., 2005). Subgenotype C4 is found exclusively in Australian Indigenous persons and its envelope protein is possibly regarded as the original "Australia Antigen" (Littlejohn et al., 2014), while C3 is found only in Melanesian populations of the South Pacific. Subgenotypes C3 and C4 most probably diverged from ancestors of subgenotypes C1 and C2 prior to being carried south and east to Melanesia and Australia (Yuen et al., 2019). Subgenotypes D1–D4 are widely distributed globally (Kimbi et al., 2004; Banerjee et al., 2006), with D3 common in Europe and the Mediterranean region while D4 is the dominant subgenotype in Polynesia and Micronesia (Jackson et al., 2020), and also found in India and the Arctic Denes (Banerjee et al., 2006; Osiowy et al., 2011).

ANCIENT HBV ISOLATES

Over the last decade, technological advances have allowed larger-scale analysis of next-generation sequencing data (Qin et al., 2010) including multiple studies utilizing fossilized and mummified tissue and skeletal remains for metagenomics composition. This has provided deeper insights into the identification of diseases from which that particular individual suffered (Cho and Blaser, 2012; Lynch and Pedersen, 2016). Although the analysis can be limited by variations in sample preservation as well as the presence of environmental contamination, studies have obtained well-preserved microbial, viral, and human DNA especially from the dental calculus (Preus et al., 2011; Warinner et al., 2014; Ozga et al., 2016). This has enabled the simultaneous study of pathogen and host. Mummified remains have also presented an opportunity to detect specific pathogens and perform human and microbial genome reconstruction, when combined with genome-wide DNA capture approaches (Bos et al., 2011) such as the

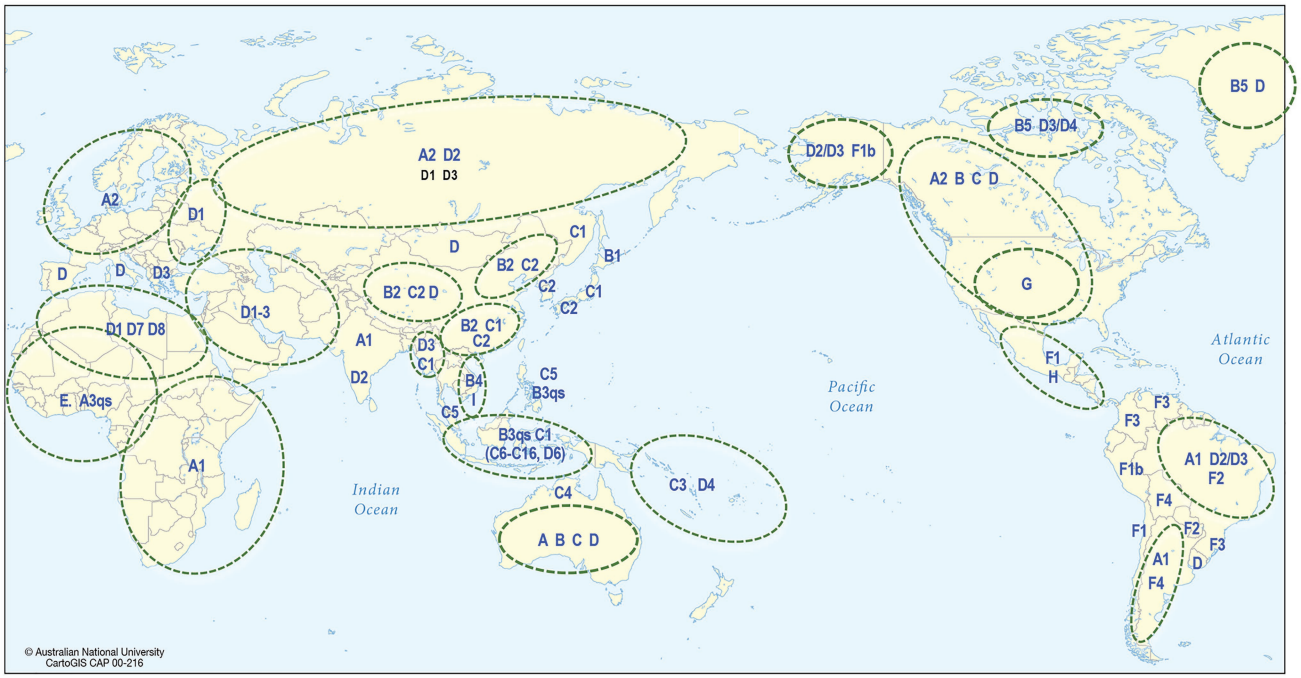

FIGURE 5 | Global distribution of HBV. The map shows the predominant genotype and subgenotypes that have been identified in different countries around the world.

identification of an ancient *Helicobacter pylori* genome from the 5,000-year-old Tyrolean Iceman (Maixner et al., 2016).

For aHBV, the oldest reconstructed genomes have been dated to the Neolithic, 5,000–7,000 years ago (kya; Krause-Kyora et al., 2018) and the Bronze Age 3–4 kya (Muhlemann et al., 2018) using dental and skeletal remains. It is likely that these represent exogenous HBV genomes rather than HBV integrated into the human genome based on the fact that the samples were extracted from teeth. aHBV has also been detected in two preserved mummies, one from Korea and the other from Italy (Kahila Bar-Gal et al., 2012; Patterson Ross et al., 2018) dated from around 400 years ago (ya). Phylogenetic reconstruction of HBV isolated from these samples revealed that the oldest HBV genomes, which were from the Neolithic period and appear extinct, are most closely related to the HBV of African NHP (**Table 1**; Krause-Kyora et al., 2018; Muhlemann et al., 2018). The oldest classifiable genotype was aHBV-A found in Russia dated around 4,200 years ago. For further comparisons, the aHBV from these two studies as well as the isolates identified from the Italian and Korean mummies can be grouped under the following time-frame categories: (a) Extinct Neolithic Genotype, (b) Bronze Age Genotype, (c) Medieval Genotype, and (d) Early Modern Genotype (**Table 1**).

The Extinct Neolithic Genotype (5–7 Kya)

Two isolates were identified from fossils across a narrow geographical range in Western Europe mainly in present-day Germany at Karsdorf (7 kya) and Sorsum (5.2 kya). Although the age of these two strains were estimated to be 2,000 years apart, they showed a higher genomic similarity (6%) to each other and African NHP HBV (<8%) than to any modern human HBV (**Figure 4**; Krause-Kyora et al., 2018). The authors identified some fragments of the two aHBV genomes showing high similarity to human HBV genotypes E and G and proposed that the aHBV from Karsdorf may have been a distant source for the younger Sorsum virus. Given the genetic distances of these aHBV were <8% from African NHP HBV they were not classified as a novel genotype. Nonetheless, they could be classified as an extinct subgenotype of HBV, given they are no longer found circulating in humans. Both these sequences displayed the sequence motifs detected in modern NHP. Although the Karsdorf sequence has a leucine at precore position 28, it also has the BCP mutations at 1762/1764 as well as a precore stop codon (PC_Q*2), which would abrogate expression of the viral HBeAg.

Bronze Age Genotype (Approx. 2–4 Kya)

Isolates RISE563 (Germany), RISE254 (Hungary), and RISE154 (Poland) identified from an independent study (Muhlemann et al., 2018) were also found to be genetically related to the Neolithic aHBV discussed above and are closely related to the African NHP HBV with the corresponding sequence motifs although the RISE563 sequence has a stop codon in the precore gene (PC_Q*2; **Table 1; Figure 4**). Given European great apes have been extinct since the late Miocene (at least 7 million years ago (mya); DeMiguel et al., 2014), the presence of HBV most closely related to African NHP HBV in these locations suggest at some stage in the evolutionary history, a possible African origin for these extinct aHBV genotypes.

Phylogenetic testing of the Bronze Age isolates showed the RISE386/387 samples, which were collected from Bulanovo Russia, had clustered as basal partners within clade genotype A. Interestingly, the RISE387 has some, but not all NHP sequence motifs, while the RISE 386 has predominantly human HBV sequence motifs (**Table 1**). Although the isolate RISE387 does not consistently cluster with genotype A as shown by Datta (2020), it is <8% genetically different from contemporary genotype A isolates. This suggests the geographical origin of genotype A may have been Central Asia, in agreement with studies using contemporary sequences, which also concluded a Middle East/Central Asia origin for genotype A (Kramvis and Paraskevis, 2013; Paraskevis et al., 2013; Kostaki et al., 2018). In this scenario, Muhlemann et al. proposed that the ancestors of subgenotype A1 and quasi-subgenotype A3 would have been carried into Africa following migration from Eurasia (Muhlemann et al., 2018). In addition, the oldest isolate of genotype D (DA51) was identified from a 2,000-year-old Iron Age fossil, and it was considered by these investigators as the ancestral genotype D (Muhlemann et al., 2018). This aHBV was collected from a fossil in Kyrgyzstan, Central Asia, and has all human HBV sequence motifs. Also of interest from this study was the identification of the isolate DA45 that was classified as subgenotype B1 from Mongolia, East Asia, indicating that this non-recombinant subgenotype may have originated in other parts of East Asia before becoming the dominant genotype B in Japan. This finding also indicates that the Asian dominant strains like genotype B had already established their geographical niche over 2 kya.

Medieval Genotype (Approx. 1 Kya)

Both the Krause-Kyora et al. (2018) and Muhlemann et al. (2018) studies identified HBV genomes from fossil samples dated around 1,000 years old. These included a HBV-D4 from Petersburg, Russia, and two HBV-D3 from Kazakhstan in Central Asia (DA29 and DA222). Of interest, the two isolates that were over 1,000-year old clustered as basal taxons in the subgenotype clades, while isolate DA29 (829-year old) clustered among the contemporary HBV-D3.

Early Modern Genotypes

Ancient DNA studies have also provided new perspectives on the more recent evolutionary history of HBV. HBV genomes have been sequenced from a Korean child mummy radiocarbon dated to 330-year BP (±70 years), translating to ca. 1,682 (1,612–1,752; Kahila Bar-Gal et al., 2012) and a 439-year BP (±60 years) Italian mummy from Naples, Italy around the same time frame (Patterson Ross et al., 2018). Phylogenetic analysis of the HBV DNA from the mummified Korean child positioned it within the cluster of modern subgenotype C2 (**Figure 4**). Likewise, the phylogenetic analysis of HBV sequence from the Italian mummy (**Figure 4**) revealed that the HBV sequence reads were of subgenotype D3. These phylogenetic results suggest low long-term mutation rates, with genotypes diversifying over many thousands of years (Paraskevis et al., 2013). This finding highlights that the genomic structure of HBV has strong selective constraints as described in section "The Family Hepadnaviruses." There was no evidence of recombination from either the Korean mummy or Italian mummy HBV genome sequences (Patterson Ross et al., 2018).

MIXED CO-INFECTIONS AND RECOMBINATION

Mixed/Co-infection of HBV

The co-existence of different HBV genotypes, resulting from multiple exposure and super-infection, can lead to exchange of genetic material between two viral strains infecting the same cell resulting in a recombination event (Fares and Holmes, 2002; Hannoun et al., 2002). The actual mechanism by which inter-genotype recombination occurs remains unknown but there are multiple occasions throughout the HBV life cycle, where this may occur (Newbold et al., 1995; Yang and Summers, 1995; reviewed in Littlejohn et al., 2016).

Modern HBV Recombinants

Several HBV recombinants have been assigned their own genotype or subgenotype and are endemic in certain geographic regions. Interestingly, the majority of known HBV recombinants involve HBV genotype C either as a major or minor parental sequence (**Figure 6**). Most notable is the HBV subgenotypes B2-B4, which are recombinants between isolates of HBV genotypes B and C (Sugauchi et al., 2002). These contain the precore and core genes of HBV genotype C and are found throughout Asia (**Figure 6**; Sugauchi et al., 2002). As highlighted above, the subgenotypes B1 and B5 are not recombinants and are mainly confined to Japan and the circumpolar Arctic, respectively, while HBV genotype I has been reported as a triple recombinant containing elements from genotypes A, G, and C and are predominantly found in Vietnam and Southern China (Tran et al., 2008). Recognized HBV recombinants that localize to specific geographical regions have also been described including the various C/D-recombinants circulating in Tibet (Cui et al., 2002; Wang et al., 2005; Zeng et al., 2005). Other genotype C recombinants are shown in **Figure 6**.

The overlapping reading frames and viral regulatory elements embedded within coding regions of HBV genomes give rise to alternating regions of variable and conserved domains that can be described as mosaic structures (Bowyer and Sim, 2000). The more variable of these domains have been associated with recombination events rather than the result of point mutation accumulation. Simmonds and Midgley (2005) have also described the HBV genome as modular, representing a blend of small segments from genomes of different human and primate HBV strains. It is possible that genome mosaicism is an important driving force in the evolutionary history of HBV genotypes. Pathogenesis of chronic hepatitis combined with host-virus interaction among different host populations can result in the selection of different combinations of allelic modules with different properties that would improve viral fitness (Littlejohn et al., 2016).

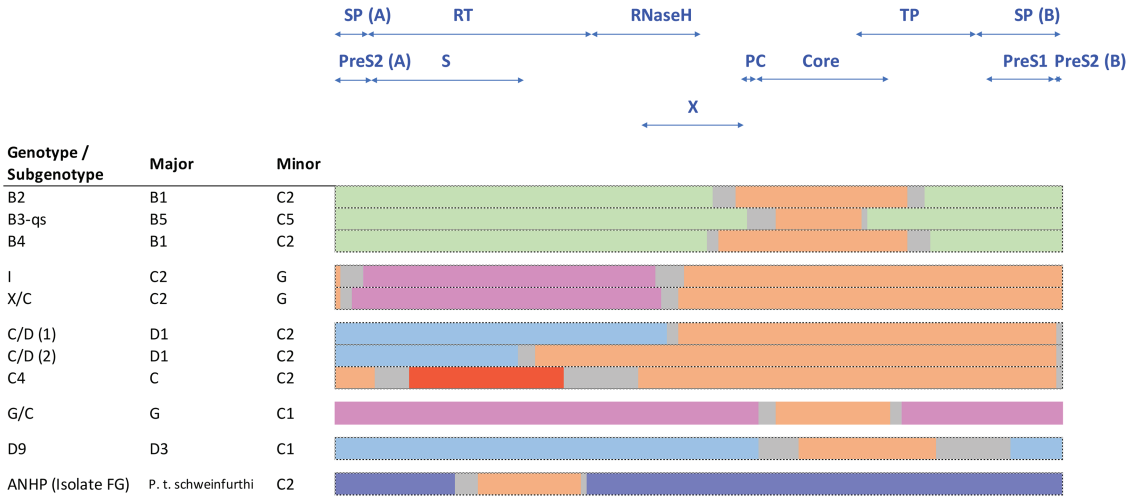

FIGURE 6 | Schematic representation of known HBV genotype C recombinants, showing the conserved regions of the major and minor components. Sequences are color coded by genotype: genotype B (green), C (orange), D (blue), G (pink), J (red), and chimpanzee (purple). The space between the conserved regions is due to uncertainty about the precise recombination breakpoints and is indicated in gray.

Simmonds and Midgley (2005) developed the TreeOrder scan methodology to enable a large number of DNA sequences to be simultaneously analyzed for potential recombination events, on HBV genomes. Most (90%) of the recombinants detected were either hybrids of genotypes B/C or A/D with A and C showing a higher tendency for recombination than other genotypes. These studies mapped up to eight breakpoint hotspots in HBV genomes, where inter-genotype recombination tend to occur. A comprehensive review on global HBV inter-genotypic recombinants has been published, drawing on over 400 complete HBV sequences (Araujo, 2015), highlighting some of the challenges in studying and developing models for the origins of HBV. Inter-species recombination events between human and chimpanzee HBV have also been described. Studies have characterized an HBV recombinant that was isolated from a wild-caught chimpanzee (*P. t. schweinfurthii*) in East Africa (Vartanian et al., 2002; **Figure 6**). The segment of human HBV subgenotype C2 detected in the HBV polymerase genome region of this isolate (FG, **Figure 6**) support the occurrence of recombination among various primate HBV strains. Such recombination events may have played an important role in the current distribution of HBV genotypes and subgenotypes following *Homo sapiens* exit from Africa and subsequent global diaspora. It is interesting to note that Datta, (2020) observed components of chimpanzees and subgenotype C2 HBV in aHBV samples from the Bronze Age (RISE 386/387) that were distantly related to the ancestral genotype A, from around 4 kya isolated from Bulanovo, Russia.

Ancient HBV Recombinants

From the sequences of aHBV published by Krause-Kyora et al. (2018) dating from the Neolithic period, a number of recombination events were identified. Fragments of the Karsdorf sequences showed high similarity to modern human HBV genotypes G and E as well as NHP HBV sequences. Part of the Sorsum HBV genome showed high similarity to the human genotypes G, E, and B. Likewise, Muhlemann et al. (2018) found strong evidence that aHBV sequence of DA51 (ancestral D) and an unknown parent recombined to form the ancient genotype A sequence. Their predicted recombination break points corresponded closely to the polymerase gene implying that the polymerase from an unknown parent and the remainder of the genome from an HBV DA51 ancestor recombined to form ancestral genotype A about 7.4–9 kya (ancestral A).

An independent analysis by Datta (2020) confirmed and extended these findings and suggested a RISE563-like sequence could have contributed to existing genotype I sequences, which is composed of genotypes A, C, and G. Again, this analysis suggested NHP-like HBV sequences were major contributors to the Neolithic and early Bronze Age aHBV sequences with minor contributions from HBV genotype D (subgenotype D6) and E-like sequences. Furthermore, sequences similar to genotype C (subgenotype C2) and gibbon HBV were detected in Neolithic and Bronze Age aHBV isolates from Europe/central Asia (Datta, 2020). Considering genotype I is a triple recombinant of genotype A, C, and G, these findings may not be that surprising after all.

The Australia Antigen and HBV-C4

Indigenous Australians have the oldest continuous living culture outside of Africa with studies showing both Melanesian and Australian Indigenous populations have ancient mitochondrial DNA haplotypes (Merriwether et al., 2005). These two major populations of Sahul are associated with infection with unique HBV genotypes, HBV-C4 for Indigenous Australians and HBV-C3 for Indigenous Melanesians. Davies et al. (2013) reported that HBV isolated from Indigenous Australians who reside in the

Northern Territory (NT) of Australia are exclusively infected with HBV-C4 and this subgenotype has not been found in any other population. HBV-C4 is a recombinant subgenotype containing a genotype C backbone with a genotype J surface gene (Littlejohn et al., 2014). Given genotype J is most similar to HBV isolated from SEA gibbons and orangutans (Tatematsu et al., 2009) and these primates are only found west of the Wallace Line (Wallace, 1876), the faunal boundary line that separates the eco-zones of Asia and Australia. The presence of the HBV-J surface gene in the HBV-C4 recombinant and the exclusivity of HBV-C4 in Indigenous Australians from the NT suggests the recombination event occurred during the migration of anatomically modern humans (AMH) to Australia approx. 50 kya (Yuen et al., 2019). The HBV-C3 in contrast is not a recombinant virus, but is almost universally associated with the Indigenous Melanesian population of Sahul and the Solomon Islands.

PRIMATE ORIGINS, EVOLUTION, AND MIGRATIONS

Given the origin of HBV in primates may be associated with host evolution, it would be reasonable to propose that primate origins, evolution, and migrations may have a role in the evolutionary history of HBV. Nonetheless, further studies are required to allow a definitive conclusion. It is worth noting that co-evolution may be unlikely as the data in **Figure 4** shows the relationships between human and NHP HBV do not follow host phylogeny, thus not supporting a co-evolution theory. A proposed timeline in primate and human evolution based on fossil evidence is shown in **Table 2**.

Asian Primates Origins: Middle Eocene

It is generally held that stem anthropoids arose in Asia 45 mya and that one or several anthropoid groups later migrated to Africa in the late Eocene (Kay, 2012); there is no convincing evidence for the existence of anthropoids in Africa before 38 mya. Specifically, Beard et al. (1994) discovered a fauna of primates from the Eocene dated to around 45 mya from deposits in Shanghuang, Jiangsu Province, China. These fossils included the first Eocene Tarsier as well as fossils subsequently shown to form the basal simians (anthropoids/higher primates), which became known as the Eosimias. These findings established Asia as the biogeographical origin for primate evolution and were subsequently confirmed from fossil discoveries made in Myanmar in 2012 with the identification of *Afrasia djidjidas*, a new form of anthropoid primate that was at least 40 million years old (Chaimanee et al., 2012). Interestingly, this fossil find resembled another early anthropoid, *Afrotaesius lybicus* that had been discovered in the Sahara Desert of Libya (Jaeger et al., 2010) and dental examinations revealed *Afrasia* to be more primitive than *Afrotarsius*, further establishing that early anthropoids originated in Asia first and then quickly colonized Africa. The movement into Africa by these early anthropoids was crucial for subsequent primate and human evolution, with the subsequent emergence of more advanced apes and hominoids there. This ancient migration required the crossing of the Tethys Sea that separated Africa from Eurasia in the late Middle Eocene (Jaeger and Marivaux, 2005). The successful colonization of Africa by Asian anthropoids across the Tethyan marine barrier was possibly by island hopping, similar to that proposed for the arrival of the ancestors of the New World monkeys into the Americas (**Table 2**).

TABLE 2 | The fossil evidence proposes the following timeline in primate and human evolution.

Timeline	Primate and human evolution
45 mya	Primates evolved in Asia and rapidly dispersed and colonized in Africa [a,b,c]
35	Old World and New World monkeys split [d]
20.4	Gibbons split [e]
15.7	Orangutans split [e]
8.8	Gorillas split from the human-bonobo-chimpanzee ancestor [e]
6.4	Australopithices/Homo ancestors split from the bonobo-chimpanzee lineage [e]
1.8	*Homo erectus* emerged from Africa and dispersed and colonized China (Peking Man) and Java (Java Man) [f]
>400 kya	*Homo sapiens neaderthalensis* emerged in the Levant [g]
>300 kya	*Homo denisova* detected in Siberia/Tibet Plateau [h]
>200 kya	*Homo sapiens sapiens* emerged in Africa [i]

[a]Beard et al. (1994); [b]Kay (2012); [c]Chaimanee et al. (2012); [d]Godinot (2020); [e]Kumar and Hedges (2011); [f]Zaim et al. (2011) and Zhu et al. (2018); [g]Green et al. (2010); [h]Reich et al. (2010); [i]Bergstrom et al. (2021).

New World Monkeys

Two New World monkeys have been found to be infected with their own hepadnaviruses; capuchin monkeys and woolly monkeys. The woolly monkeys comprise the genus *Lagothrix* with four species: *Lagothrix lagotricha*, *L. cana*, *L. lugens*, and *L. poeppigii*. These animals are found throughout the northern humid rainforests of South America in Bolivia, Colombia, Ecuador, Venezuela, and Peru. The woolly monkey species *L. lagotricha* was shown in 1998 to be infected with the woolly monkey HBV (WMHBV) found in captive woolly monkeys housed in the Louisville KY zoo (Lanford et al., 1998). Interestingly, six monkeys had been obtained from a zoo in Scotland in 1985. Inoculation of WMHBV into spider monkeys and chimpanzees resulted in the detection of virus replication in both animals and the development of anti-HBc positivity in the spider monkeys. However, no evidence of raised liver function tests (LFT) was obtained detectable hepatic disease. More recently, Souza and colleagues (de Carvalho Dominguez Souza et al., 2018) sampled multiple sites in Bahia State, Brazil and identified a new HBV in the species *Sapajus* (family Cebidae), the capuchin monkey (or the "organ grinder" monkey). Maximum likelihood phylogenetic analysis of full-length primate HBV genomes revealed that the capuchin monkey HBV (CMHBV) and WMHBV clustered together with high statistical support, forming a basal sister lineage to the HBV genotypes (**Figure 4**; de Carvalho Dominguez Souza et al., 2018). The high level of nucleotide divergence between these New World

NHP HBV and other primate HBV may be associated with an early divergence followed by very long-term isolation, possibly soon after the split between Old and New World monkeys around 35 mya. Importantly, though the modern human HBV genotypes F and H are predominantly found in the Americas, the phylogenetic analysis strongly suggests their origin is in the Old World with no evidence of cross-species transmission with the New World NHP HBV.

A frequently asked question is "How did these New World animals arrive in South America from West Africa?" The order *Simiiformes* split into the Platyrrhini (New World monkeys) and Catarrhini (apes and Old World monkeys) about 33–35 mya in Africa and Eurasia. Furthermore, it has been suggested that the ancestors of the Platyrrhini migrated to South America either on a raft of vegetation or *via* a land bridge (Beard, 2004). Two possible rafting routes have been suggested, either across the Atlantic Ocean from Africa (then the Afro-Arabian plate; Godinot, 2020; Seiffert et al., 2020) or across the Caribbean from North America. The former would have the New World monkey ancestors sail on the south equatorial paleo current (SEC), 35–32 mya during the Oligocene (Godinot, 2020). The land bridge route would need to rely on the existence of Atlantic Ocean ridges and a significant fall in the sea level during the Oligocene. At that time, the Atlantic Ocean was less than its present 2,800 km width, by about one-third, at around 1,000 km or less (Beard, 2004). This may have resulted in either a single land bridge or a series of mid-Atlantic islands that could act as stepping stones.

African Origins and Out-of-Africa: Again and Again

Around 6–8 mya, the ancestors of modern humans began to evolve in Africa (**Table 2**), approximately around the same time that *Australopithices* split from the bonobo-chimpanzee lineage. *Homo sapiens* emerged in Africa around 200 kya while the Neanderthals (*H. sapiens neanderthalensis*) evolved from *H. heidelbergensis* around 400 kya, not in Africa but probably in the Levant (Middle East) or Western Asia region. Neanderthals went extinct around 30 kya with the last evidence for their existence in caves in and around Gibraltar. Complete nuclear genomes of archaic humans have been successfully reconstructed from fossil remains of not only Neanderthals (Green et al., 2010), but also Denisovans from the Denisova cave in Southern Siberia (Reich et al., 2010), and an archaic human who lived in Nigeria (Hammer et al., 2011). All three genomes were more than 35,000 years old with clear genetic evidence of interbreeding among archaic humans on at least three occasions (Reich, 2018). Potentially then, as AMH evolved in and left Africa by around 40 kya (northern dispersal route), they interbred with Neanderthals in the Middle East or Arabia before spreading out into Asia and Europe (Green et al., 2010; **Figure 7**). Another group of humans had earlier left Africa around 70 kya (Rasmussen et al., 2011; southern dispersal route) and these humans headed east along the coastal route toward Australia and Melanesia interbreeding with the Denisovans on the way (**Figure 7**). Thus, Indigenous Melanesians inherited DNA from both the Neanderthals and Denisovans during the first wave of expansion into Melanesia, having up to 8% of their DNA coming from these archaic peoples (Reich et al., 2010), which is distinct to the second wave during the Neolithic period (10,000–3,000 BC) when people migrated from East Asia during the Austronesian Expansion to expand into Polynesia (Kirch, 2017). Other populations in Southeast Asia and Oceania have also been found to carry Denisovan admixture, including Indigenous Australians, Polynesians, Fijians, East Indonesians, and Mamanwa but not mainland East Asians, Western Indonesians, Jehai, or Onge (Reich et al., 2011). The successful sequencing of an Indigenous Australian's 100-year-old lock of hair enabled Rasmussen et al. (2011) to confirm the genomic findings discussed above (**Figure 7**). Importantly, they were unable to detect any evidence of European admixture and concluded that Indigenous Australians were descendants of that earlier human dispersal into eastern Asia 62–75 kya. Furthermore, these investigators demonstrated that this dispersal was separate from the one that gave rise to modern Asians 25–38 kya and supported the view that present-day Indigenous Australians descended directly from the earliest humans to occupy Australia, thus representing the oldest continuous population outside of Africa. Finally, the data do support multiple dispersal models with separate dispersals of AMH into eastern Asia; that is, the founder group for the Indigenous Australians and the ancestors of most present-day East Asians.

Spread of Modern Humans: From Great Steppe Nomads to Bronze Age Farmers

The Great Steppe (Mammoth Steppe) stretched from the Atlantic Coast of Europe in the west to Hokkaido in the east. It was the vast steppe ecoregion of Eurasia with extensive temperate grasslands, savannas, and shrublands biome. It supported almost limitless herds of large herbivores such as wild horses, reindeer, musk-ox, and saiga antelopes as well as the woolly mammoth, steppe bison, and woolly rhinoceros. The environment readily supported the hunter-gatherer lifestyle (Oppenheimer, 2003), but it was bitterly cold. By 30 kya, a substantial population of hunter-gatherers was living on and along the Great Steppe. However, by around 18 kya, the Last Glacial Maxim (LGM) began to drive temperatures down even further, causing massive movement of peoples south into Central and Southern Europe taking refuge from the coming of the Ice Age, and adopting a predominantly farming lifestyle by around 5 kya. Agriculture arose independently in at least 11 "centres of origins" scattered over all inhabited continents except Australia from around 11–12 kya (Larson et al., 2014). The benefits of farming over hunter-gathering methods for food triggered the agricultural expansion which spanned over 6,000 years. However, many controversies exist regarding the dispersal process from these regions. The Fertile Crescent (Mesopotamia and Ancient Egypt) in the Levant is considered as one of

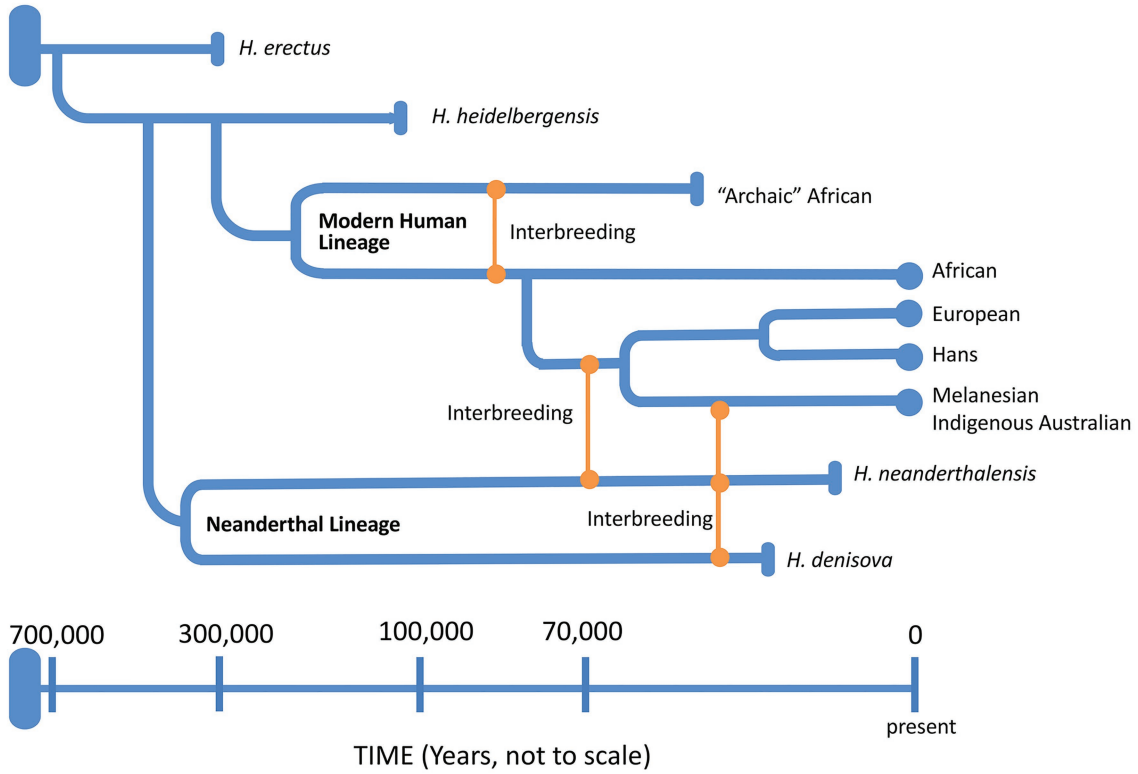

FIGURE 7 | A schematic representation of the genus *Homo* evolution timeline over the last 700,000 years, showing the admixture relationships between *Homo sapiens*, during migration out of Africa, and other archaic humans. At the end of each lineage, circle marks represent modern humans and rod marks represent archaic humans who are now extinct.

the earliest "cradle of civilization," and the early domestication of plants and animals began in the region around 11 kya. Based on craniofacial form studies, it was also one of the earliest "centre of origin," where the early farming populations migrated westward into Europe and North Africa, northward to Crimea, and also eastward to Mongolia (Brace et al., 2006). Of these, the Neolithic transition from the Near East into Europe that spanned between 9 and 5.5 kya has been the most studied. However, both archeological (Brace et al., 2006) and genetic (Dupanloup et al., 2004) studies have confirmed that the Neolithic and Bronze Age peoples of Europe are not closely related to the modern inhabitants. Nonetheless, these studies supported by mathematical modeling (Fort, 2015) revealed that the transition took place *via* demic and cultural diffusion. The contribution levels of Paleolithic and Neolithic populations to the gene pool of modern European populations were sufficient to support the theory that as the Neolithic populations diffused from the Near East into Europe, they had lived and interbred with the existing hunter-gatherers who had inhabited the regions since Late Pleistocene, and who over time had absorbed the cultures and agricultural way of life of these migrants. Although the dispersion routes from the other "centres of origins" around the globe have not been studied as extensively, it is likely that demic and cultural diffusion had also played an important role in the geographical distribution of modern-day humans, and by association, the distribution of HBV genotypes.

THEORIES OF HBV EVOLUTION

A number of theories have been proposed for the origin and evolution of HBV, as reviewed by Simmonds (2001). However, the discovery of hepadnaviruses in a number of reptile and amphibian species has radically altered the reasoning behind these theories, revealing deep ancestry across the animal kingdom. The presence of endogenous hepadnavirus species in birds and reptiles suggests an origin >200 mya followed by co-evolution with the host species within the reptile/fish/bird host lineages (Geoghegan et al., 2017; Lauber et al., 2017). The orthohepadnaviruses detected from a number of bat species, which are genetically most related to primate HBV, also support co-evolution of *hepadnaviridae* in the mammalian lineage (Drexler et al., 2013).

When it comes to HBV evolution within the primate lineage, a theory involving co-evolution is unlikely. The lack of distinct evolutionary pathways between human HBV and those from African and Southeast Asian NHP species undermines it (Rasche et al., 2016). Cross-species transmission cases have been reported in multiple studies (Grethe et al., 2000; Hu et al., 2000; Takahashi et al., 2000; Magiorkinis et al., 2005;

Simmonds and Midgley, 2005; Dupinay et al., 2013). It is also worth noting that geographical regions, where there are increased opportunities for HBV transmission between human and NHP tend to be areas, where HBV prevalence is high among humans (Southeast Asia and Africa), thus supporting the notion that cross-species transmission may have played an important role in the evolution of modern-day primate HBV genotypes.

Until the discovery of aHBV, it was presumed the modern human HBV genotypes co-evolved and spread with AMH out of Africa (Norder et al., 1994; Magnius and Norder, 1995). However, the finding of five Neolithic and Bronze Age aHBV in Central and Western Europe that could be classified as an extinct genotype of NHP-like aHBV supports a substantially more complex evolutionary history.

Primate HBV is the youngest lineage of the family *Hepadnaviridae*, and its origin is likely to be sometime in the Middle Pleistocene era. Paraskevis et al. had estimated an origin time of 33.6 kya for the primate HBV of humans and Old World NHP species (Paraskevis et al., 2013). However, Yuen et al. had estimated an origin time of 59 kya for the HBV subgenotype C4 alone (Yuen et al., 2019), thus implying the origin time of primate HBV to be substantially older. Such a large discrepancy between the two studies was probably the consequence of different fossil calibration data and HBV sequence datasets used during the phylogenetic inferencing process. Nonetheless, combining the recent findings of aHBV from Krause-Kyora et al. (2018) and Muhlemann et al. (2018) with the knowledge of AMH migration histories have enabled the proposition of an alternative model for the origin of primate HBV. We propose the origin and evolution of modern-day human and Old World NHP HBV could be driven by major AMH migration events, including the early human migration out of Africa during the Upper Paleolithic era (Bergstrom et al., 2021) and the agricultural expansion during the Neolithic (Brace et al., 2006; Larson et al., 2014).

In the scenario, that the unique subgenotype of Indigenous Australians HBV-C4 had originated in Island Southeast Asia and entered Australia around 50 kya, as inferred by Yuen et al. (2019), imply two important events. Firstly, ancestors of the Asian dominant HBV (genotypes B and C) had already been established in Asia during this time period. Secondly, the origin time of Old World primate HBV is substantially older, supporting the theory of an Out of Africa origin. The current geographical distribution of human HBV genotypes also supports the two main routes of human migration out of Africa, with the northern route introducing the ancestral strains of the European dominant HBV into Eurasia (genotypes A and D) and the ancestral strains of Asian dominant HBV introduced along the Indian Ocean coastline into Oceania and Southeast Asia, often referred to as the southern dispersal route. Although the origin of HBV genotypes F and H remains enigmatic, the phylogenetic analysis supports their origin being in the Old World and had diverged from the same common ancestral strain of HBV that gave rise to the present-day human and Old World NHP HBV, and possibly followed by long-term isolation. The long-term isolation event may be associated with the "Beringian standstill" hypothesis (Tamm et al., 2007) that suggests the ancestors of Native Americans had been geographically isolated on the Beringian land bridge for millennia during the past glacial maximum before dispersal into the American continents.

The agricultural revolution that occurred throughout the Neolithic and Bronze Age (Larson et al., 2014) had played an important role in the global dispersion of HBV genotypes. This has been previously noted in phylogeographic studies of genotypes A and D (Kostaki et al., 2018). Prior to this, ancient hunter-gatherer societies were likely to be small and sparsely dispersed. The spread of farming societies during these time periods coincided with population growth worldwide, which in turn enabled the global spread of the ancestral strains of human HBV. The discovery of a potentially extinct genotype most closely related to African NHP HBV in human fossils from the Neolithic and Bronze Age suggests numerous ghost lineages may have existed in the past, and that the landscape of HBV genotypes and subgenotypes in this era was vastly different to today. This is an important fact to take into account during the inferencing process of HBV evolutionary history. The extinct African NHP-like genotype isolates detected in human fossils with ages that spanned over 2,000 years in Central and Western Europe also suggest that it may have been endemic in the region and could have been introduced from the Fertile Crescent, a region with frequent human traffic between Africa and the Levant, as farming cultures diffused from the Near East into Europe (Brace et al., 2006).

The final shaping of the current geographical distribution of HBV genotypes was the result of numerous more specific population-movements post-Bronze Age, rather than by a single major migration event. These include the Medieval Great Migrations in Europe during the first millennium AD, early historical invasions, and slave trades.

SUMMARY AND CONCLUSION

The recent papers on aHBV from Krause-Kyora et al. (2018) and Muhlemann et al. (2018) provide a unique contribution to the HBV field and novel insights into the origin and dispersal of HBV over the last 7,000 years. These Neolithic and Bronze Age HBV genomes are most closely related to those infecting African NHP, have no close relatives today, possibly went extinct and appear to be generated by recombination. Both papers clearly demonstrate that humans throughout Eurasia have been widely infected with HBV for 1,000 years. Surprisingly, the geographical locations of a number of aHBV genotypes did not match present-day genotypic distribution. For example, genotype A aHBV was identified in South West Russia over 4 kya and the authors propose that it may have descended from an ancient Eurasian tribal group derived from the western Scythian culture, who themselves descended from a Western steppe population (Damgaard et al., 2018), and not from African ancestors as previously thought (Hannoun et al., 2005; Zehender et al., 2015). Interestingly, the three oldest genotype A aHBV strains

studied (DA195, RISE386, and RISE387; Muhlemann et al., 2018) lacked the 6 nt insertion characteristic of modern genotype A (Kramvis, 2014) as did the youngest isolate DA119 discovered in this study. These investigators go on to suggest that the ancestors of subgenotypes A1 and A3 would have then been carried *via* migration from Western Eurasia (Pickrell et al., 2014) subsequently, into Africa. The introduction of genotype A into Africa from Western Eurasia is supported by the phylogeographic analysis performed by Kostaki et al. (2018).

All genotype D isolates were found in Central Asia including the ancestral genotype D isolate DA51 from Kyrgyzstan around 2.2 kya supporting the notion of subsequent spread into India and onto the Pacific. The two Neolithic isolates discovered by Krause-Kyora et al. have closely linked to African NHP over 7 kya with no human association. These NHP sequences provided a major contribution to the Neolithic and early Bronze Age aHBV, suggesting an African origin for these extinct genotypes (see **Table 1**). Finally, Muhlemann et al. (2018) identified ancestral A in isolates RISE386/387, while Datta (2020) further identified chimpanzee and genotype C2 sequences. The ancestors of DA51 with an unknown parent sequence combined to form ancestral genotype A, the oldest known genotype. It is thus hard not to accept the important role of NHP in the evolutionary history of human HBV (**Table 1**). Modern studies have consistently shown that genotype C is the oldest of the modern HBVs but the data from the aHBV sequences would support a series of successful zoonoses from primates to humans, followed by recombination events with other HBVs, both primate and human, in the 9 genotypes of human HBVs recognizable today.

AUTHOR CONTRIBUTIONS

SL: conceptualized and drafted the manuscript. LY: generation of figures. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

The authors would like to thank Tania Candy for her editorial work and support.

REFERENCES

Aiewsakun, P., and Katzourakis, A. (2016). Time-dependent rate phenomenon in viruses. *J. Virol.* 90, 7184–7195. doi: 10.1128/JVI.00593-16

Aiewsakun, P., and Katzourakis, A. (2017). Marine origin of retroviruses in the early Palaeozoic Era. *Nat. Commun.* 8:13954. doi: 10.1038/ncomms13954

Araujo, N. M. (2015). Hepatitis B virus intergenotypic recombinants worldwide: an overview. *Infect. Genet. Evol.* 36, 500–510. doi: 10.1016/j.meegid.2015.08.024

Banerjee, A., Kurbanov, F., Datta, S., Chandra, P. K., Tanaka, Y., Mizokami, M., et al. (2006). Phylogenetic relatedness and genetic diversity of hepatitis B virus isolates in Eastern India. *J. Med. Virol.* 78, 1164–1174. doi: 10.1002/jmv.20677

Beard, K. C. (2004). *The Hunt for the Dawn Monkey: Unearthing the Origins of Monkeys, Apes and Humans*. LA: University of California Press.

Beard, K. C., Qi, T., Dawson, M. R., Wang, B., and Li, C. (1994). A diverse new primate fauna from middle Eocene fissure-fillings in southeastern China. *Nature* 368, 604–609. doi: 10.1038/368604a0

Bergstrom, A., Stringer, C., Hajdinjak, M., Scerri, E. M. L., and Skoglund, P. (2021). Origins of modern human ancestry. *Nature* 590, 229–237. doi: 10.1038/s41586-021-03244-5

Bock, C. T., Schranz, P., Schroder, C. H., and Zentgraf, H. (1994). Hepatitis B virus genome is organized into nucleosomes in the nucleus of the infected cell. *Virus Genes* 8, 215–229. doi: 10.1007/BF01703079

Bos, K. I., Schuenemann, V. J., Golding, G. B., Burbano, H. A., Waglechner, N., Coombes, B. K., et al. (2011). A draft genome of Yersinia pestis from victims of the black death. *Nature* 478, 506–510. doi: 10.1038/nature10549

Bowyer, S. M., and Sim, J. G. (2000). Relationships within and between genotypes of hepatitis B virus at points across the genome: footprints of recombination in certain isolates. *J. Gen. Virol.* 81, 379–392. doi: 10.1099/0022-1317-81-2-379

Brace, C. L., Seguchi, N., Quintyn, C. B., Fox, S. C., Nelson, A. R., Manolis, S. K., et al. (2006). The questionable contribution of the Neolithic and the bronze age to European craniofacial form. *Proc. Natl. Acad. Sci. U. S. A.* 103, 242–247. doi: 10.1073/pnas.0509801102

Chaimanee, Y., Chavasseau, O., Beard, K. C., Kyaw, A. A., Soe, A. N., Sein, C., et al. (2012). Late middle Eocene primate from Myanmar and the initial anthropoid colonization of Africa. *Proc. Natl. Acad. Sci. U. S. A.* 109, 10293–10297. doi: 10.1073/pnas.1200644109

Chan, H. L., Tsui, S. K., Tse, C. H., Ng, E. Y., Au, T. C., Yuen, L., et al. (2005). Epidemiological and virological characteristics of 2 subgroups of hepatitis B virus genotype C. *J. Infect. Dis.* 191, 2022–2032. doi: 10.1086/430324

Chen, C. J., Iloeje, U. H., and Yang, H. I. (2007). Long-term outcomes in hepatitis B: the REVEAL-HBV study. *Clin. Liver Dis.* 11, 797–816. doi: 10.1016/j.cld.2007.08.005

Cho, I., and Blaser, M. J. (2012). The human microbiome: at the interface of health and disease. *Nat. Rev. Genet.* 13, 260–270. doi: 10.1038/nrg3182

Chu, C. J., Keeffe, E. B., Han, S. H., Perrillo, R. P., Min, A. D., Soldevila-Pico, C., et al. (2003). Hepatitis B virus genotypes in the United States: results of a nationwide study. *Gastroenterology* 125, 444–451. doi: 10.1016/S0016-5085(03)00895-3

Cui, C., Shi, J., Hui, L., Xi, H., Zhuoma, Quni, et al. (2002). The dominant hepatitis B virus genotype identified in Tibet is a C/D hybrid. *J. Gen. Virol.* 83, 2773–2777. doi: 10.1099/0022-1317-83-11-2773

Damgaard, P. B., Marchi, N., Rasmussen, S., Peyrot, M., Renaud, G., Korneliussen, T., et al. (2018). 137 ancient human genomes from across the Eurasian steppes. *Nature* 557, 369–374. doi: 10.1038/s41586-018-0094-2

Datta, S. (2020). Excavating new facts from ancient hepatitis B virus sequences. *Virology* 549, 89–99. doi: 10.1016/j.virol.2020.08.002

Davies, J., Littlejohn, M., Locarnini, S. A., Whiting, S., Hajkowicz, K., Cowie, B. C., et al. (2013). Molecular epidemiology of hepatitis B in the indigenous people of northern Australia. *J. Gastroenterol. Hepatol.* 28, 1234–1241. doi: 10.1111/jgh.12177

de Carvalho Dominguez Souza, B. F., Konig, A., Rasche, A., de Oliveira Carneiro, Stephan, N., Corman, V. M., et al. (2018). A novel hepatitis B virus species discovered in capuchin monkeys sheds new light on the evolution of primate hepadnaviruses. *J. Hepatol.* 68, 1114–1122. doi: 10.1016/j.jhep.2018.01.029

DeMiguel, D., Alba, D. M., and Moya-Sola, S. (2014). Dietary specialization during the evolution of Western Eurasian hominoids and the extinction of European great apes. *PLoS One* 9:e97442. doi: 10.1371/journal.pone.0097442

Drexler, J. F., Geipel, A., Konig, A., Corman, V. M., van Riel, Leijten, L. M., et al. (2013). Bats carry pathogenic hepadnaviruses antigenically related to hepatitis B virus and capable of infecting human hepatocytes. *Proc. Natl. Acad. Sci. U. S. A.* 110, 16151–16156. doi: 10.1073/pnas.1308049110

Dupanloup, I., Bertorelle, G., Chikhi, L., and Barbujani, G. (2004). Estimating the impact of prehistoric admixture on the genome of Europeans. *Mol. Biol. Evol.* 21, 1361–1372. doi: 10.1093/molbev/msh135

Dupinay, T., Gheit, T., Roques, P., Cova, L., Chevallier-Queyron, P., Tasahsu, S. I., et al. (2013). Discovery of naturally occurring transmissible chronic hepatitis B virus infection among Macaca fascicularis from Mauritius Island. *Hepatology* 58, 1610–1620. doi: 10.1002/hep.26428

Fares, M. A., and Holmes, E. C. (2002). A revised evolutionary history of hepatitis B virus (HBV). *J. Mol. Evol.* 54, 807–814. doi: 10.1007/s00239-001-0084-z

Fort, J. (2015). Demic and cultural diffusion propagated the Neolithic transition across different regions of Europe. *J. R. Soc. Interface* 12. doi: 10.1098/rsif.2015.0166

Geoghegan, J. L., Duchene, S., and Holmes, E. C. (2017). Comparative analysis estimates the relative frequencies of co-divergence and cross-species transmission within viral families. *PLoS Pathog.* 13:e1006215. doi: 10.1371/journal.ppat.1006215

Godinot, M. (2020). Rafting on a wide and wild ocean. *Science* 368, 136–137. doi: 10.1126/science.abb4107

Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., et al. (2010). A draft sequence of the Neandertal genome. *Science* 328, 710–722. doi: 10.1126/science.1188021

Grethe, S., Heckel, J. O., Rietschel, W., and Hufert, F. T. (2000). Molecular epidemiology of hepatitis B virus variants in nonhuman primates. *J. Virol.* 74, 5377–5381. doi: 10.1128/JVI.74.11.5377-5381.2000

Hammer, M. F., Woerner, A. E., Mendez, F. L., Watkins, J. C., and Wall, J. D. (2011). Genetic evidence for archaic admixture in Africa. *Proc. Natl. Acad. Sci. U. S. A.* 108, 15123–15128. doi: 10.1073/pnas.1109300108

Hannoun, C., Horal, P., and Lindh, M. (2000). Long-term mutation rates in the hepatitis B virus genome. *J. Gen. Virol.* 81, 75–83. doi: 10.1099/0022-1317-81-1-75

Hannoun, C., Krogsgaard, K., Horal, P., and Lindh, M.Interpred Trial Group (2002). Genotype mixtures of hepatitis B virus in patients treated with interferon. *J. Infect. Dis.* 186, 752–759. doi: 10.1086/342599

Hannoun, C., Soderstrom, A., Norkrans, G., and Lindh, M. (2005). Phylogeny of African complete genomes reveals a West African genotype A subtype of hepatitis B virus and relatedness between Somali and Asian A1 sequences. *J. Gen. Virol.* 86, 2163–2167. doi: 10.1099/vir.0.80972-0

Harrison, A., Lemey, P., Hurles, M., Moyes, C., Horn, S., Pryor, J., et al. (2011). Genomic analysis of hepatitis B virus reveals antigen state and genotype as sources of evolutionary rate variation. *Viruses* 3, 83–101. doi: 10.3390/v3020083

Ho, S. Y., Lanfear, R., Bromham, L., Phillips, M. J., Soubrier, J., Rodrigo, A. G., et al. (2011). Time-dependent rates of molecular evolution. *Mol. Ecol.* 20, 3087–3101. doi: 10.1111/j.1365-294X.2011.05178.x

Hoang, D. T., Chernomor, O., von Haeseler, Minh, B. Q., and Vinh, L. S. (2018). UFBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* 35, 518–522. doi: 10.1093/molbev/msx281

Hu, X., Javadian, A., Gagneux, P., and Robertson, B. H. (2001). Paired chimpanzee hepatitis B virus (ChHBV) and mtDNA sequences suggest different ChHBV genetic variants are found in geographically distinct chimpanzee subspecies. *Virus Res.* 79, 103–108. doi: 10.1016/s0168-1702(01)00334-3

Hu, X., Margolis, H. S., Purcell, R. H., Ebert, J., and Robertson, B. H. (2000). Identification of hepatitis B virus indigenous to chimpanzees. *Proc. Natl. Acad. Sci. U. S. A.* 97, 1661–1664. doi: 10.1073/pnas.97.4.1661

Huy, T. T., Ushijima, H., Quang, V. X., Win, K. M., Luengrojanakul, P., Kikuchi, K., et al. (2004). Genotype C of hepatitis B virus can be classified into at least two subgroups. *J. Gen. Virol.* 85, 283–292. doi: 10.1099/vir.0.19633-0

Hyams, K. C. (1995). Risks of chronicity following acute hepatitis B virus infection: a review. *Clin. Infect. Dis.* 20, 992–1000. doi: 10.1093/clinids/20.4.992

Jackson, K., Tekoaua, R., Holgate, T., Edwards, R., Yuen, L., Lee, A., et al. (2020). Hepatitis B and D in the Pacific Islands of Kiribati. *J. Clin. Virol.* 129:104527. doi: 10.1016/j.jcv.2020.104527

Jaeger, J. J., Beard, K. C., Chaimanee, Y., Salem, M., Benammi, M., Hlal, O., et al. (2010). Late middle Eocene epoch of Libya yields earliest known radiation of African anthropoids. *Nature* 467, 1095–1098. doi: 10.1038/nature09425

Jaeger, J. J., and Marivaux, L. (2005). Paleontology. Shaking the earliest branches of anthropoid primate evolution. *Science* 310, 244–245. doi: 10.1126/science.1118124

Kahila Bar-Gal, G., Kim, M. J., Klein, A., Shin, D. H., Oh, C. S., Kim, J. W., et al. (2012). Tracing hepatitis B virus to the 16th century in a Korean mummy. *Hepatology* 56, 1671–1680. doi: 10.1002/hep.25852

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, and Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14, 587–589. doi: 10.1038/nmeth.4285

Kay, R. F. (2012). Evidence for an Asian origin of stem anthropoids. *Proc. Natl. Acad. Sci. U. S. A.* 109, 10132–10133. doi: 10.1073/pnas.1207933109

Kimbi, G. C., Kramvis, A., and Kew, M. C. (2004). Distinctive sequence characteristics of subgenotype A1 isolates of hepatitis B virus from South Africa. *J. Gen. Virol.* 85, 1211–1220. doi: 10.1099/vir.0.19749-0

Kirch, P. V. (2017). *On the Road of the Winds: An Archaeological History of the Pacific Islands Before European Contact*. London: University of California Press.

Kostaki, E. G., Karamitros, T., Stefanou, G., Mamais, I., Angelis, K., Hatzakis, A., et al. (2018). Unravelling the history of hepatitis B virus genotypes A and D infection using a full-genome phylogenetic and phylogeographic approach. *Elife* 7:e36709. doi: 10.7554/eLife.36709

Kramvis, A. (2014). Genotypes and genetic variability of hepatitis B virus. *Intervirology* 57, 141–150. doi: 10.1159/000360947

Kramvis, A., and Kew, M. C. (2005). Relationship of genotypes of hepatitis B virus to mutations, disease progression and response to antiviral therapy. *J. Viral Hepat.* 12, 456–464. doi: 10.1111/j.1365-2893.2005.00624.x

Kramvis, A., and Paraskevis, D. (2013). Subgenotype A1 of HBV-tracing human migrations in and out of Africa. *Antivir. Ther.* 18, 513–521. doi: 10.3851/IMP2657

Krause-Kyora, B., Susat, J., Key, F. M., Kuhnert, D., Bosse, E., Immel, A., et al. (2018). Neolithic and medieval virus genomes reveal complex evolution of hepatitis B. *Elife* 7. doi: 10.7554/eLife.36666

Kumar, S., and Hedges, S. B. (2011). TimeTree2: species divergence times on the iPhone. *Bioinformatics* 27, 2023–2024. doi: 10.1093/bioinformatics/btr315

Lanford, R. E., Chavez, D., Brasky, K. M., Burns, R. B. 3rd, and Rico-Hesse, R. (1998). Isolation of a hepadnavirus from the woolly monkey, a New World primate. *Proc. Natl. Acad. Sci. U. S. A.* 95, 5757–5761. doi: 10.1073/pnas.95.10.5757

Larson, G., Piperno, D. R., Allaby, R. G., Purugganan, M. D., Andersson, L., Arroyo-Kalin, M., et al. (2014). Current perspectives and the future of domestication studies. *Proc. Natl. Acad. Sci. U. S. A.* 111, 6139–6146. doi: 10.1073/pnas.1323964111

Lauber, C., Seitz, S., Mattei, S., Suh, A., Beck, J., Herstein, J., et al. (2017). Deciphering the origin and evolution of hepatitis B viruses by means of a family of non-enveloped fish viruses. *Cell Host Microbe* 22, 387.e6–399.e6. doi: 10.1016/j.chom.2017.07.019

Letunic, I., and Bork, P. (2019). Interactive tree of life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* 47, W256–W259. doi: 10.1093/nar/gkz239

Li, S., Wang, Z., Li, Y., and Ding, G. (2017). Adaptive evolution of proteins in hepatitis B virus during divergence of genotypes. *Sci. Rep.* 7:1990. doi: 10.1038/s41598-017-02012-8

Liaw, Y. F., and Chu, C. M. (2009). Hepatitis B virus infection. *Lancet* 373, 582–592. doi: 10.1016/S0140-6736(09)60207-5

Littlejohn, M., Davies, J., Yuen, L., Edwards, R., Sozzi, T., Jackson, K., et al. (2014). Molecular virology of hepatitis B virus, sub-genotype C4 in northern Australian indigenous populations. *J. Med. Virol.* 86, 695–706. doi: 10.1002/jmv.23888

Littlejohn, M., Locarnini, S., and Yuen, L. (2016). Origins and evolution of hepatitis B virus and hepatitis D virus. *Cold Spring Harb. Perspect. Med.* 6:a021360. doi: 10.1101/cshperspect.a021360

Locarnini, S., Littlejohn, M., Aziz, M. N., and Yuen, L. (2013). Possible origins and evolution of the hepatitis B virus (HBV). *Semin. Cancer Biol.* 23, 561–575. doi: 10.1016/j.semcancer.2013.08.006

Lynch, S. V., and Pedersen, O. (2016). The human intestinal microbiome in health and disease. *N. Engl. J. Med.* 375, 2369–2379. doi: 10.1056/NEJMra1600266

MacDonald, D. M., Holmes, E. C., Lewis, J. C., and Simmonds, P. (2000). Detection of hepatitis B virus infection in wild-born chimpanzees (*Pan troglodytes* verus): phylogenetic relationships with human and other primate genotypes. *J. Virol.* 74, 4253–4257. doi: 10.1128/JVI.74.9.4253-4257.2000

Magiorkinis, E. N., Magiorkinis, G. N., Paraskevis, D. N., and Hatzakis, A. E. (2005). Re-analysis of a human hepatitis B virus (HBV) isolate from an

East African wild born *Pan troglodytes schweinfurthii*: evidence for interspecies recombination between HBV infecting chimpanzee and human. *Gene* 349, 165–171. doi: 10.1016/j.gene.2004.12.021

Magnius, L., Mason, W. S., Taylor, J., Kann, M., Glebe, D., Deny, P., et al. (2020). ICTV virus taxonomy profile: Hepadnaviridae. *J. Gen. Virol.* 101, 571–572. doi: 10.1099/jgv.0.001415

Magnius, L. O., and Norder, H. (1995). Subtypes, genotypes and molecular epidemiology of the hepatitis B virus as reflected by sequence variability of the S-gene. *Intervirology* 38, 24–34. doi: 10.1159/000150411

Maixner, F., Krause-Kyora, B., Turaev, D., Herbig, A., Hoopmann, M. R., Hallows, J. L., et al. (2016). The 5300-year-old helicobacter pylori genome of the iceman. *Science* 351, 162–165. doi: 10.1126/science.aad2545

McMahon, B. J., Alward, W. L., Hall, D. B., Heyward, W. L., Bender, T. R., Francis, D. P., et al. (1985). Acute hepatitis B virus infection: relation of age to the clinical expression of disease and subsequent development of the carrier state. *J. Infect. Dis.* 151, 599–603. doi: 10.1093/infdis/151.4.599

Merriwether, D. A., Hodgson, J. A., Friedlaender, F. R., Allaby, R., Cerchio, S., Koki, G., et al. (2005). Ancient mitochondrial M haplogroups identified in the Southwest Pacific. *Proc. Natl. Acad. Sci. U. S. A.* 102, 13034–13039. doi: 10.1073/pnas.0506195102

Miyakawa, Y., and Mizokami, M. (2003). Classifying hepatitis B virus genotypes. *Intervirology* 46, 329–338. doi: 10.1159/000074988

Mizokami, M., Orito, E., Ohba, K., Ikeo, K., Lau, J. Y., and Gojobori, T. (1997). Constrained evolution with respect to gene overlap of hepatitis B virus. *J. Mol. Evol.* 44, S83–S90. doi: 10.1007/pl00000061

Muhlemann, B., Jones, T. C., Damgaard, P. B., Allentoft, M. E., Shevnina, I., Logvin, A., et al. (2018). Ancient hepatitis B viruses from the bronze age to the medieval period. *Nature* 557, 418–423. doi: 10.1038/s41586-018-0097-z

Newbold, J. E., Xin, H., Tencza, M., Sherman, G., Dean, J., Bowden, S., et al. (1995). The covalently closed duplex form of the hepadnavirus genome exists in situ as a heterogeneous population of viral minichromosomes. *J. Virol.* 69, 3350–3357. doi: 10.1128/JVI.69.6.3350-3357.1995

Nguyen, L. T., Schmidt, H. A., von Haeseler, and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300

Norder, H., Courouce, A. M., Coursaget, P., Echevarria, J. M., Lee, S. D., Mushahwar, I. K., et al. (2004). Genetic diversity of hepatitis B virus strains derived worldwide: genotypes, subgenotypes, and HBsAg subtypes. *Intervirology* 47, 289–309. doi: 10.1159/000080872

Norder, H., Courouce, A. M., and Magnius, L. O. (1994). Complete genomes, phylogenetic relatedness, and structural proteins of six strains of the hepatitis B virus, four of which represent two new genotypes. *Virology* 198, 489–503. doi: 10.1006/viro.1994.1060

Norder, H., Ebert, J. W., Fields, H. A., Mushahwar, I. K., and Magnius, L. O. (1996). Complete sequencing of a gibbon hepatitis B virus genome reveals a unique genotype distantly related to the chimpanzee hepatitis B virus. *Virology* 218, 214–223. doi: 10.1006/viro.1996.0181

Okamoto, H., Tsuda, F., Sakugawa, H., Sastrosoewignjo, R. I., Imai, M., Miyakawa, Y., et al. (1988). Typing hepatitis B virus by homology in nucleotide sequence: comparison of surface antigen subtypes. *J. Gen. Virol.* 69, 2575–2583. doi: 10.1099/0022-1317-69-10-2575

Olinger, C. M., Jutavijittum, P., Hubschen, J. M., Yousukh, A., Samountry, B., Thammavong, T., et al. (2008). Possible new hepatitis B virus genotype, Southeast Asia. *Emerg. Infect. Dis.* 14, 1777–1780. doi: 10.3201/eid1411.080437

Oppenheimer, S. (2003). *Out of Eden*. London: Robinson.

Osiowy, C., Larke, B., and Giles, E. (2011). Distinct geographical and demographic distribution of hepatitis B virus genotypes in the Canadian Arctic as revealed through an extensive molecular epidemiological survey. *J. Viral Hepat.* 18, e11–e19. doi: 10.1111/j.1365-2893.2010.01356.x

Ozga, A. T., Nieves-Colon, M. A., Honap, T. P., Sankaranarayanan, K., Hofman, C. A., Milner, G. R., et al. (2016). Successful enrichment and recovery of whole mitochondrial genomes from ancient human dental calculus. *Am. J. Phys. Anthropol.* 160, 220–228. doi: 10.1002/ajpa.22960

Paraskevis, D., Magiorkinis, G., Magiorkinis, E., Ho, S. Y., Belshaw, R., Allain, J. P., et al. (2013). Dating the origin and dispersal of hepatitis B virus infection in humans and primates. *Hepatology* 57, 908–916. doi: 10.1002/hep.26079

Patterson Ross, Z., Klunk, J., Fornaciari, G., Giuffra, V., Duchene, S., Duggan, A. T., et al. (2018). The paradox of HBV evolution as revealed from a 16th century mummy. *PLoS Pathog.* 14:e1006750. doi: 10.1371/journal.ppat.1006750

Pickrell, J. K., Patterson, N., Loh, P. R., Lipson, M., Berger, B., Stoneking, M., et al. (2014). Ancient West Eurasian ancestry in Southern and Eastern Africa. *Proc. Natl. Acad. Sci. U. S. A.* 111, 2632–2637. doi: 10.1073/pnas.1313787111

Preus, H. R., Marvik, O. J., Selvig, K. A., and Bennike, P. (2011). Ancient bacterial DNA (aDNA) in dental calculus from archaeological human remains. *J. Archaeol. Sci.* 38, 1827–1831. doi: 10.1016/j.jas.2011.03.020

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., et al. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 464, 59–65. doi: 10.1038/nature08821

Rasche, A., Souza, B., and Drexler, J. F. (2016). Bat hepadnaviruses and the origins of primate hepatitis B viruses. *Curr. Opin. Virol.* 16, 86–94. doi: 10.1016/j.coviro.2016.01.015

Rasmussen, M., Guo, X., Wang, Y., Lohmueller, K. E., Rasmussen, S., Albrechtsen, A., et al. (2011). An aboriginal Australian genome reveals separate human dispersals into Asia. *Science* 334, 94–98. doi: 10.1126/science.1211177

Reich, D. (2018). *Who We Are and How We Got Here: Ancient DNA and the New Science of the Human Past*.

Reich, D., Green, R. E., Kircher, M., Krause, J., Patterson, N., Durand, E. Y., et al. (2010). Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* 468, 1053–1060. doi: 10.1038/nature09710

Reich, D., Patterson, N., Kircher, M., Delfin, F., Nandineni, M. R., Pugach, I., et al. (2011). Denisova admixture and the first modern human dispersals into Southeast Asia and Oceania. *Am. J. Hum. Genet.* 89, 516–528. doi: 10.1016/j.ajhg.2011.09.005

Revill, P. A., Tu, T., Netter, H. J., Yuen, L. K. W., Locarnini, S. A., and Littlejohn, M. (2020). The evolution and clinical impact of hepatitis B virus genome diversity. *Nat. Rev. Gastroenterol. Hepatol.* 17, 618–634. doi: 10.1038/s41575-020-0296-6

Robertson, B. H., and Margolis, H. S. (2002). Primate hepatitis B viruses - genetic diversity, geography and evolution. *Rev. Med. Virol.* 12, 133–141. doi: 10.1002/rmv.348

Rodriguez-Frias, F., Buti, M., Jardi, R., Cotrina, M., Viladomiu, L., Esteban, R., et al. (1995). Hepatitis B virus infection: precore mutants and its relation to viral genotypes and core mutations. *Hepatology* 22, 1641–1647. doi: 10.1002/hep.1840220605

Sall, A. A., Starkman, S., Reynes, J. M., Lay, S., Nhim, T., Hunt, M., et al. (2005). Frequent infection of *Hylobates pileatus* (pileated gibbon) with species-associated variants of hepatitis B virus in Cambodia. *J. Gen. Virol.* 86, 333–337. doi: 10.1099/vir.0.80274-0

Seiffert, E. R., Tejedor, M. F., Fleagle, J. G., Novo, N. M., Cornejo, F. M., Bond, M., et al. (2020). A parapithecid stem anthropoid of African origin in the Paleogene of South America. *Science* 368, 194–197. doi: 10.1126/science.aba1135

Simmonds, P. (2001). The origin and evolution of hepatitis viruses in humans. *J. Gen. Virol.* 82, 693–712. doi: 10.1099/0022-1317-82-4-693

Simmonds, P., Aiewsakun, P., and Katzourakis, A. (2019). Prisoners of war–host adaptation and its constraints on virus evolution. *Nat. Rev. Microbiol.* 17, 321–328. doi: 10.1038/s41579-018-0120-2

Simmonds, P., and Midgley, S. (2005). Recombination in the genesis and evolution of hepatitis B virus genotypes. *J. Virol.* 79, 15467–15476. doi: 10.1128/JVI.79.24.15467-15476.2005

Simmonds, P., and Smith, D. B. (1999). Structural constraints on RNA virus evolution. *J. Virol.* 73, 5787–5794. doi: 10.1128/JVI.73.7.5787-5794.1999

Stanaway, J. D., Flaxman, A. D., Naghavi, M., Fitzmaurice, C., Vos, T., Abubakar, I., et al. (2016). The global burden of viral hepatitis from 1990 to 2013: findings from the global burden of disease study 2013. *Lancet* 388, 1081–1088. doi: 10.1016/S0140-6736(16)30579-7

Starkman, S. E., MacDonald, D. M., Lewis, J. C., Holmes, E. C., and Simmonds, P. (2003). Geographic and species association of hepatitis B virus genotypes in non-human primates. *Virology* 314, 381–393. doi: 10.1016/S0042-6822(03)00430-6

Stuyver, L., De Gendt, Van Geyt, Zoulim, F., Fried, M., Schinazi, R. F., et al. (2000). A new genotype of hepatitis B virus: complete genome and phylogenetic relatedness. *J. Gen. Virol.* 81, 67–74. doi: 10.1099/0022-1317-81-1-67

Sugauchi, F., Kumada, H., Sakugawa, H., Komatsu, M., Niitsuma, H., Watanabe, H., et al. (2004). Two subtypes of genotype B (Ba and Bj) of hepatitis B virus in Japan. *Clin. Infect. Dis.* 38, 1222–1228. doi: 10.1086/382885

Sugauchi, F., Orito, E., Ichida, T., Kato, H., Sakugawa, H., Kakumu, S., et al. (2002). Hepatitis B virus of genotype B with or without recombination with genotype C over the precore region plus the core gene. *J. Virol.* 76, 5985–5992. doi: 10.1128/JVI.76.12.5985-5992.2002

Summers, J., and Mason, W. S. (1982). Replication of the genome of a hepatitis B--like virus by reverse transcription of an RNA intermediate. *Cell* 29, 403–415. doi: 10.1016/0092-8674(82)90157-X

Takahashi, K., Brotman, B., Usuda, S., Mishiro, S., and Prince, A. M. (2000). Full-genome sequence analyses of hepatitis B virus (HBV) strains recovered from chimpanzees infected in the wild: implications for an origin of HBV. *Virology* 267, 58–64. doi: 10.1006/viro.1999.0102

Tamm, E., Kivisild, T., Reidla, M., Metspalu, M., Smith, D. G., Mulligan, C. J., et al. (2007). Beringian standstill and spread of native American founders. *PLoS One* 2:e829. doi: 10.1371/journal.pone.0000829

Tatematsu, K., Tanaka, Y., Kurbanov, F., Sugauchi, F., Mano, S., Maeshiro, T., et al. (2009). A genetic variant of hepatitis B virus divergent from known human and ape genotypes isolated from a Japanese patient and provisionally assigned to new genotype J. *J. Virol.* 83, 10538–10547. doi: 10.1128/JVI.00462-09

Tedder, R. S., Bissett, S. L., Myers, R., and Ijaz, S. (2013). The "Red Queen" dilemma-running to stay in the same place: reflections on the evolutionary vector of HBV in humans. *Antivir. Ther.* 18, 489–496. doi: 10.3851/IMP2655

Tiollais, P., Pourcel, C., and Dejean, A. (1985). The hepatitis B virus. *Nature* 317, 489–495. doi: 10.1038/317489a0

Tran, T. T., Trinh, T. N., and Abe, K. (2008). New complex recombinant genotype of hepatitis B virus identified in Vietnam. *J. Virol.* 82, 5657–5663. doi: 10.1128/JVI.02556-07

Vartanian, J. P., Pineau, P., Henry, M., Hamilton, W. D., Muller, M. N., Wrangham, R. W., et al. (2002). Identification of a hepatitis B virus genome in wild chimpanzees (*Pan troglodytes schweinfurthi*) from East Africa indicates a wide geographical dispersion among equatorial African primates. *J. Virol.* 76, 11155–11158. doi: 10.1128/JVI.76.21.11155-11158.2002

Wallace, A. R. (1876). *The Geographical Distribution of Animals*. New York: Harper and brothers.

Wang, Z., Liu, Z., Zeng, G., Wen, S., Qi, Y., Ma, S., et al. (2005). A new intertype recombinant between genotypes C and D of hepatitis B virus identified in China. *J. Gen. Virol.* 86, 985–990. doi: 10.1099/vir.0.80771-0

Warinner, C., Rodrigues, J. F., Vyas, R., Trachsel, C., Shved, N., Grossmann, J., et al. (2014). Pathogens and host immunity in the ancient human oral cavity. *Nat. Genet.* 46, 336–344. doi: 10.1038/ng.2906

Warren, K. S., Heeney, J. L., Swan, R. A., Heriyanto, and Verschoor, E. J. (1999). A new group of hepadnaviruses naturally infecting orangutans (*Pongo pygmaeus*). *J. Virol.* 73, 7860–7865. doi: 10.1128/JVI.73.9.7860-7865.1999

World Health Organisation (2017). Global Hepatitis Report, 2017.

Yang, W., and Summers, J. (1995). Illegitimate replication of linear hepadnavirus DNA through nonhomologous recombination. *J. Virol.* 69, 4029–4036. doi: 10.1128/JVI.69.7.4029-4036.1995

Yuen, L. K. W., Littlejohn, M., Duchene, S., Edwards, R., Bukulatjpi, S., Binks, P., et al. (2019). Tracing ancient human migrations into Sahul using hepatitis B virus genomes. *Mol. Biol. Evol.* 36, 942–954. doi: 10.1093/molbev/msz021

Yuen, M. F., Wong, D. K., Fung, J., Ip, P., But, D., Hung, I., et al. (2008). HBsAg Seroclearance in chronic hepatitis B in Asian patients: replicative level and risk of hepatocellular carcinoma. *Gastroenterology* 135, 1192–1199. doi: 10.1053/j.gastro.2008.07.008

Zaim, Y., Ciochon, R. L., Polanski, J. M., Grine, F. E., Bettis, E. A. 3rd, Rizal, Y., et al. (2011). New 1.5 million-year-old Homo erectus maxilla from Sangiran (Central Java, Indonesia). *J. Hum. Evol.* 61, 363–376. doi: 10.1016/j.jhevol.2011.04.009

Zehender, G., Svicher, V., Gabanelli, E., Ebranati, E., Veo, C., Lo Presti, A., et al. (2015). Reliable timescale inference of HBV genotype A origin and phylodynamics. *Infect. Genet. Evol.* 32, 361–369. doi: 10.1016/j.meegid.2015.03.009

Zeng, G., Wang, Z., Wen, S., Jiang, J., Wang, L., Cheng, J., et al. (2005). Geographic distribution, virologic and clinical characteristics of hepatitis B virus genotypes in China. *J. Viral Hepat.* 12, 609–617. doi: 10.1111/j.1365-2893.2005.00657.x

Zhu, Z., Dennell, R., Huang, W., Wu, Y., Qiu, S., Yang, S., et al. (2018). Hominin occupation of the Chinesel plateau since about 2.1 million years ago. *Nature* 559, 608–612. doi: 10.1038/s41586-018-0299-4

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

REVIEW
published: 21 June 2021
doi: 10.3389/fmicb.2021.652962

Hepatitis Delta Virus (HDV) and Delta-Like Agents: Insights Into Their Origin

Hans J. Netter[1,2], Marilou H. Barrios[1,3], Margaret Littlejohn[1] and Lilly K. W. Yuen[1]*

[1] *Victorian Infectious Diseases Reference Laboratory (VIDRL), Melbourne Health, The Peter Doherty Institute, Melbourne, VIC, Australia,* [2] *School of Science, Royal Melbourne Institute of Technology (RMIT) University, Melbourne, VIC, Australia,* [3] *The Peter Doherty Institute, University of Melbourne, Melbourne, VIC, Australia*

Hepatitis delta virus (HDV) is a human pathogen, and the only known species in the genus *Deltavirus*. HDV is a satellite virus and depends on the hepatitis B virus (HBV) for packaging, release, and transmission. Extracellular HDV virions contain the genomic HDV RNA, a single-stranded negative-sense and covalently closed circular RNA molecule, which is associated with the HDV-encoded delta antigen forming a ribonucleoprotein complex, and enveloped by the HBV surface antigens. Replication occurs in the nucleus and is mediated by host enzymes and assisted by *cis*-acting ribozymes allowing the formation of monomer length molecules which are ligated by host ligases to form unbranched rod-like circles. Recently, meta-transcriptomic studies investigating various vertebrate and invertebrate samples identified RNA species with similarities to HDV RNA. The delta-like agents may be representatives of novel subviral agents or satellite viruses which share with HDV, the self-complementarity of the circular RNA genome, the ability to encode a protein, and the presence of ribozyme sequences. The widespread distribution of delta-like agents across different taxa with considerable phylogenetic distances may be instrumental in comprehending their evolutionary history by elucidating the transition from transcriptome to cellular circular RNAs to infectious subviral agents.

Keywords: hepatitis delta virus, delta-like agent, satellite virus, hepatitis B virus, helper virus, subviral agent

OPEN ACCESS

Edited by:
Julie Lucifora,
Institut National de la Santé et de la Recherche Médicale (INSERM), France

Reviewed by:
Paul Dény,
Institut National de la Santé et de la Recherche Médicale (INSERM), France
Lucie Etienne,
UMR 5308 Centre International de Recherche en Infectiologie (CIRI), France

****Correspondence:***
Hans J. Netter
hans.netter@mh.org.au

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *13 January 2021*
Accepted: *12 May 2021*
Published: *21 June 2021*

Citation:
Netter HJ, Barrios MH, Littlejohn M and Yuen LKW (2021) Hepatitis Delta Virus (HDV) and Delta-Like Agents: Insights Into Their Origin. Front. Microbiol. 12:652962. doi: 10.3389/fmicb.2021.652962

LIFE CYCLE OF HDV AND ITS DEPENDENCE ON HBV AS HELPER VIRUS

Hepatitis delta virus (HDV) is a unique human pathogen, and has been the only known species in the genus *Deltavirus* (Magnius et al., 2018), but was reclassified in a new family *Kolmioviridae*, genus *Deltavirus* within one new realm *Ribozyviria* [International Committee on Taxonomy of Viruses (ICTV), 2020]. Due to the possession of a circular RNA genome and its mechanism of replication, similarities exist with viroids, which represent a large family of subviral plant pathogens (Flores et al., 2016; Adkar-Purushothama and Perreault, 2019). But HDV is clearly distinguished from the viroids by its larger genome size and the ability to encode a protein. The recent discovery of delta-like agents in various animal species has broadened the views on the evolutionary history of HDV (Wille et al., 2018; Chang et al., 2019; Hetzel et al., 2019; Paraskevopoulou et al., 2020; Bergner et al., 2021; Iwamoto et al., 2021).

The existence of HDV was discovered in 1977 by the identification of a new antigen, the delta antigen (HDAg), in liver biopsies and sera from patients with a severe form of hepatitis B

(Rizzetto et al., 1977). Experimental transmission studies (Rizzetto et al., 1980; Ponzetto et al., 1984), then the cloning of the HDV genome (Denniston et al., 1986; Wang et al., 1986; Makino et al., 1987) demonstrated that the HDAg is associated with a separate transmissible agent. HDV is a satellite virus and depends on the human hepatitis B virus (HBV) surface proteins (HBsAg) for packaging, release, and transmission.

Extracellular HDV virions exclusively contain genomic HDV RNA, a single-stranded negative-sense and covalently closed circular RNA molecule with a size of 1,668–1,697 nucleotides, depending on the genotype (Le Gal et al., 2017) in contrast to viroids with a size range between approximately 250–400 nucleotides (Flores et al., 2016). HDV does not encode for an RNA-dependent-RNA polymerase (RdRp), but depends on host DNA-dependent RNA polymerases (DdRp) to facilitate RNA-directed RNA synthesis for transcribing and replicating the genome in the cell nucleus (Modahl et al., 2000; Chang et al., 2008; Sureau and Negro, 2016). The negative-sense, genomic RNA is replicated by a rolling circle mechanism to generate its complement, the positive-sense, antigenomic RNA, which serves as a replication intermediate for the synthesis of the genomic RNA (**Figure 1A**). Both *de novo* synthesized genomic and antigenomic RNA molecules are processed into monomers by intrinsic ribozymes, which are formed by nested double pseudoknot structures (Kuo et al., 1988; Perrotta and Been, 1991; Ferré-D'Amaré et al., 1998; Sureau and Negro, 2016). Monomer synthesis is followed by the formation of circular RNA molecules by intramolecular ligation (Sharmeen et al., 1989; Reid and Lazinski, 2000).

The HDV genomic RNA and its antigenomic complement are highly self-complementary and form unbranched rod-like structures in native conditions with approximately 74% intramolecular base pairing (Kos et al., 1986; Wang et al., 1986). HDV shares the features of a circular self-complementary RNA genome with viroids (Flores et al., 2016; Adkar-Purushothama and Perreault, 2019). Contrary to viroids, HDV generates a mRNA transcript which encodes the HDAg (Hsieh et al., 1990). HDV expresses two HDAg forms, which interact with HDV RNA to form ribonucleoprotein (RNP) complexes (Ryu et al., 1993) but they play different roles during the HDV life cycle. The small form of the HDAg (HDAg-S) is expressed throughout the HDV infection, it is essential for HDV replication and HDV RNA accumulation but does not exhibit a RdRp activity and stimulates DdRp II elongation (Kuo et al., 1989; Hsieh et al., 1990; Yamaguchi et al., 2001). During infection and HDV replication, a site-specific editing event occurs at some antigenomic RNAs, which allows the generation of mRNAs with the UAG stop codon replaced with the tryptophan UGG codon (**Figure 1B**). The extended HDAg open reading frame by 19 amino acids encodes the large form of HDAg (HDAg-L) (Luo et al., 1990; Polson et al., 1996; Wong and Lazinski, 2002). Following the editing event, HDAg-L accumulates during HDV replication and facilitates HDV assembly in the presence of the HBV HBsAg proteins (Chang et al., 1991). The additional 19 amino acids contain the carboxyterminal "CXXX" sequence motif allowing HDAg-L farnesylation, which is essential for the assembly process (Glenn et al., 1992; Hwang and Lai, 1993; O'Malley and Lazinski, 2005). The farnesyl group facilitates RNP attachment to the membrane of the endoplasmic reticulum (ER), the site of HBsAg synthesis. The proximity of the hepatitis delta RNPs and HBV surface proteins allows the envelopment of the hepatitis delta RNPs at the ER membrane to generate hepatitis delta virions. HDAg-L in the absence of HDV RNA and HDAg-S can be packaged by HBsAg proteins indicating that HDAg-L is the driver of HDV assembly and release (Chang et al., 1991; Chen et al., 1992; Ryu et al., 1992).

HDV depends on the HBV envelope proteins for its life cycle completion and the production of infectious HDV particles. HBV encodes three related surface proteins, the shortest surface protein HBsAg-S is composed only of the S-domain, the middle and large HBsAg proteins (HBsAg-M and HBsAg-L) contain additional N-terminal extensions of the S-domain. In particular, the dual topology of the preS1 domain of HBsAg-L is essential for hepatitis B virion assembly and attachment to the host cell. The preS1 domain of newly synthesized HBsAg-L proteins are on the cytosolic side and interact with the hepatitis B nucleocapsid allowing virion assembly. During the HBV budding process, the HBsAg-L preS1 domain translocates across the viral lipid layer to be surface exposed (Bruss et al., 1994; Ostapchuk et al., 1994), which is essential for the preS1 binding site to be able to engage the viral receptor, "sodium taurocholate cotransporting polypeptide" (NTCP) (Yan et al., 2012). As for HBV, HDV requires the preS1 domain of HBsAg-L for binding to the NTCP receptor (Yan et al., 2012). But in contrast to the HBV morphogenesis, the preS1 domain is not required for the HDV assembly and budding process. The presence of only HBsAg-S proteins is sufficient for HDV assembly and release of HDV particles, but in the absence of the HBsAg-L preS1 domain, they are non-infectious (Sureau et al., 1993). Several HBV envelope S-domains are involved in HDV assembly and secretion including the internal, cytosolic loop and a tryptophan-rich C-terminal sequence (Jenna and Sureau, 1999; Komla-Soukha and Sureau, 2006). Compared to HBV with a diameter of approximately 42 nm, HDV is less dense and slightly smaller in size, approximately 39 nm in diameter (He et al., 1989; Hu and Liu, 2017).

The dynamics of HDV infections following orthotopic liver transplantations provided evidence that HDV can cause subclinical helper-independent or mono-infections (Ottobrelli et al., 1991; Samuel et al., 1995) but HDV viremia requires the HBV helper function (Smedile et al., 1998). The detection of HDAg in the liver in the absence of HBV markers is possibly an indicator of HDV latency in the liver (Mederacke et al., 2012). Consistently, animal models demonstrated that HDV replicates and also persists in helper independent- or mono-infection contexts. The absence of HBV HBsAg or envelope proteins from a closely related mammalian hepatitis B virus, woodchuck hepatitis B virus (WHV), allowed HDV replication but no progression to viral assembly and release (Netter et al., 1993; 1994; Giersch et al., 2014). Studies with the woodchuck animal model and humanized mice showed that HDV during a mono-infection phase could be rescued by inoculating the corresponding WHV or HBV helper virus resulting in HDV viremia (Netter et al., 1994; Giersch et al., 2014). Alternatively, integrated HBV DNA can provide functional HBsAg-L and HBsAg-S transcripts and proteins to

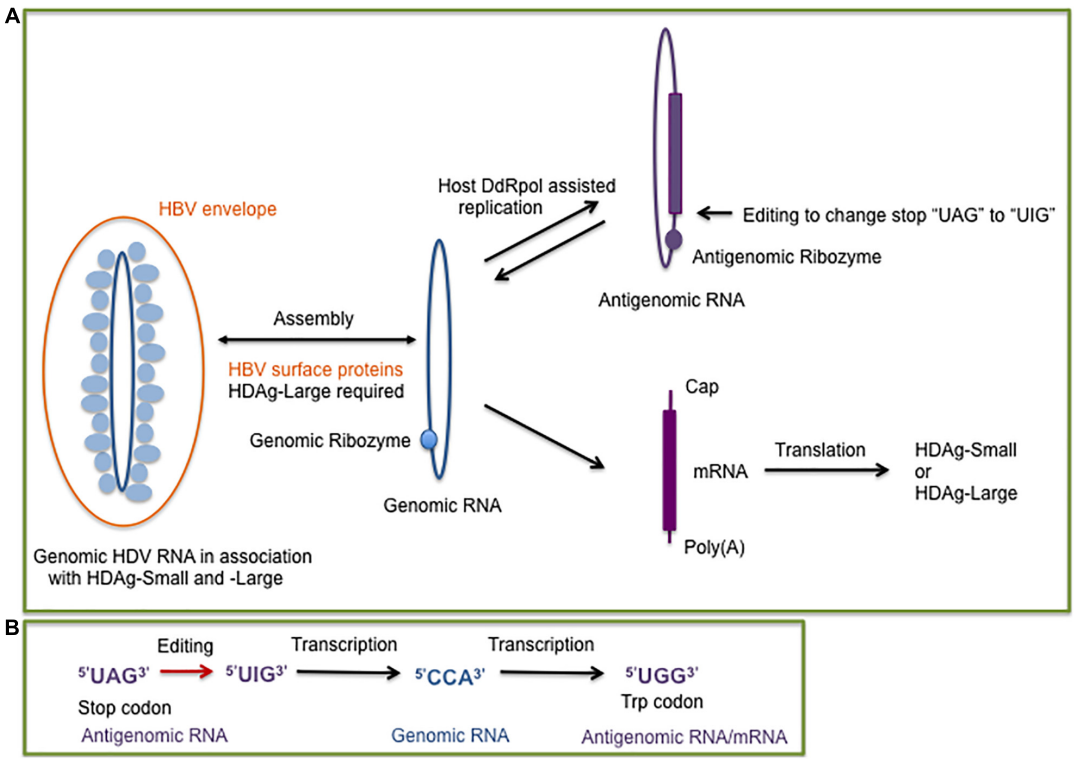

FIGURE 1 | (A) Overview of the hepatitis delta virus (HDV) replication cycle supported by the host DNA dependent RNA polymerase (DdRp) to replicate the genomic (blue) and antigenomic (purple) circular single-stranded RNA molecules. The presence of the genomic and antigenomic ribozyme is indicated by filled circles. The genomic RNA serves as a template for the synthesis of the mRNA which encodes the small or large delta antigen (HDAg-Small, HDAg-Large). An editing event at the antigenomic RNA results in the formation of genomic RNAs for the synthesis of mRNA molecules encoding HDAg-Small or HDAg-Large. The dependence of HDV packaging on the hepatitis B virus (HBV) envelope (surface) proteins is indicated. **(B)** Steps of the editing outcome which converts the stop codon of the HDAg-Small open reading frame to a tryptophan codon. The editing of the antigenomic RNA converts the adenosine of the "UAG" stop codon to an inosine, which prefers to pair with cytosine during replication. The edited genomic RNA transcribes into a mRNA with the "UGG" tryptophan codon, which allows a 19 amino acid extension of the open reading frame for HDAg-Large synthesis (Polson et al., 1996).

facilitate formation of infectious HDV in the absence of HBV replication (Freitas et al., 2014).

A cell culture based study showed that HDV can be pseudotyped with envelope proteins derived from various viruses, including vesicular stomatitis virus (VSV), hepatitis C virus (HCV), and Dengue virus. The VSV and HCV envelope proteins supported the release and assembly of genomic HDV RNA, also depending on HDAg-L farnesyl-mediated targeting of cell membranes similar to the assembly process supported by HBV envelope proteins (Perez-Vargas et al., 2019). The HDV RNPs pseudotyped with HCV envelope proteins and Dengue glycoproteins generated infectious HDV particles supporting entry and replication in human hepatoma cells HuH7.5 and insect C6/36 cell lines, respectively. However, the clinical relevance of these findings remain uncertain. A recent study that involved 323 HCV RNA positive and HBsAg-negative patients could only detect HDV markers in eight HBV core antibody (anti-core) positive patients (evidence of past acute HBV infections) and not among the remaining HBV core antibody negative patients suggesting the occurrence of replicative HDV infections in HCV mono-infected patients is low (Pflüger et al., 2020). A similar study investigating a cohort of 160 Venezuelan patients infected with HCV in the absence of molecular markers for HBV detected two patients with anti-HDAg antibodies, and for one patient low-level circulating HDV RNA (Chemin et al., 2020), also indicating that if HCV provides helper functions, it does not seem to be an effective or potent helper virus. Furthermore, HDV RNA and HDAg have been detected in the salivary glands of patients with a primary Sjögren's syndrome in the absence of a past or current HBV infection, which leaves the unanswered question of how HDV established an infection in these patients (Weller et al., 2016). Interestingly, Perez-Vargas et al. (2019) confirmed earlier studies that HDV replication is not restricted to human liver cells. HDV replication in the absence of HBV helper function for assembly and release has been reported for human embryonic kidney cells, mouse skeletal muscle cells, and hamster kidney cells (Bichko et al., 1994; Polo et al., 1995; Chang et al., 2005) but HDV replication is restricted in avian cells due to the cytotoxicity of the delta antigen (Chang et al., 2000). The ability of HDV to replicate in different cell types, the recent identificaton of HDV-like agents in diverse vertebrate and invertebrate species (Chang et al., 2019) and the presence of cellular ribozymes with structural similarities to the

HDV-ribozyme (Riccitelli and Lupták, 2013) possibly suggests that HDV originates from the cell transcriptome, due to a process related to the biogenesis of cellular circular RNAs found in eukaryotes (Kristensen et al., 2019).

HDV Genotypes and Delta-Like Agents

Natural HDV infections have only been described in humans, and hence HDV most likely co-evolved with the helper HBV in the human lineage. Experimental transmission of HDV and HBV to chimpanzees (Rizzetto et al., 1980), and the acceptance of mammalian hepatitis B viruses (genus *Orthohepadnavirus*), such as the woodchuck hepatitis B virus (WHV) as alternative helper viruses for HDV assembly and transmission allowed the establishment of animal models to study HDV replication and pathogenesis (Ponzetto et al., 1984; Netter et al., 1994; Gerin, 2001; Aldabe et al., 2015).

Eight distinct HDV genotypes have been documented in human populations. The HDV genotypes differ in their genomic sequence by 19–40% (Le Gal et al., 2017), and can be further sub-categorized into two to four subgenotypes with the exception of HDV genotype 3 (HDV-3). Their global distribution is geographically distinct except for HDV-1d (**Figure 2**; Le Gal et al., 2017). The subgenotype HDV-1d is prevalent worldwide, and represents the dominant HDV strain in Europe and North America. In contrast, HDV-1a and -1b are predominantely found in Africa and the Middle East, while HDV-1c is the dominant strain in the Western Pacific region (Han et al., 2014). The subgenotypes HDV-2a, HDV-4a, and HDV-4b are mainly distributed in Southeast Asia, China, Japan, and Taiwan; HDV-2b in Russia (Siberia). HDV-3 is mainly located in the Northern part of South America, and HDV-5 to HDV-8 are found in Africa (Le Gal et al., 2017).

Based on full-length genome and deduced HDAg amino acid sequences, HDV-3 seems to be the most distantly related human HDV genotype and has the lowest similarity score when compared to the other HDV genotyes (Casey et al., 1993). The distant genetic relationship of HDV-3 to the other genotypes is further indicated by the inability of the HDAg encoded by HDV-3 to support replication of HDV-1, and vice versa (Casey and Gerin, 1998). Depending on the genotype, HDV can cause disease manifestations of different severity, HDV-2 is normally associated with a milder disease progession than HDV-1. Infections with HDV-3 in combination with HBV genotype F, which is the predominant HBV in the northern part of South America are associated with an enhanced risk of fulminant hepatitis (Casey et al., 1996). With the presence of HDV genotypes 5–8, and sub-genotypes 1a and 1b, central Africa around Cameroon is possibly the main site of HDV diversication resulting in an ancient radiation of the African lineages (Radjef et al., 2004; Le Gal et al., 2017).

Advances in metagenomics have lead to the discovery of delta-like agents from the transcriptome libraries generated for a number of non-human vertebrates and invertebrates in recent years (Wille et al., 2018; Chang et al., 2019; Hetzel et al., 2019; Paraskevopoulou et al., 2020; Bergner et al., 2021; Iwamoto et al., 2021). Of the large number of libraries screened in these studies to date, delta-like agents have been identified in at least nine taxa including birds (gray teal, chestnut teal, Pacific black ducks, zebra finch, Gouldian finch, canary, yellow-bellied tit), termite (subterranean termite), fish (pooled sample from multiple species), toad (Asiatic toad), newt (Chinese fire belly newt), snake (boa constrictor, water python), rat (Tome's spiny rat), woodchuck (Eastern woodchuck), bat (common vampire bats, lesser dog-like bat), and deer (white-tailed deer). Surprisingly, delta-like agents were not detected in non-human primates indicating that HDV is the only known representative infecting the order Primates (Bergner et al., 2021). The natural habitats of the taxa mentioned above in relation to the geographical location of the human HDV genotype and subgenotypes are shown in **Figure 2**. The ability for these non-human delta-like agents to transmit and replicate in nature remain unclear, but it has been shown that the rodent delta-like agent is able to replicate *in vitro* (Paraskevopoulou et al., 2020) and that the bat delta-like agent can transmit to other members of the same colony based on prevalence studies of bat colonies (Bergner et al., 2021). As analyzed for delta-like agents identified in the Eastern woodchuck (*Marmota monax*), canary (*Serinus canaria*), Zebra finch (*Taeniopygia guttata*) and white-tailed deer (*Odocoileus virginianus*), the read depths of predicted transcribed regions (the coding region for the delta-like antigen) were greater than those of other genomic regions indicating that most delta reads were derived from delta mRNAs suggesting that the novel delta-like agents replicate in their hosts (Iwamoto et al., 2021). Nonetheless, no definitive helper virus for assembly and transmission of these delta-like agents within their hosts have been identified to date. It has been suggested that an alternative supply of envelope proteins for the non-human delta-like agents may be provided by endogenous viral elements (EVEs), which are encoded within the host genome, or they may utilize strategies distinct from those employed by HDV (Iwamoto et al., 2021). HBV EVEs have been identitifed in the genome of birds of the order *Passeriforme* including zebra finch (Gilbert and Feschotte, 2010) and budgerigars of the order *Psittaciformes* (Cui and Holmes, 2012), however *in vitro* studies confirmed that the zebra finch delta-like agent, also the woodchuck delta-like agent, did not use the small HBsAg supplied *in trans* to generate infectious virions (Iwamoto et al., 2021). Interestingly, the Eastern woodchuck (*Marmota monax*) is the host for the woodchuck hepatitis B virus (WHV) which is an efficient helper virus providing the envelope proteins for the assembly of infectious HDV in the woodchuck animal model (Netter et al., 1994; Gerin, 2001). No natural HDV infection has been reported in this animal population, and there is no evidence that HBV supports packaging and release of infectious woodchuck-derived delta-like agents (Iwamoto et al., 2021). This strongly indicates that the presence of a potential helper virus and a delta-like agent does not necessarily drive the emergence of a satellite RNA virus. This finding supports the proposition that non-human delta-like agents may use a novel mechanism for assembly and transmission.

The phylogenetic relationships between HDV and the delta-like agents have been consistent between studies (Paraskevopoulou et al., 2020; Bergner et al., 2021; Iwamoto et al., 2021), though minor clustering variations do occur between the full-genome and HDAg phylogenetic trees. Based on full-genome

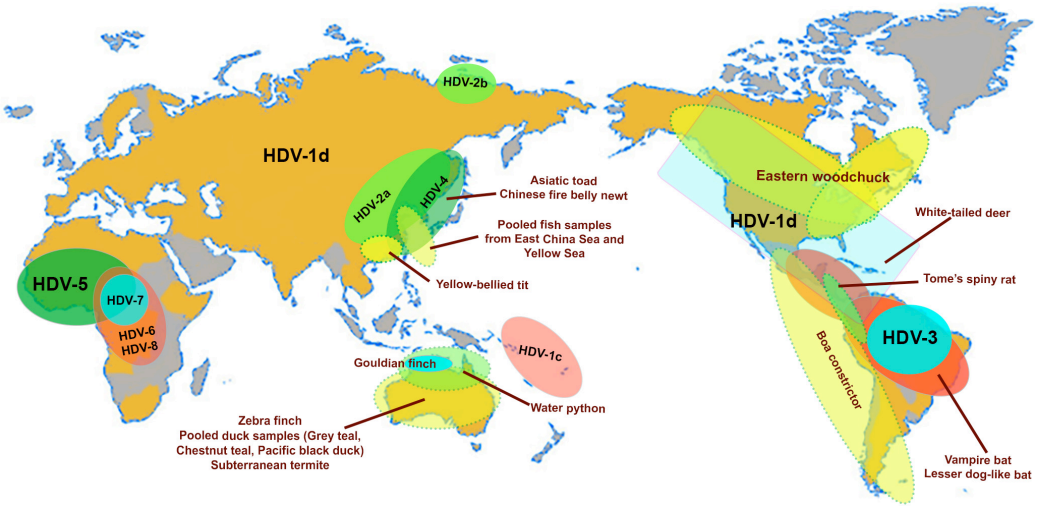

FIGURE 2 | Map highlights the geographical distribution of HDV genotypes (HDV1-8) and subgenotypes (HDV1c, 1d; HDV-2a, b) and the worldwide distribution of animal hosts of identified delta-like agents (Asiatic toad–*Bufo gargarizans*; Newt–*Cynops orientalis*; Yellow-bellied tit–*Pardaliparus venustulus*; Gouldian finch–*Erythrura gouldiae*; water pyton–*Liasis mackloti*; Zebra finch–*Taeniopygia guttata*; subterranean termite–*Schedorhinotermes intermedius*, lesser dog-like bat–*Peropteryx macrotis*; vampire bat–*Desmodus rotundus*; Tome's spiny rat–*Proechimys semispinosus*; white-tailed deer–*Odocoileus virginianus*; Eastern woodchuck–*Marmota monax* (Paraskevopoulou et al., 2020; Bergner et al., 2021; Iwamoto et al., 2021). The pooled fish samples contain species from the classes Actinopterygii, Chondrichthyes and Agnatha (Chang et al., 2019), and the waterfowl samples were collected from gray teal (*Anas gracilis*), chestnut teal (*Anas castanea*), and pacific black duck (*Anas superciliosa*) (Wille et al., 2018). The map is adapted from Le Gal et al. (2017).

phylogenetic trees, HDV and the non-human delta-like agents formed two distinct clusters (Paraskevopoulou et al., 2020). In this study, the delta-like agent genomes further clustered into two distinct groups (the snake and rodent delta-like agents vs. the newt, toad, fish, termite, and duck delta-like agents). Interestingly, the HDAg amino acid sequences of the newly discovered deer (*Odocoileus virginianus*), woodchuck (*Marmota monax*), and one of the bat (*Desmodus rotundus*) genotypes of delta-like agents were found to form a sister clade with the human HDV (Bergner et al., 2021). The HDAg of these delta-like agents shared a common ancestor with the HDAg of a second bat (*Desmodus rotundus*) genotype, rodent (*Proechimys semispinosus*), snake (*Boa constrictor*) and duck delta-like agents, while the remaining delta-like agents formed a more distal clade (derived from fish, newt and toad). However, despite the low level of identity between the HDAg protein sequences of delta-like agents from different taxa (14–67%) (Iwamoto et al., 2021), no clear distinction between the vertebrate and invertebrate delta-like agents could be detected by phylogenetic analyses. In addition to the finding that the bat and rodent delta-like agents were paraphyletic, cophylogenetic analyses between delta-like agent and host trees did not support the theory that delta-like agents had co-speciated with their hosts but may have evolved by host shifting (Bergner et al., 2021; Iwamoto et al., 2021).

Hypotheses of Origin

(i) Delta-Like Agents and Cellular Circular RNAs as Potential Precurors of Subviral Agents.

Viroids were the first circular RNA discovered (Sänger et al., 1976), then followed by the discovery of cellular circular RNA located in the cytoplasm of eukaryotic cells (Hsu and Coca-Prados, 1979). Circular RNAs (circRNAs) can be generated by alternative splicing or backsplicing and results in the formation of different circRNA species composed of only exon sequences (exonic circRNAs), circRNAs with both intron and exon sequences (exon-intron circRNAs), and circRNAs with only intronic sequences (ciRNAs). Alternatively, the formation of circRNA can be assisted by ribozymes, such as the HDV-related ribozymes, which have a widespread occurence in eukaryotic genomes (**Figure 3**). CircRNAs with hammerhead ribozymes have been identified in eukaryotic genomes, and it has been hypothesized that they have given origin to infectious circRNAs (De la Pena and Cervera, 2017). The covalently closed ring structure confers stability avoiding exonuclease-mediated degradation (Kristensen et al., 2019; Lasda and Parker, 2014). An abundance of circRNAs have been identified in divergent animal and plant species. For human fibroblast cells it was reported that circRNA molecules originate from more than 14% of transcribed genes (Jeck et al., 2013). The biogenesis of circular cellular RNAs is antagonized by the RNA adenosine deaminase (ADAR) by editing endogenous double-stranded (ds) RNA sequences. ADAR faciliates in combination with an ATP-dependent RNA helicase A (DHX9) the melting of dsRNA stems and hence prevents the looping of intron sequences and the formation of circRNA (Ivanov et al., 2015). Interestingly, the circular HDV RNA depends on ADAR editing for the completion of the viral reproductive cycle. HDV antigenomic RNA requires editing to generate mRNAs with an extended open reading for the synthesis of HDAg-L, which is essential for the HDV envelopment by the HBV surface proteins. ADAR targets the partially double-stranded editing site of the circular HDV antigenomic RNA

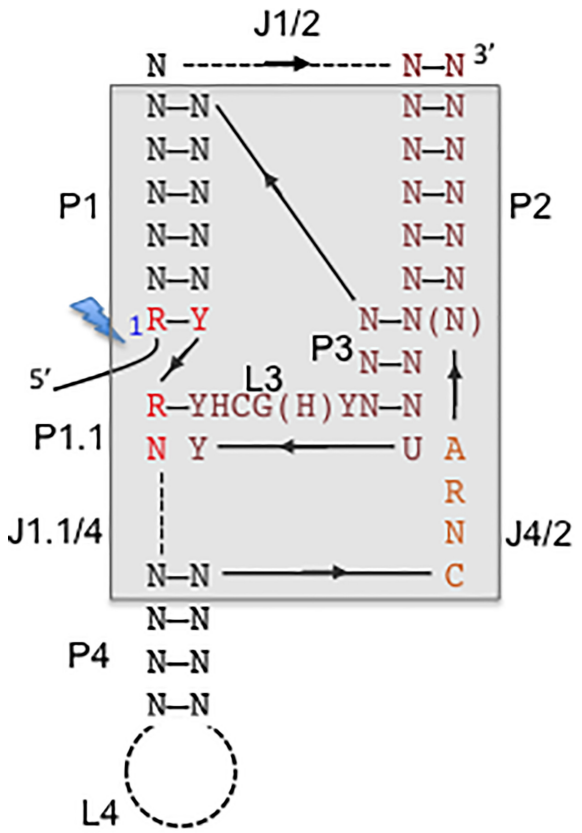

FIGURE 3 | Consensus secondary structure of hepatitis delta virus (HDV)-like ribozymes. The catalytic core is boxed with gray background. The base-paired regions are numbered with P1–P4, which are linked by single-stranded regions, J1/2, J4/2, and J1.1/4. L stands for single-stranded loop regions. Depending on the ribozyme, J1/2 region and P4 helix can be expanded indicated by the dashed lines allowing sequence length variations. The solid line represents direct connections without nucleotide insertions. The arrows indicate the direction of the strands from 5′ to 3′. The 5′ end highlighted by a lightning bolt marks the 5′-OH end after self-cleavage. N stands for any nucleotide, R for purine, Y for pyrimidine, H for adenine, cytidine or uracyl. The figure is adapted from Ferré-D'Amaré et al. (1998) and Webb and Lupták (2011).

(Polson et al., 1996) to ultimately convert the amber stop codon to a tryptophan codon (**Figure 1B**). HDV editing by ADAR which targets partially double-stranded RNA molecules and RNA loop structures points together with the widespread distribution of the HDV-like ribozyme and the prevalence of circRNA molecules in eukaryotes toward a long evolutionary history of HDV.

The recent discovery of delta-like agents in metagenomic samples from birds and snake, then in fish, amphibians and invertebrates demonstrates that the RNA genomes are highly divergent (Wille et al., 2018; Chang et al., 2019; Hetzel et al., 2019; Bergner et al., 2021; Iwamoto et al., 2021). The delta-like agents share common features with HDV, the circular genome with a size of approximately 1.7 kb (between 1,547 and 1,735 nucleotides), self-complementarity to fold into rod-like structures, and the presence of an open reading frame (ORF) (Chang et al., 2019). Delta-like agents identified in tissues from snakes (Hetzel et al., 2019), a rodent species (*Proechimys semispinosus*) (Paraskevopoulou et al., 2020) and in combined oropharangeal and cloacal samples from teals and ducks (Wille et al., 2018) contain HDV-like ribozymes (**Table 1**). Importantly, the presence of delta-like agents does not seem to be associated with members of the hepatitis B virus family indicating that HDV as a delta-agent possibly co-evolved with HBV in humans and optimized the assembly efficacy and release. The HBV envelope proteins provide hepatocyte-specific tropism (Bonino et al., 1986; Yan et al., 2012) but the HDV genome retained the ability to replicate in non-liver cells (Bichko et al., 1994; Polo et al., 1995; Chang et al., 2005). In contrast, the snake delta-like agent (sDLA) does not exhibit strict assembly requirements as clearly demonstrated for HDV and its dependence on the HBV envelope proteins. The inoculation of boa kidney cells with sDLAs derived from an infected brain homogenate passaged sDLAs in the presence of coinfecting Arenaviruses (Hartmanivirus, Reptarenavirus). Transfection experiments with glycoproteins from Arenaviruses and Orthohantavirus allowed the formation of infectious sDLAs (Szirovicza et al., 2020). The sDLA antigen (sDL-Ag) was identified in different tissues of infected animals, and consistently, sDLA replication similar to HDV replication, is supported by different cell types (Bichko et al., 1994; Polo et al., 1995; Chang et al., 2005; Szirovicza et al., 2020). Rodent DLA (rDLA) was identified in blood samples, not linked to liver tropism, and interestingly, predominantly detected in reproductively active males living in continuous forest sites, suggesting horizontal transmission linked to competitive behavior. The study did not identify a helper virus and transmission is possibly assisted by envelope proteins provided by an unkown agent, or by extracellular vesicles, which include microvesicles and exosomes (Kim et al., 2017; Paraskevopoulou et al., 2020). It was proposed for retroviruses to exploit the exosome exchange for a low efficiency mode of infection (Gould et al., 2003). Hepatitis C virus (HCV)-RNA containing exosomes have been identified for the export of viral RNA to plasmacytoid denritic cells (Dreux et al., 2012), and importantly mediate transmission of HCV between hepatocytes (Ramakrishnaiah et al., 2013). Similarly, HBV virions were detected in extracellular vesicles collected from infected patients (Kakizaki et al., 2018), and HBV RNA in extracellular vesicles from HBV transfected hepatocytes (Kouwaki et al., 2016). CircRNAs are present in exosomes (Kim et al., 2017; Fanale et al., 2018; Veziroglu and Mias, 2020), and hence delta-like agents could be transmitted with low efficiency as proposed for retroviruses according to the Trojan exosome hypothesis (Gould et al., 2003). The exact mechanisms by which RNAs are loaded into exosomes remain unclear but certain motifs and double-stranded stem-loop secondary structure were proposed to be important for packaging (Villarroya-Beltri et al., 2013; Kossinova et al., 2017).

Both the sDLAs and rDLAs studies demonstrated the expression of the corresponding DL-Ags confirming that the ORF is translated (Paraskevopoulou et al., 2020; Szirovicza et al., 2020). Transfection of expression vectors with dimer genomes for delta-like agents derived from woodchuck and Zebra finch resulted in the expression of the corresponding delta-like antigens (Iwamoto et al., 2021). The identified delta-like agents from vertebrates

TABLE 1 | Characteristics of hepatitis delta virus (HDV), delta-like agents, cell-encoded circular RNAs, and plant viroids.

	HDV	Delta-like agents	Cell-encoded circular RNA	Viroids	
				Pospiviroidae	Avsunviroidae
Single-stranded circular RNA	✓	✓	✓	✓	✓
Self-complementarity	✓	✓	Formation of stem/loop structures possible	✓	✓
Unbranched rodlike structure	✓	✓		✓	×
Host DdRpol assisted replication	✓	✓‡	Derivatives from splicing events	✓	✓
Ribozyme	✓	✓	With or without ribozyme (hammerhead)	×	✓
HDV ribozyme/HDV-like ribozyme	✓	✓*			×
ORF, encoding of protein	✓	✓	With or without ORF	×	×

*DdRpol, DNA-dependent RNA polymerase; ORF, open reading frame; ‡delta-like agents do not encode a polymerase, and most likely replication assisted by DdRpol. *Delta-like ribozymes were not identified for delta-like agents derived from the toad (Bufo gargarizans), termite (Schedorhinotermes intermedius), newt (Cynops orientalis) and fish (derived from a pool of fish species, Chang et al., 2019) hosts.*

and invertebrates encode DL-Ags including additional in-frame reading frames downstream of the stop codon or in alternative reading frames demonstrating a high plasticity of the information content of the genomes. The presence of exons in many circular cellular RNAs supports the view that the HDAg and DL-Ags originated from a host organism (Kristensen et al., 2019). A host-derived protein-encoding sequence as proposed for the delta-interacting protein A (DIPA) may have been incorporated into a HDV ancestral genome (Brazas and Ganem, 1996; Long et al., 1997). For both sDLAs and rDLAs proteins, larger versions of the DL-Ags were not detected in contrast to the HDAg-S/-L versions expressed by HDV. The presence of ORFs and also regulatrory elements is not unexpected if the delta-like agents are related to circRNAs (Chen et al., 2016). Remarkably, circRNAs can be translated through different mechanisms, internal ribosome entry site (IRES) dependent or supported by base modification (Chen and Sarnow, 1995; Abe et al., 2015; Wang and Wang, 2015; Pamudurti et al., 2017; Yang et al., 2017).

(ii) HDV and Its Ribozyme

Self-cleaving RNA motifs or ribozymes play important roles for facilitating rolling circle replication of circular RNAs such as HDV RNA, viroids, or satellite RNAs. Self-cleaving ribozymes are classified based on the secondary and tertiary structures of the catalytic RNA motifs, which are unique for each family, such as hairpin, hammerhead, HDV-like and twister (Jimenez et al., 2015). HDV-like ribozymes are widespread and have been identified in retrotransposons and various genomic loci suggesting miscellaneous biological functions, possibly providing an extra level of control for expression of the genes in which they are located (Webb and Lupták, 2011; Riccitelli and Lupták, 2013). The HDV-ribozyme and related ribozymes are modeled into pseudoknotted secondary structures (Perrotta and Been, 1991; Webb and Lupták, 2011), the crystal structure of the related HDV genomic and antigenomic ribozymes revealed the presence of a nested, double pseudoknot structure (Ferré-D'Amaré et al., 1998; Wadkins et al., 1999) (**Figure 3**). A ribozyme with structural similarity to the HDV ribozymes was first identified in an intron of the human cytoplasmic polyadenylation element-binding protein (*CPEB3*) gene and is possibly involved in co-transcriptional processing of *CPEB3* primary RNA transcripts (Salehi-Ashtiani et al., 2006). The CPEB3 ribozyme is highly conserved among mammals, but the CPEB3 ribozyme sequences are substantially different to the HDV genomic and antigenomic ribozyme sequences. Approximately 60 nucleotides are required to form the conserved nested double-pseudoknot structure, only six nucleotides are invariant (Webb et al., 2009). Structure-based searches identified CPEB3 related sequences in non-mammalian genomes demonstrating their wide distribution. HDV-like ribozymes have been identified in all branches of life, and found in the genomes of the *Anopheles gambia*, *Drosophila* species, the insect virus Chilo iridescent virus, sea urchin *Strongylocentrotus purpuratus*, lamprey *Petromyzon marinus*, lancelet *Branchiostoma floridae*, nematodes *Caenorhabditis japonica* and *Pristionchus pacificus*, and outside of the eukaryotics in the bacterium *Faecalibacterium prausnitzii* (Webb et al., 2009; Eickbush and Eickbush, 2010; Ruminski et al., 2010). Interestingly, HDV-like ribozymes have been located at the 5′-end of retrotransposons suggesting that it represents an ancient element, possibly involved in the cotranscriptional processing of retroelements, and is spread by retrotransposition (Webb and Lupták, 2011). For instance, HDV-like ribozymes have been identified at the 5′-untranslated region of the R2 non-long terminal repeat (LTR) retrotransposon in *Drosophila* (Eickbush and Eickbush, 2010; Ruminski et al., 2010), and at the 5′-end of the L1 and NAR retrotransposons identified in *Trypanosoma cruzi* (L1Tc and NARTc, respectively) (Bringaud et al., 2002; Sánchez-Luque et al., 2012; 2011). The R2 elements are site-specific retrotransposons inserted into the 28S ribosomal RNA (rRNA) genes of most insect species, and are co-transcibed with the 28S rRNA (Moss et al., 2011). The presence of the HDV-like ribozyme in the 5′-untranslated region of the R2 allows self-cleavage of the 28S-R2 cotranscript at the junction between the 28S rRNA and the R2 element (Eickbush and Eickbush, 2010; Ruminski et al., 2010). Translation initiation of the open reading frame of the uncapped R2 transcript is possibly facilitated by the HDV-like ribozyme, which is thought to act also as an IRES (Ruminski et al., 2011). The interaction of the R2 protein with the R2 RNA to form a protein-RNA complex is required to target the 28S rDNA to faciliate insertion, which involves binding to the 3′- and 5′-sequences including the R2 pseudoknot structure at the 5′-end. The R2 elements have been active in the Drosophila lineage since the origin

of the genus 50 million years ago. Comparative studies on the HDV-like ribozymes from different Drosophila species revealed considerable sequence changes. Remarkably, 21 out of 27 nucleotides of the R2 ribozyme catalytic core are the same as those in the HDV ribozyme (Eickbush and Eickbush, 2010). Similarly, the L1Tc retrotransposon of *Trypanosoma cruzi* contains a HDV-like ribozyme, which facilitates the release of the transposon from a polycistronic RNA (Sánchez-Luque et al., 2012; 2011). Interestingly, L1Tc contains a dual promoter and ribozyme system. The 77 nucleotides of the L1Tc HDV-like ribozyme also act as an internal promoter (Pr77) at the DNA level. The HDV-like ribozyme cleaves upstream of its catalytic core (Ferré-D'Amaré et al., 1998; Perrotta and Been, 1991) and hence, the regulatory sequence, which contains ribozyme and promoter is preserved in the L1Tc RNA and after transposition (Sánchez-Luque et al., 2012, 2011). Sequences downstream of the L1Tc ribozyme can induce structural changes which interfere with the ribozyme activity and may promote an conformational switch to a possible IRES structure to facilitate initiation of translation, as in the case of the R2 element (Sánchez-Luque et al., 2012). The HDV ribozyme and HDV-related ribozymes are common in diverse biological systems which are at different levels of the evolutionary ladder with suggested various biological roles supporting rolling circle replication, mRNA biogenesis and gene regulation. The HDV ribozyme plays an integral role in HDV replication to generate monomeric genomes and antigenomes. The presence of HDV-like ribozymes in specific non-LTR retrotransposons and the ability of HDV-like ribozymes to retain their intact catalytic core after cleavage may have facilitated their spread by retrotransposition. It is unclear whether the nested, double knot ribozymes result from converged evolution. The sequence is highly divergent but preserved a highly complex structure with a higher level of constraints for retaining cleavage activity compared to other classes of ribozymes (Legiewicz et al., 2006; Nehdi and Perreault, 2006). The widespread distribution of HDV-related ribozymes and their ability to provide multiple functions beyond cleavage activity strongly suggests that several factors contributed to the selection for the HDV-like ribozyme. It is most likely that the ribozyme present in the HDV RNA is derived from a host transcriptome, and then optimized for cleavage to support the efficient replication of the HDV RNAs. Although Hepadnaviruses from various mammals can provide the helper function to support HDV assembly, natural HDV infections have been so far only observed in humans. HDV possibly originated from a circular RNA molecule, probably a delta-like agent, which co-evolved with HBV in the human lineage. The discovery of delta-like agents in divergent organisms (Wille et al., 2018; Chang et al., 2019; Hetzel et al., 2019; Paraskevopoulou et al., 2020), the existence of cellular cirular RNAs, and the presence of HDV-like ribozymes in cellular transcripts points toward a cellular origin of HDV.

(iii) HDV and Viroids

Similarities between HDV and viroids have been recognized based on their single stranded circular RNAs which possess a high degree on self-complementarity. Intramolecular base-pairing allows the formation of secondary structures, rod-like structures (HDV and viroids of the family *Pospiviroidae*) or Y-shaped or branched structures (viroids of the family *Avsunviroidae*). The viroids replicate in two different cellular compartments, the nucleus for the *Pospiviroidae*, and the chloroplast for the *Asunviroidae*. HDV replication is restricted to the nucleus (Flores et al., 2016; Adkar-Purushothama and Perreault, 2019). Similar to HDV, the non-protein coding viroid RNAs attract host DNA-dependent RNA polymerases for replication, which is possibly facilitated by the rod- and branched-liked secondary structures of the RNA molecules (**Table 1**). Members of the *Avsunviridae* replicate by a symmetric rolling circle cycle generating (+) and complement (-) circles, similar to the HDV replication cycle producing circular genomic and antigenomic HDV RNA molecules. This is in contrast to the asymmetric replication cycle of viroids of the family *Pospiviroidae*, which generate only (+) circles followed by cleavage of oligomeric strands by a host RNase III (Flores et al., 2009). HDV RNA cleavage of the oligomeric (+) and (−) strands, or co-transcriptional self-cleavage is assisted by *cis*-acting pseudoknot-like ribozymes (Ferré-D'Amaré et al., 1998; Perrotta and Been, 1991). Structurally distinct *cis*-acting hammerhead ribozymes are present in viroids of the family *Avsunviroidae*. The secondary structures and short double strandedness of the RNA molecules may have provided an evolutionary advantage to minimize detection by double-stranded RNA-dependent protein kinases, to provide resistance against endonucleases, and to exhibit motifs for replication, and in addition for HDV, the editing site for generating the extended ORF encoding HDAg-L. HDV shares similarities with both viroid families, the rod-like secondary structure and replication in the nucleus like the members of the *Pospiviroidae*, and the symmetric replication mechanism and cis-acting ribozyme activity is shared with the *Avsunviridae*, although the ribozyme activiy is provided by differently structured ribozyme types. Sequence complementary of viroids with the cellular 7S RNA, a component of the signal recognition particle, and a corresponding sequence similarity of the HDV antigenomic RNA and its human counterpart 7SL RNA revealed additional similarities (Negro et al., 1989; Symons, 1989; Young and Hicke, 1990). Based on the similarities between viroids and HDV circles and their intramolecular base-pairing, a concept was proposed that HDV consists of two distinct domains. One domain contains the coding sequence for the delta antigen, and the second the viroid-like domain with a size of approximately 350 nucleotides with the sequences required for replication and also the self-cleaving ribozyme activity (Lai, 1995; Lafontaine et al., 1997). The presence of different ribozyme motifs specific for HDV and delta-like agents vs. *Avsunviridae* possibly indicates that HDV is not a viroid derivative generated by recombination events.

CONCLUDING REMARKS

The findings that (i) circular RNA species are abundant in eukaryotic cells, (ii) HDV-like ribozymes are present in highly divergent organisms with important roles in many molecular pathways, and (iii) delta-like circular RNA agents are found in vertebrates and invertebrates, suggest that HDV is derived from the cellular transcriptome. Delta-like agents have not yet

been identified in non-human primates strongly indicating that a transspecies transmission event of a delta-like agent into a HBV-infected human host allowed a host shift and the emergence of HDV. The HDV sequence diversity is possibly a consequence of human migration and geographic vicariance.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

REFERENCES

Abe, N., Matsumoto, K., Nishihara, M., Nakano, Y., Shibata, A., Maruyama, H., et al. (2015). Rolling circle translation of circular RNA in living human cells. *Scientific. Rep.* 5:16435.

Adkar-Purushothama, C. R., and Perreault, J. P. (2019). Current overview on viroid-host interactions. *WIREs RNA* 11:e1570.

Aldabe, R., Suárez-Amarán, L., Usai, C., and González-Aseguinolaza, G. (2015). Animal models of chronic hepatitis delta virus infection host-virus immunological interactions. *Pathogens* 4, 46–65. doi: 10.3390/pathogens4010046

Bergner, L. M., Orton, R. J., Broos, A., Tello, C., Becker, D. J., Carrera, J. E., et al. (2021). Diversification of mammalian deltaviruses by host shifting. *Proc. Natl. Acad. Sci. USA* 118:e2019907118. doi: 10.1073/pnas.2019907118

Bichko, V., Netter, H. J., and Taylor, J. (1994). Introduction of hepatitis delta virus into animal cell lines via cationic liposomes. *J. Virol.* 68, 5247–5252. doi: 10.1128/jvi.68.8.5247-5252.1994

Bonino, F., Heermann, K. H., Rizzetto, M., and Gerlich, W. H. (1986). Hepatitis delta virus: Protein composition of delta antigen and its hepatitis B virus-derived envelope. *J. Virol.* 58, 945–950. doi: 10.1128/jvi.58.3.945-950.1986

Brazas, R., and Ganem, D. (1996). A cellular homolog of hepatitis delta antigen: Implications for viral replication and evolution. *Science* 274, 90–94. doi: 10.1126/science.274.5284.90

Bringaud, F., García-Pérez, J. L., Heras, S. R., Ghedin, E., El-Sayed, N. M., Andersson, B., et al. (2002). Identification of non-autonomous non-LTR retrotransposons in the genome of Trypanosoma cruzi. *Mol. Biochem. Parasitol.* 124, 73–78. doi: 10.1016/s0166-6851(02)00167-6

Bruss, V., Lu, X., Thomssen, R., and Gerlich, W. H. (1994). Post-translational alterations in transmembrane topology of the hepatitis B virus large envelope protein. *EMBO J.* 13, 2273–2279. doi: 10.1002/j.1460-2075.1994.tb06509.x

Casey, J. L., and Gerin, J. L. (1998). Genotype-specific complementation of hepatitis delta virus RNA replication by hepatitis delta antigen. *J. Virol.* 72, 2806–2814. doi: 10.1128/jvi.72.4.2806-2814.1998

Casey, J. L., Niro, G. A., Engle, R. E., Vega, A., Gomez, H., McCarthy, M., et al. (1996). Hepatitis B virus (HBV)/Hepatitis D virus (HDV) coinfection in outbreaks of acute hepatitis in the Peruvian Amazon basin: The roles of HDV genotype III and HBV genotype F. *J. Infect. Dis.* 174, 920–926. doi: 10.1093/infdis/174.5.920

Casey, J. L., Brown, T. L., Colan, E. J., Wignall, F. S., and Gerin, J. L. (1993). A genotype of hepatitis D virus that occurs in northern South America. *Proc. Natl. Acad. Sci. USA* 90, 9016–9020. doi: 10.1073/pnas.90.19.9016

Chang, W. S., Pettersson, J. H. O., Le Lay, C., Shi, M., Lo, N., Wille, M., et al. (2019). Novel hepatitis D-like agents in vertebrates and invertebrates. *Virus Evol.* 5:vez021.

Chang, J., Nie, X., Chang, H. E., Han, Z., and Taylor, J. (2008). Transcription of hepatitis delta virus RNA by RNA polymerase II. *J. Virol.* 82, 1118–1127. doi: 10.1128/jvi.01758-07

Chang, J., Gudima, S. O., Tarn, C., Nie, X., and Taylor, J. M. (2005). Development of a novel system to study hepatitis delta virus genome repliction. *J. Virol.* 79, 8182–8188. doi: 10.1128/jvi.79.13.8182-8188.2005

Chang, J., Moraleda, G., and Taylor, J. (2000). Limitations to replication of hepatitis delta virus in avian cells. *J. Virol.* 74, 8861–8866. doi: 10.1128/jvi.74.19.8861-8866.2000

Chang, F. L., Chen, P. J., Tu, S. J., Wang, C. J., and Chen, D. S. (1991). The large form of hepatitis δ antigen is crucial for assembly of hepatitis δ virus. *Proc. Natl. Acad. Sci. USA* 88, 8490–8494. doi: 10.1073/pnas.88.19.8490

Chemin, I., Pujol, F. H., Scholtès, C., Loureiro, C. L., Amirache, F., Levrero, M., et al. (2020). Preliminary evidence for hepatitis delta virus exposure in patients who are apparently not infected with hepatitis B virus. *Hepatology* 73, 861–864. doi: 10.1002/hep.31453

Chen, C. Y., and Sarnow, P. (1995). Initiation of protein synthesis by the eukaryotic translational apparatus on circular RNAs. *Science* 268, 415–417. doi: 10.1126/science.7536344

Chen, X., Han, P., Zhou, T., Guo, X., Song, X., and Li, Y. (2016). circRNADb: A comprehensive database for human circular RNAs with protein-coding annotations. *Sci. Rep.* 6:34985.

Chen, P. J., Chang, F. L., Wang, C. J., Lin, C. J., Sung, S. Y., and Chen, D. S. (1992). Functional study of hepatitis delta virus large antigen in packaging and replication inhibition: Role of the amino-terminal leucine zipper. *J. Virol.* 66, 2853–2859. doi: 10.1128/jvi.66.5.2853-2859.1992

Cui, J., and Holmes, E. C. (2012). Endogenous hepadnaviruses in the genome of the Budgerigar (Melopsittacus undulates) and the evolution of avian hepadnaviruses. *J. Virol.* 86, 7688–7691. doi: 10.1128/jvi.00769-12

De la Pena, M., and Cervera, A. (2017). Circular RNAs with hammerhead ribozymes encoded in eukaryotic genomes: The enemy at home. *RNA Biol.* 14, 985–991. doi: 10.1080/15476286.2017.1321730

Denniston, K. J., Hoyer, B. H., Smedile, A., Wells, F. V., Nelson, J., and Gerin, J. L. (1986). Cloned fragment of the hepatitis delta virus RNA genome: Sequence and diagnostic application. *Science* 232, 873–875. doi: 10.1126/science.3704630

Dreux, M., Garaigorta, U., Boyd, B., Décembre, E., Chung, J., Whitten-Bauer, C., et al. (2012). Short-range exosomal transfer of viral RNA from infected cells to plasmacytoid dendritic cells triggers innate immunity. *Cell Host Microb.* 12, 558–570. doi: 10.1016/j.chom.2012.08.010

Eickbush, D. G., and Eickbush, T. H. (2010). R2 retrotransposons encode a self-cleaving ribozyme for processing from an rRNA cotranscript. *Mol. Cell. Biol.* 30, 3142–3150. doi: 10.1128/mcb.00300-10

Ferré-D'Amaré, A. R., Zhou, K., and Doudna, J. A. (1998). Crystal structure of a hepatitis delta virus ribozyme. *Nature* 395, 567–574.

Fanale, D., Taverna, S., Russo, A., and Bazan, V. (2018). "Circular RNA in exosomes," in *Circular RNAs, Advances in Experimental Medicine and Biology 1087*, ed. J. Xiao (Singapore: Springer Nature), 109–117. doi: 10.1007/978-981-13-1426-1_9

Flores, R., Owens, R. A., and Taylor, J. (2016). Pathogenesis by subviral agents: viroids and hepatitis delta virus. *Curr. Opin. Virol.* 17, 87–94. doi: 10.1016/j.coviro.2016.01.022

Flores, R., Gas, M. E., Molina-Serrano, D., Nohales, M. A., Carbonell, A., Gago, S., et al. (2009). Viroid replication: Rolling-circles, enzymes and ribozymes. *Viruses* 1, 317–334. doi: 10.3390/v1020317

Freitas, N., Cunha, C., Menne, S., and Gudima, S. O. (2014). Envelope proteins derived from naturally integrated hepatitis B virus DNA support assembly and release of infectious hepatitis delta virus particles. *J. Virol.* 88, 5742–5754. doi: 10.1128/jvi.00430-14

Gerin, J. L. (2001). Animal models of hepatitis delta virus infection and disease. *ILAR J.* 42, 103–106. doi: 10.1093/ilar.42.2.103

Giersch, K., Helbig, M., Volz, T., Allweiss, L., Mancke, L. V., Lohse, A. W., et al. (2014). Persistent hepatitis D virus mono-infection in humanized mice is efficiently converted by hepatitis B virus to a productive co-infection. *J. Hepatol.* 60, 538–544. doi: 10.1016/j.jhep.2013.11.010

Gilbert, C., and Feschotte, C. (2010). Genomic fossils calibrate the long-term evolution of hepadnaviruses. *PLoS Biol.* 8:e1000495. doi: 10.1371/journal.pbio.1000495

Glenn, J. S., Watson, J. A., Havel, C. M., and White, J. M. (1992). Identification of a prenylation site in delta virus large antigen. *Science* 256, 1331–1333. doi: 10.1126/science.1598578

Gould, S. J., Booth, A. M., and Hildreth, J. E. K. (2003). The Trojan exosome hypothesis. *Proc. Natl. Acad. Sci. USA* 100, 10592–10597. doi: 10.1073/pnas.1831413100

Han, M., Littlejohn, M., Yuen, L., Edwards, R., Devi, U., Bowden, S., et al. (2014). Molecular epidemiology of hepatitis delta virus in the Western Pacific region. *J. Clin. Virol.* 61, 34–39. doi: 10.1016/j.jcv.2014.05.021

He, L. F., Ford, E., Purcell, R. H., London, W. T., Phillips, J., and Gerin, J. L. (1989). The size of the hepatitis delta agent. *J. Med. Virol.* 27, 31–33. doi: 10.1002/jmv.1890270107

Hetzel, U., Szirovicza, L., Smura, T., Prähauser, B., Vapalahti, O., Kipar, A., et al. (2019). Identification of a novel deltavirus in Boa constrictors. *Mbio* 10, e00014–e00019.

Hsieh, S. Y., Chao, M., Coates, L., and Taylor, J. (1990). Hepatitis delta virus genome replication: a polyadenylated mRNA for delta antigen. *J. Virol.* 64, 3192–3198. doi: 10.1128/jvi.64.7.3192-3198.1990

Hsu, M. T., and Coca-Prados, M. (1979). Electron microscopic evidence for the circular form of RNA in the cytoplasm of eukaryotic cells. *Nature* 280, 339–340. doi: 10.1038/280339a0

Hu, J., and Liu, K. (2017). Complete and incomplete hepatitis B virus particles: Formation, function, and application. *Viruses* 9:56. doi: 10.3390/v9030056

Hwang, S. B., and Lai, M. M. C. (1993). Isoprenylation mediates direct protein-protein interactions between hepatitis large delta antigen and hepatitis B virus surface antigen. *J. Virol.* 67, 7659–7662. doi: 10.1128/jvi.67.12.7659-7662.1993

International Committee on Taxonomy of Viruses (ICTV) (2020). *Deltavirus*. Available online at: https://talk.ictvonline.org/taxonomy/p/taxonomy-history?taxnode_id=202005347

Ivanov, A., Memczak, S., Wyler, E., Torti, F., Porath, H. T., Orejuela, M. R., et al. (2015). Analysis of intron sequences reveals hallmarks of circular RNA biogenesis in animals. *Cell Rep.* 10, 170–177. doi: 10.1016/j.celrep.2014.12.019

Iwamoto, M., Shibata, Y., Kawasaki, J., Kojima, S., Li, Y.-T., Iwami, S., et al. (2021). Identification of novel avian and mammalian deltaviruses provides new insights into deltavirus evolution. *Virus Evol.* 7:veab003.

Jeck, W. R., Sorrentino, J. A., Wang, K., Slevin, M. K., Burd, C. E., Liu, J., et al. (2013). Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA* 19, 141–157. doi: 10.1261/rna.035667.112

Jenna, S., and Sureau, C. (1999). Mutations in the carboxyl-terminal domain of the small hepatitis B virus envelope protein impair the assembly of hepatitis delta virus particles. *J. Virol.* 73, 3351–3358. doi: 10.1128/jvi.73.4.3351-3358.1999

Jimenez, R. M., Polanco, J. A., and Lupták, A. (2015). Chemistry and biology of self-cleaving ribozymes. *Trends Biochem. Sci.* 40, 648–661. doi: 10.1016/j.tibs.2015.09.001

Kakizaki, M., Yamamoto, Y., Yabuta, S., Kurosaki, N., Kagawa, T., and Kotani, A. (2018). The immunological function of extracellular vesicles in hepatitis B virus-infected hepatocytes. *PLoS One* 13:e0205886. doi: 10.1371/journal.pone.0205886

Kim, K. M., Abdelmohsen, K., Mustapic, M., Kapogiannis, D., and Gorospe, M. (2017). RNA in extracellular vesicles. *WIREs RNA* 8:e1413. doi: 10.1002/wrna.1413

Komla-Soukha, I., and Sureau, C. (2006). A tryptophan-rich motif in the carboxyl terminus of the small envelope protein of hepatitis B virus is central to the assembly of hepatitis delta virus particles. *J. Virol.* 80, 4648–4655. doi: 10.1128/jvi.80.10.4648-4655.2006

Kos, A., Dijkema, R., Arnberg, A. C., van der Meide, P. H., and Schellekens, H. (1986). The hepatitis delta (δ) virus possesses a circular RNA. *Nature* 323, 558–560. doi: 10.1038/323558a0

Kossinova, O. A., Gopanenko, A. V., Tamkovich, S. N., Krasheninina, O. A., Tupikin, A. E., Kiseleva, E., et al. (2017). Cytosolic YB-1 and NSUN2 are the only proteins recognizing specific motifs present in mRNAs enriched in exosomes. *BBA - Proteins Proteom.* 1865, 664–673. doi: 10.1016/j.bbapap.2017.03.010

Kouwaki, T., Fukushima, Y., Daito, T., Sanada, T., Yamamoto, N., Mifsud, E. J., et al. (2016). Extracellular vesicles including exosomes regulate innate immune responses to hepatitis B virus infection. *Front. Immunol.* 7:335.

Kristensen, L. S., Andersen, M. S., Stagsted, L. V. W., Ebbesen, K. K., Hansen, T. B., and Kjems, J. (2019). The biogenesis, biology and characterization of circular RNAs. *Nat. Rev. Genet.* 20, 675–691.

Kuo, M. Y. P., Chao, M., and Taylor, J. (1989). Initiation of replication of the human hepatitis delta virus genome from cloned DNA: Role of delta antigen. *J. Virol.* 63, 1945–1950. doi: 10.1128/jvi.63.5.1945-1950.1989

Kuo, M. Y. P., Sharmeen, L., Dinter-Gottlieb, G., and Taylor, J. (1988). Characterization of self-cleaving RNA sequences on the genome and antigenome of human hepatitis delta virus. *J. Virol.* 62, 4439–4444. doi: 10.1128/jvi.62.12.4439-4444.1988

Lai, M. M. C. (1995). The molecular biology of hepatitis delta virus. *Ann. Rev. Biochem.* 64, 259–286.

Lafontaine, D., Mercure, S., and Perreault, J. P. (1997). Update of the viroid and viroid-like sequence database: addition of a hepatitis delta virus RNA section. *Nucl. Acids Res.* 25, 123–125. doi: 10.1093/nar/25.1.123

Lasda, E., and Parker, R. (2014). Circular RNAs: diversity of form and function. *RNA* 20, 1829–1842. doi: 10.1261/rna.047126.114

Le Gal, F., Brichler, S., Drugan, T., Alloui, C., Roulot, D., Pawlotsky, J. M., et al. (2017). Genetic diversity and worldwide distribution of the Deltavirus genus: A study of 2,152 clinical strains. *Hepatology* 66, 1826–1841. doi: 10.1002/hep.29574

Legiewicz, M., Wichlacz, A., Brzezicha, B., and Ciesiolka, J. (2006). Antigenomic delta ribozyme variants with mutations in the catalytic core obtained by the *in vitro* selection method. *Nucl. Acids Res.* 34, 1270–1280. doi: 10.1093/nar/gkl018

Long, M., de Souza, S. J., and Gilbert, W. (1997). Delta-interacting protein A and the origin of hepatitis delta antigen. *Science* 276, 824–825. doi: 10.1126/science.276.5313.824

Luo, G., Chao, M., Hsieh, S. Y., Sureau, C., Nishikura, K., and Taylor, J. (1990). A specific base transition occurs on replicating hepatitis delta virus RNA. *J. Virol.* 64, 1021–1027. doi: 10.1128/jvi.64.3.1021-1027.1990

Magnius, L., Taylor, J., Mason, W. S., Sureau, C., Dény, P., Norder, H., et al. (2018). ICTV virus taxonomy profile: Deltavirus. *J. Gen. Virol.* 99, 1565–1566. doi: 10.1099/jgv.0.001150

Makino, S., Chang, M. F., Shieh, C. K., Kamahora, T., Vannier, D. M., Govindarajan, S., et al. (1987). Molecular cloning and sequencing of a human hepatitis delta (δ) virus RNA. *Nature* 329, 343–346. doi: 10.1038/329343a0

Mederacke, I., Filmann, N., Yurdaydin, C., Bremer, B., Puls, F., Zacher, B. J., et al. (2012). Rapid early HDV RNA decline in the peripheral blood but prolonged intrahepatic hepatitis delta antigen persistence after liver transplantation. *J. Hepatol.* 56, 115–122. doi: 10.1016/j.jhep.2011.06.016

Modahl, L. E., Macnaughton, T. B., Zhu, N., Johnson, D. L., and Lai, M. M. C. (2000). RNA-dependent replication and transcription of hepatitis delta virus RNA involve distinct cellular RNA polymerases. *Mol. Cell. Biol.* 20, 6030–6039. doi: 10.1128/mcb.20.16.6030-6039.2000

Moss, W. N., Eickbush, D. G., Lopez, M. J., Eickbush, T. H., and Turner, D. H. (2011). The R2 retrotransposon RNA families. *RNA Biol.* 8, 714–718. doi: 10.4161/rna.8.5.16033

Negro, F., Gerin, J. L., Purcell, R. H., and Miller, R. H. (1989). Basis of hepatitis delta virus disease? *Nature* 341, 111. doi: 10.1038/341111b0

Nehdi, A., and Perreault, J. P. (2006). Unbiased *in vitro* selection reveals the unique character of the self-cleaving antigenomic HDV RNA sequence. *Nucl. Acids Res.* 34, 584–592. doi: 10.1093/nar/gkj463

Netter, H. J., Gerin, J. L., Tennant, B. C., and Taylor, J. M. (1994). Apparent helper-independent infection of woodchucks by hepatitis delta virus and subsequent rescue with woodchuck hepatitis virus. *J. Virol.* 68, 5344–5350. doi: 10.1128/jvi.68.9.5344-5350.1994

Netter, H. J., Kajino, K., and Taylor, J. M. (1993). Experimental transmission of human hepatitis delta virus to the laboratory mouse. *J. Virol.* 67, 3357–3362. doi: 10.1128/jvi.67.6.3357-3362.1993

O'Malley, B., and Lazinski, D. W. (2005). Roles of carboxyl-terminal and farnesylated residues in the functions of the large hepatitis delta antigen. *J. Virol.* 79, 1142–1153. doi: 10.1128/jvi.79.2.1142-1153.2005

Ostapchuk, P., Hearing, P., and Ganem, D. (1994). A dramatic shift in the transmembrane topology of a viral envelope glycoprotein accompanies hepatitis B viral morphogenesis. *EMBO J.* 13, 1048–1057. doi: 10.1002/j.1460-2075.1994.tb06353.x

Ottobrelli, A., Marzano, A., Smedile, A., Recchia, S., Salizzoni, M., Cornu, C., et al. (1991). Patterns of hepatitis delta virus reinfection and disease in liver transplantation. *Gastroenterology* 101, 1649–1655. doi: 10.1016/0016-5085(91)90404-9

Pamudurti, N. R., Bartok, O., Jens, M., Ashwal-Fluss, R., Stottmeister, C., Ruhe, L., et al. (2017). Translation of circRNAs. *Mol. Cell* 66, 9–21.

Paraskevopoulou, S., Pirzer, F., Goldmann, N., Schmid, J., Corman, V. M., Gottula, L. T., et al. (2020). Mammalian deltavirus without hepadnavirus coinfection in the neotropical rodent Proechimys semispinosus. *Proc. Natl. Acad. Sci. USA* 117, 17977–17983.

Perez-Vargas, J., Amirache, F., Boson, B., Mialon, C., Freitas, N., Sureau, C., et al. (2019). Enveloped viruses distinct from HBV induce dissemination of hepatitis D virus *in vivo*. *Nat. Comm.* 10:2098.

Perrotta, A. T., and Been, M. D. (1991). A pseudoknot-like structure required for efficient self-cleavage of hepatitis delta virus RNA. *Nature* 350, 434–436. doi: 10.1038/350434a0

Pflüger, L. S., Schulze zur Wiesch, J., Polywka, S., and Lütgehetmann, M. (2020). Hepatitis delta virus propagation enabled by hepatitis C virus – Scientifically intriguing, but is it relevant to clinical practice? *J. Viral. Hepat.* 28, 213–216. doi: 10.1111/jvh.13385

Polo, J. M., Jeng, K. S., Lim, B., Govindarajan, S., Hofman, F., Sangiorgi, F., et al. (1995). Transgenic mice support replication of hepatitis delta virus RNA in multiple tissues, particularly in skeletal muscle. *J. Virol.* 69, 4880–4887. doi: 10.1128/jvi.69.8.4880-4887.1995

Polson, A. G., Bass, B. L., and Casey, J. L. (1996). RNA editing of hepatitis delta virus antigenome by dsRNA-adenosine deaminase. *Nature* 380, 454–456. doi: 10.1038/380454a0

Ponzetto, A., Cote, P. J., Popper, H., Hoyer, B. H., London, W. T., Ford, E. C., et al. (1984). Transmission of the hepatitis B virus-associated δ agent to the eastern woodchuck. *Proc. Natl. Acad. Sci. USA* 81, 2208–2212. doi: 10.1073/pnas.81.7.2208

Radjef, N., Gordien, E., Ivaniushina, V., Gault, E., Anaïs, P., Drugan, T., et al. (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, 2537–2544. doi: 10.1128/jvi.78.5.2537-2544.2004

Ramakrishnaiah, V., Thumann, C., Fofana, I., Habersetzer, F., Pan, Q., de Ruiter, P. E., et al. (2013). Exosome-mediated transmission of hepatitis C virus between human hepatoma Huh7.5 cells. *Proc. Natl. Acad. Sci. USA* 110, 13109–13113. doi: 10.1073/pnas.1221899110

Reid, C. E., and Lazinski, D. W. (2000). A host-specific function is required for ligation of a wide variety of ribozyme-processed RNAs. *Proc. Natl. Acad. Sci. USA* 97, 424–429. doi: 10.1073/pnas.97.1.424

Riccitelli, N., and Lupták, A. (2013). "HDV family of self-cleaving ribozymes," in *Progress in Molecular Biology and Translational Science*, Vol. 120, ed. G. A. Soukup (Burlington, MA: Academic Press), 123–171.

Rizzetto, M., Canese, M. G., Gerin, J. L., London, W. T., Sly, D. L., and Purcell, R. H. (1980). Transmission of the hepatitis B virus-associated delta antigen to Chimpanzees. *J. Inf. Dis.* 141, 590–602. doi: 10.1093/infdis/141.5.590

Rizzetto, M., Canese, M. G., Aricò, S., Crivelli, O., Trepo, C., Bonino, F., et al. (1977). Immunofluorescence detection of a new antigen-antibody system (δ/anti-δ) associated to hepatitis B virus in liver and in serum of HBsAg carriers. *Gut* 18, 997–1003. doi: 10.1136/gut.18.12.997

Ruminski, D. J., Webb, C. H. T., Riccitelli, N. J., and Lupták, A. (2011). Processing and translation initiation of non-long terminal repeat retrotransposons by hepatitis delta virus (HDV)-like self-cleaving ribozymes. *J. Biol. Chem.* 286, 41286–41295. doi: 10.1074/jbc.m111.297283

Ruminski, D., Webb, C. H. T., Riccitelli, N. J., and Lupták, A. (2010). Processing of insect retrotransposons by self-cleaving ribozymes. *Nat. Precedings* 2010:4333. doi: 10.1038/ npre.2010.4333.1

Ryu, W. S., Netter, H. J., Bayer, M., and Taylor, J. (1993). Ribonucleoprotein complexes of hepatitis delta virus. *J. Virol.* 67, 3281–3287. doi: 10.1128/jvi.67.6.3281-3287.1993

Ryu, W. S., Bayer, M., and Taylor, J. (1992). Assembly of hepatitis delta virus particles. *J. Virol.* 66, 2310–2315. doi: 10.1128/jvi.66.4.2310-2315.1992

Sänger, H. L., Klotz, G., Riesner, D., Gross, H. J., and Kleinschmidt, A. K. (1976). Viroids are single-stranded covalently closed circular RNA molecules existing as highly base-paired rod-like structures. *Proc. Natl. Acad. Sci. USA* 73, 3852–3856. doi: 10.1073/pnas.73.11.3852

Salehi-Ashtiani, K., Lupták, A., Litovchick, A., and Szostak, J. W. (2006). A genomewide search for ribozymes reveals an HDV-like sequence in the human *CPEB3* gene. *Science* 313, 1788–1792. doi: 10.1126/science.1129308

Samuel, D., Zignego, A. L., Reynes, M., Feray, C., Arulnaden, J. L., David, M. F., et al. (1995). Long-term clinical and virological outcome after liver transplantation for cirrhosis caused by chronic delta hepatitis. *Hepatology* 21, 333–339. doi: 10.1002/hep.1840210211

Sánchez-Luque, F. J., López, M. C., Macias, F., Alonso, C., and Thomas, M. C. (2012). Pr77 and L1TcRz. A dual system within the 5′end of L1Tc retrotransposon, internal promoter and HDV-like ribozyme. *Mobile Genet. Elements* 2, 1–7. doi: 10.4161/mge.19233

Sánchez-Luque, F. J., López, M. C., Macias, F., Alonso, C., and Thomas, M. C. (2011). Identification of an hepatitis delta virus-like ribozyme at the mRNA 5′end of the L1Tc retrotransposon from Trypanosoma cruzi. *Nucl. Acids. Res.* 39, 8065–8077. doi: 10.1093/nar/gkr478

Sharmeen, L., Kuo, M. Y. P., and Taylor, J. (1989). Self-ligating RNA sequences on the antigenome of human hepatitis delta virus. *J. Virol.* 63, 1428–1430. doi: 10.1128/jvi.63.3.1428-1430.1989

Smedile, A., Casey, J. L., Cote, P. J., Durazzo, M., Lavezzo, B., Purcell, R. H., et al. (1998). Hepatitis D viremia following orthotopic liver transplantation involves a typical HDV virion with a hepatitis B surface antigen envelope. *Hepatology* 27, 1723–1729. doi: 10.1002/hep.510270636

Sureau, C., and Negro, F. (2016). The hepatitis delta virus: replication and pathogenesis. *J. Hepatol.* 64, S102–S116.

Sureau, C., Guerra, B., and Lanford, R. E. (1993). Role of the large hepatitis B virus envelope protein in infectivity of the hepatitis delta virion. *J. Virol.* 67, 366–372. doi: 10.1128/jvi.67.1.366-372.1993

Symons, B. (1989). Pathogenesis by antisense. *Nature* 338, 542–543. doi: 10.1038/338542a0

Szirovicza, L., Hetzel, U., Kipar, A., Martinez-Sobrido, L., Vapalahti, O., and Hepojoki, J. (2020). Snake deltavirus utilizes envelope proteins of different viruses to generate infectious particles. *mBio* 11, e3250-19.

Veziroglu, E. M., and Mias, G. I. (2020). Characterizing extracellular vesicles and their diverse RNA contents. *Front. Genetics* 11:700.

Villarroya-Beltri, C., Gutiérrez-Vázquez, C., Sánchez-Cabo, F., Pérez-Hernández, D., Vázquez, J., Martin-Cofreces, N., et al. (2013). Sumoylated hnRNPA2B1 controls the sorting of miRNAs into exosomes through binding to specific motifs. *Nat. Comm.* 4:2980.

Wadkins, T. S., Perrotta, A. T., Ferré-D'Amaré, A. R., Doudna, J. A., and Been, M. D. (1999). A nested double pseudoknot is required for self-cleavage activity of both the genomic and antigenomic hepatitis delta virus ribozymes. *RNA* 5, 720–727. doi: 10.1017/s1355838299990209

Wang, Y., and Wang, Z. (2015). Efficient backsplicing produces translatable circular mRNAs. *RNA* 21, 172–179. doi: 10.1261/rna.048272.114

Wang, K.-S., Choo, Q.-L., Weiner, A. J., Ou, J. H., Najarian, R. C., Thayer, R. M., et al. (1986). Structure, sequence and expression of the hepatitis delta (δ) viral genome. *Nature* 323, 508–514. doi: 10.1038/323508a0

Webb, C. H. T., and Lupták, A. (2011). HDV-like self-cleaving ribozymes. *RNA Biol.* 8, 719–727. doi: 10.4161/rna.8.5.16226

Webb, C. H. T., Riccitelli, N., Ruminski, D. J., and Lupták, A. (2009). Widespread occurrence of self-cleaving ribozymes. *Science* 326, 953. doi: 10.1126/science.1178084

Weller, M. L., Gardener, M. R., Bogus, Z. C., Smith, M. A., Astorri, E., Michael, D. G., et al. (2016). Hepatitis delta virus detected in Salivary glands of Sjögren's syndrome patients and recapitulates a Sjögren's syndrome-lik phenotype *in vivo*. *Pathog Immun.* 1, 12–33.

Wille, M., Netter, H. J., Littlejohn, M., Yuen, L., Shi, M., Eden, J. S., et al. (2018). A divergent hepatitis D-like agent in birds. *Viruses* 10:720. doi: 10.3390/v10120720

Wong, S. K., and Lazinski, D. W. (2002). Replicating hepatitis delta virus RNA is edited in the nucleus by the small form of ADAR1. *Proc. Natl. Acad. Sci. USA* 99, 15118–15123. doi: 10.1073/pnas.232416799

Yamaguchi, Y., Filipovska, J., Yano, K., Furuya, A., Inukai, N., Narita, T., et al. (2001). Stimulation of the RNA polymerase II elongation by hepatitis delta antigen. *Science* 293, 124–127. doi: 10.1126/science.1057925

Yan, H., Zhong, G., Xu, G., He, W., Jing, Z., Gao, Z., et al. (2012). Sodium taurocholate cotransporting polypeptide is a

functional receptor for human hepatitis B and D virus. *eLIFE* 1: e00049.

Yang, Y., Fan, X., Mao, M., Song, X., Wu, P., Zhang, Y., et al. (2017). Extensive translation of circular RNAs driven by N6-methyladenosine. *Cell Res.* 27, 626–641. doi: 10.1038/cr.2017.31

Young, B., and Hicke, B. (1990). Delta virus as a cleaver. *Nature* 343, 28. doi: 10.1038/343028a0

Conflict of Interest: The author HJN has an equity interest in, and serves as a consultant to ClearB Therapeutics. ClearB Therapeutics had no role in the writing of the article.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 08 July 2021
doi: 10.3389/fmicb.2021.671466

Strong Replication Interference Between Hepatitis Delta Viruses in Human Liver Chimeric Mice

Katja Giersch[1†], Lennart Hermanussen[1†], Tassilo Volz[1], Annika Volmari[1], Lena Allweiss[1,2], Camille Sureau[3], John Casey[4], Jiabin Huang[5], Nicole Fischer[5], Marc Lütgehetmann[2,5†] and Maura Dandri[1,2†]*

[1] Department of Internal Medicine, University Medical Center Hamburg-Eppendorf, Hamburg, Germany, [2] German Center for Infection Research (DZIF), Hamburg-Lübeck-Borstel-Riems Site, Hamburg, Germany, [3] Institut National de la Transfusion Sanguine, Paris, France, [4] Georgetown University Medical Center, Washington, DC, United States, [5] Department of Medical Microbiology, Virology and Hygiene, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

OPEN ACCESS

Edited by:
Lilly Yuen, Victorian Infectious Diseases Reference Laboratory, Australia

Reviewed by:
Eloi R. Verrier, INSERM UMR_S1110 Institute de Recherche sur les Maladies Virales et Hepatiques, France
Philip Meuleman, Ghent University, Belgium

***Correspondence:**
Maura Dandri
m.dandri@uke.de

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *23 February 2021*
Accepted: *08 June 2021*
Published: *08 July 2021*

Citation:
Giersch K, Hermanussen L, Volz T, Volmari A, Allweiss L, Sureau C, Casey J, Huang J, Fischer N, Lütgehetmann M and Dandri M (2021) Strong Replication Interference Between Hepatitis Delta Viruses in Human Liver Chimeric Mice. Front. Microbiol. 12:671466. doi: 10.3389/fmicb.2021.671466

Background: Hepatitis D Virus (HDV) is classified into eight genotypes with distinct clinical outcomes. Despite the maintenance of highly conserved functional motifs, it is unknown whether sequence divergence between genotypes, such as HDV-1 and HDV-3, or viral interference mechanisms may affect co-infection in the same host and cell, thus hindering the development of HDV inter-genotypic recombinants. We aimed to investigate virological differences of HDV-1 and HDV-3 and assessed their capacity to infect and replicate within the same liver and human hepatocyte *in vivo*.

Methods: Human liver chimeric mice were infected with hepatitis B virus (HBV) and with one of the two HDV genotypes or with HDV-1 and HDV-3 simultaneously. In a second set of experiments, HBV-infected mice were first infected with HDV-1 and after 9 weeks with HDV-3, or vice versa. Also two distinct HDV-1 strains were used to infect mice simultaneously and sequentially. Virological parameters were determined by strain-specific qRT-PCR, RNA *in situ* hybridization and immunofluorescence staining.

Results: HBV/HDV co-infection studies indicated faster spreading kinetics and higher intrahepatic levels of HDV-3 compared to HDV-1. In mice that simultaneously received both HDV strains, HDV-3 became the dominant genotype. Interestingly, antigenomic HDV-1 and HDV-3 RNA were detected within the same liver but hardly within the same cell. Surprisingly, sequential super-infection experiments revealed a clear dominance of the HDV strain that was inoculated first, indicating that HDV-infected cells may acquire resistance to super-infection.

Conclusion: Infection with two largely divergent HDV genotypes could be established in the same liver, but rarely within the same hepatocyte. Sequential super-infection with distinct HDV genotypes and even with two HDV-1 isolates was strongly impaired, suggesting that virus interference mechanisms hamper productive replication in the same cell and hence recombination events even in a system lacking adaptive immune responses.

Keywords: HDV, genotypes, human liver chimeric mice, humanized mice, *in vivo*, replication, recombination, hepatitis delta

INTRODUCTION

Around 20 million people worldwide are infected with the hepatitis delta virus (HDV) and recent reports suggested that the number of HDV-positive individuals may be even substantially higher (Chen et al., 2018). Its viral genome is a circular, negative-sense, single-stranded RNA, which forms a characteristic unbranched rod-like structure with paired nucleotides (Wang et al., 1986). HDV redirects the host RNA polymerase II to replicate via a double rolling-circle amplification process (Lai, 2005) leading to the intracellular accumulation of two additional RNAs: the antigenomic RNA (AG HDV RNA), which is an exact complement of the genomic RNA (G HDV RNA) and the smaller linear mRNA encoding for the only viral protein, the hepatitis delta antigen (HDAg). HDAg binds specifically to the HDV RNA and occurs in two different forms: the small HDAg (S-HDAg, 24 kDa) is important for virus replication, whereas the large variant (L-HDAg, 27 kDa) inhibits replication and promotes virus assembly (Chang et al., 1991; Wang et al., 1991). The L-HDAg is generated during virus replication by post-transcriptional RNA editing at adenosine 1012 (amber/W site), which is mediated by RNA-specific adenosine deaminase (ADAR). HDV requires the envelope proteins of the hepatitis B virus (HBV) in order to assemble into infectious particles and spread (Rizzetto et al., 1980; Freitas et al., 2014). Hence, HBV plays an essential role as helper virus for HDV transmission. Recently, the envelope proteins of other HBV-unrelated viruses (such as dengue virus, hepatitis C virus (HCV), or west nile virus) were shown to act as alternative helper viruses enabling coating of HDV *in vitro* (Perez-Vargas et al., 2019). However, clinical relevance for HCV/HDV co-infection appears unlikely, since three different studies analyzing chronic HCV-infected patients revealed only one case with detectable HDV RNA in the absence of HBV (Cappy et al., 2020; Chemin et al., 2020; Pfluger et al., 2021). Both HBV and HDV enter human hepatocytes via the sodium taurocholate co-transporting polypeptide (NTCP) (Yan et al., 2012) and active HDV infection can occur either upon simultaneous co-infection with HBV or as a super-infection in patients already infected with HBV. Chronic HDV infections are associated with severe liver disease and progression to cirrhosis, liver decompensation, hepatocellular carcinoma and death (Koh et al., 2019). HDV is classified into eight genotypes with distinct clinical courses (Le Gal et al., 2006; Botelho-Souza et al., 2017; Delfino et al., 2018). Sequence divergence among the genotypes is as high as 40% over the entire RNA genome with the greatest difference observed between HDV genotype 1 (HDV-1) and genotype 3 (HDV-3) (Deny, 2006). However, a recent systematic analysis of HDV sequences revealed highly conserved functional nucleotide and amino acid motifs across all genotypes, indicating strong conservatory constraints on the structure and function of the HDV genome and the HDAg proteins (Le Gal et al., 2017). HDV-1 is distributed around the world and causes a broad spectrum of chronic diseases. HDV-3 is almost exclusively detected in patients from the Amazonian region, where HDV infection is associated with particularly severe disease (Botelho-Souza et al., 2017). An *in vitro* study suggested that the severity of hepatic inflammation might be associated with the distinct efficacies of HDV genotypes to generate L-HDAg, package and secrete viral particles (Hsu et al., 2002). To date, most *in vitro* and *in vivo* experiments had been performed with a certain HDV-1 strain, which was obtained from an infected patient, serially passaged through chimpanzees and then in 1988 isolated and cloned after being transmitted to a woodchuck (Kuo et al., 1988, 1989). Ten years later, an HDV-3 isolate (Peru-1) was obtained from an 18-year-old man from Peru, who developed severe acute hepatitis, and cloned by Casey et al. (1993). However, little is known about the behavior and kinetics of different HDV isolates and divergent strains such as HDV-3 and HDV-1 *in vivo*. Such analyses may help understand differences in HDV pathogenesis and disease outcome.

Similar to other RNA viruses co-infections and inter-genotypic recombination between different HDV genotypes can occur. However, little is known about the ability of distinct HDV genotypes to infect and actively replicate within the same hepatocyte and, to our knowledge, recombination events have only been reported between HDV-1 and HDV genotype 2 (HDV-2), which are closely related (Wu et al., 1999; Wang and Chao, 2005; Sy et al., 2015). Co-infections including HDV-3, such as HDV1/HDV-3 co-infections in patients, have not been reported yet (Cicero et al., 2016), perhaps because HDV-3 is geographically rather isolated in the Amazonas region or because the high sequence divergence of HDV-3 to other genotypes prevents productive co-infection and inter-genotypic recombination. Virus interference limiting a productive co-infection of the same cell with HDV of distinct genotypes might play an important role in HDV epidemiology and geographical distribution.

The aim of this study was to investigate virological differences between two distinct cloned genotypes (HDV-1 and HDV-3) in human liver chimeric mice in the presence of HBV. Furthermore, to assess whether HDV-1 and HDV-3 can efficiently spread within the same liver and even replicate within the same hepatocyte, we performed both simultaneous and sequential super-infections experiments with HDV-1 and HDV-3 strains *in vivo* in a system lacking adaptive immune responses.

Abbreviations: AG, antigenomic; ALT, alanine aminotransferase; cc, cell culture derived; cccDNA, covalently closed circular DNA; CK18, cytokeratin 18; CXCL10, C-X-C motif chemokine ligand 10; G, genomic; GAPDH, glyceraldehyde-3 phosphate dehydrogenase; HBsAg, hepatitis B surface antigen; HBV, hepatitis B virus; HDAg, hepatitis delta antigen (S, small, L, large); HCV, hepatitis C virus; HDV, hepatitis delta virus; HIV, human immunodeficiency virus; HLA, human leukocyte antigen; HSA, human serum albumin; ISGs, interferon stimulated genes; LLoD, lower limit of detection; MxA, Myxovirus resistance gene A; NK, natural killer; NTCP, natrium taurocholate polypeptide; p, patient derived; OAS1, 2′-5′-oligoadenylate synthetase 1; pgRNA, pregenomic RNA; qRT-PCR, quantitative real time polymerase chain reaction; SCID, severe combined immunodeficiency; uPA, urokinase plasminogen activator; USG, uPA/SCID/beige/IL2RG$^{-/-}$.

MATERIALS AND METHODS

Generation of Humanized USG Mice

Human liver chimeric urokinase-type plasminogen activator (uPA)/severe combined immunodeficiency (SCID)/beige/interleukin-2 receptor gamma chain negative

(IL2RG$^{-/-}$) mice (short USG mice) were generated by transplanting one million thawed cryo-preserved human hepatocytes into homozygous USG mice as previously reported (Lutgehetmann et al., 2012). Repopulation rates were estimated by determining human serum albumin (HSA) in mouse sera (ELISA; Bethyl Laboratories, Biomol GmbH, Hamburg, Germany) and human beta-globin in mouse liver DNA (qRT-PCR; Taqman Gene Expression Assay Hs00758889_s1; Applied Biosystems, Carlsbad, CA, United States) (Lutgehetmann et al., 2012). Animals displaying high levels of human chimerism (>2 mg/ml HSA in serum) were used for the study. All mice were sacrificed at the end of the experiment (at different time-points as indicated in the results), blood was collected and liver specimens were snap-frozen in chilled isopentane and cryo-conserved at −80°C for further histological and molecular analyses. Mice were maintained under specific pathogen free conditions in accordance with institutional guidelines under approved protocols. All animal experiments were conducted in accordance with the European Communities Council Directive (86/609/EEC) and were approved by the City of Hamburg, Germany.

Virus Generation and Infections

Human liver chimeric USG mice were either co-infected with HBV/HDV-1cc or HBV/HDV-3cc, or first infected with HBV and after 9 weeks super-infected simultaneously with HDV-1cc and HDV-3cc or HDV-1p. For each co-infection setting and for simultaneous HDV1/HDV3 super-infection, mice received a single peritoneal injection of cell culture derived HBV genotype D (1×10^7 HBV genome equivalents/mouse, kindly provided by Dieter Glebe, Gießen, Germany), cell culture derived HDV-1 (HDV-1cc) (1×10^7 HDV genome equivalents/mouse) and/or cell culture derived HDV-3 (HDV-3cc) (1×10^7 HDV genome equivalents/mouse). For simultaneous super-infection with two different HDV-1 isolates chronic HBV-infected mice (>9 weeks) received patient derived HDV-1 (HDV-1p) (1×10^5 HDV genome equivalents/mouse) and HDV-1cc (1×10^6 HDV genome equivalents/mouse). For sequential HDV super-infection experiments, chronic HBV-infected mice (>9 weeks) were infected with HDV-3cc and after 9 weeks super-infected with HDV-1cc and vice versa - using the same inocula as described above. HBV-infected mice were also infected with a patient-derived HDV-1 strain (HDV-1p) and then super-infected with either HDV-3cc or HDV-1cc. Cell culture derived HDV-1cc and HDV-3cc particles were generated in HuH7 cells as previously described (Gudima et al., 2007). In brief, cells were transfected with 1 μg of the HDV recombinant plasmid pSVL(D3) for HDV-1 (kindly provided by John Taylor, Philadelphia, PA, United States) (Kuo et al., 1989) or pCMV3-Peru-1.2 for HDV-3 (Casey and Gerin, 1998) and 1 μg of the HBV envelope-expressing vector pT7HB2.7 using Fugene HD Transfection Reagent (Promega, Madison, United States). HDV-1p was isolated from a chronic HBV/HDV co-infected patient and passaged through humanized USG mice. 50 μl HBV/HDV-1p-positive mouse serum, which contained 7.3×10^6 HBV DNA copies/ml and 1.3×10^8 HDV RNA copies/ml, was used for infection. The inocula (1×10^7 HDV genome equivalents/mouse) corresponded to a MOI of approximately 0.3 by estimating an average of 3×10^7 human hepatocytes per mouse liver (Dandri et al., 2008).

Virological Measurements in Serum and Liver

Viral DNA and RNA were extracted from serum samples using the QiAmp MinElute Virus Spin kit (Qiagen, Hilden, Germany) and from liver tissues using the MasterPure™ Complete DNA and RNA Purification Kit (Epicentre, Madison, WI, United States) and the Qiagen RNeasy Mini Kit. HDV viremia and intrahepatic HDV RNA levels were determined by reverse transcription and qRT-PCR using the ABI Fast 1-Step Virus Master (Applied Biosystems, Waltham, MA, United States) and HDV Taqman primers and probes on an ABI Viia7 (Applied Biosystems, Waltham, MA, United States) as previously described (Giersch et al., 2019). Primers and probes recognized either all HDV genotypes (Ferns et al., 2011) or specifically HDV-1 (all isolates) (Mederacke et al., 2010), HDV-1cc (HDV1cc-fw: 5′-TCA CGG TAA AGA GCA TTG, HDV1cc-rv: 5′-TTC CCC TTC CAG AGA TTC-3′, HDV1cc probe: 5′-/56-FAM/CGT CCG CTT/ZEN/CCT GAG ACC), HDV-1p (HDV1p-fw: 5′-AGG AGT AAG ATC ATA GCG ATA-3′, HDV1p-rv: 5′-CTG CTC TCT TTG CTT TCC-3′, HDV1p probe: 5′-/56-FAM/CGC CTC GGT/ZEN/CTC CTC TAA-3′), or HDV-3 (HDV3-fw: 5′-GGT CCG TCG TTC CAT CCT TT-3′, HDV3-rv: 5′-GTA GCT CCC TCG GAT CGT TG-3′, HDV3 probe: 5′-/56-FAM/CTT ACC TCG TGG CCG GC/3BHQ_1/-3′). Primers and probes were designed to specifically detect the used HDV isolates and did not detect the respective other isolates, e.g., HDV-1 primers did not allow amplification of HDV-3 RNA, and HDV-3 primers did not detect HDV-1 RNA. HBV viremia and intrahepatic HBV pregenomic (pg) RNA levels were determined by qRT-PCR using specific primers and probe (Taqman Gene Expression Assay Pa03453406_s1, Applied Biosystems) (Malmstrom et al., 2012) under conditions previously described (Giersch et al., 2019). HBsAg quantification were performed on the Abbott Alinity I platforms (quantitative HBsAg kit, Abbott, Ireland, Diagnostic Division) after diluting the mouse serum 1:200 in the dilution serum (Abbott) as recommended by manufacturer. Alanine aminotransferase (ALT) was measured by using the Roche Cobas c111 System (Roche, Basel, Switzerland). For the measurements 5 μl of mouse serum was used. In HBV/HDV co-infection experiments, known amounts of an HDV containing plasmid were used as standard for serum HDV quantification. In all super-infection experiments, standard curves for HDV quantification were generated by extracting RNA from transfected HuH7 cell culture supernatant, which contained either HDV-1cc, HDV-, or HDV-3cc virions, using the QiAmp MinElute Virus Spin kit (Qiagen, Hilden, Germany). Residual plasmids were removed from the extracts by DNAse I (Epicentre/Lucigen, Madison, WI, United States) digestion followed by isopropanol precipitation. The HDV RNA standard concentration was then determined by qRT-PCR using HDV specific primers and probe (Ferns et al., 2011) and known amounts of an HDV containing plasmid.

The HDV RNA standard curves for HDV-1 (all isolates), and HDV-3cc were prepared in 1:5 dilutions and can be found in **Supplementary Figure 1**. Known amounts of an HBV-containing plasmid was used as standard for serum HBV quantification in all experiments (Lutgehetmann et al., 2012). Steady-state levels of intracellular viral RNA amounts were normalized to human specific hGAPDH (Taqman Gene Expression Assay Hs99999905_m1, Applied Biosystems).

For genomic and antigenomic HDV RNA quantification, intracellular RNA was reverse transcribed using biotinylated genomic and antigenomic HDV RNA primers (Ferns et al., 2011) as previously described (Giersch et al., 2019). Biotinylated cDNA was purified with the MinElute PCR Purification Kit (Qiagen, Venlo, Netherlands) and isolated with dynabeads specifically interacting with biotin (Dynal Kilobase Binder Kit, Invitrogen, Carlsbad, CA, United States) following the manufacturer's instructions. For qRT-PCR with purified biotinylated cDNA bound to dynabeads, HDV-specific primers and probes (Ferns et al., 2011) and the ABI Fast Advanced Master (Applied Biosystems) were used under the conditions described above.

cccDNA Quantification

Total DNA was extracted from liver specimens using the Master Pure DNA purification kit (Epicentre/Lucigen) following the manufacturer's instructions, which include a proteinase K digestion step. To remove rcDNA 1 μg of extracted liver DNA was digested with plasmid-safe ATP-dependent DNase (Epicentre/Lucigen) (30 U) at 37°C for 2 h. After heat inactivation (30 min at 70°C) and isopropanol precipitation, qRT-PCR with cccDNA-selective primers and probe (Malmstrom et al., 2012) (final concentration forward primer: 100 nmol/l, reverse primer: 800 nmol/l) was performed under the following conditions: 10 min initial denaturation at 95°C; 40 cycles: 1 s at 95°C, 1 min at 65°C). cccDNA copies were normalized to the number of human hepatocytes, which was estimated by measuring the single copy gene human hemoglobin beta (Taqman Gene Expression Assay Hs00758889_s1, Applied Biosystems). Known amounts of cccDNA containing plasmid and human genomic DNA (Roche Applied Science, Mannheim, Germany) were used as a standard for quantification.

Sequencing

For HDV genome sequencing of the amber/W site (position 1012) cDNA was synthesized with the Transcriptor First Strand cDNA Synthesis Kit (Roche, Mannheim, Germany) using random hexamer primer according to the manufacturer's instructions. HDV region R1 [location: 305–1285, product size: 980 bp, according to the numeration of Wang et al. (1986)) were amplified by PCR using cDNA, respective primers (R1 fw 305-327: CCA GAG GAC CCC TTC AGC GAA C, R1 rv 1285-1261: GAA GGA AGG CCC TCG AGA ACA AGA) (Ivaniushina et al., 2001] and a Red-Taq Polymerase (Sigma-Aldrich, St. Louis, United States) under the following conditions: 4 min initial denaturation at 94°C; 45 cycles: 1 min at 94°C, 1 min at 62°C, and 2 min at 72°C. PCR product length was analyzed on a 0.8% agarose gel and DNA fragments were extracted with the MinElute PCR Purification Kit (Qiagen) as recommended by the manufacturer. The forward and reverse strand was sequenced with Sanger sequencing (Mix2seq kit) performed by Eurofins Genomics (Ebersberg, Germany).

RNA Library Preparation/Next Generation Sequencing (NGS)

RNA Illumina NGS libraries were prepared from each sample using SMARTer Stranded Total RNA-Seq Kit v2 - Pico Input Mammalian (Takara Bio Europe, Saint-Germain-en-Laye, France). All libraries were multiplex-sequenced on an Illumina MiSeq instrument (300 cycles, PE protocol). Total numbers of 3,249,720 and 3,193,042 paired-end 151-nucleotide (nt) reads for the samples HDV1/3-liver, and HDV1/1-liver were generated. Adapter sequences of the reads and bases with a score of less than Q30 were trimmed, and any reads shorter than 40 nt removed using Trimmomatic v0.36 (Bolger et al., 2014). The high-quality paired-end reads were mapped to the host genome (10 mm, GCA_000001635.2) using the Bowtie2 package (Langmead and Salzberg, 2012). The unaligned reads were converted to fastq files using the SamTofastq tool in Picard tools. The tool FLASH (Magoc and Salzberg, 2011) was used to stitch the paired-end reads since over 70 percent of the reads overlap the reads generated from the opposite end of the same DNA fragment. The resulting stitched sequences together with the remaining paired-end reads were used as input data for the SPAdes assembler (v3.13.0) (Bankevich et al., 2012).

Expression of Human Specific Interferon Stimulated Genes (ISGs)

To determine intrahepatic expression levels of human specific interferon stimulated genes (ISGs) in USG mice, liver RNA was extracted as described above and cDNA was synthesized with the Transcriptor First Strand cDNA Synthesis Kit (Roche, Mannheim, Germany) using oligo-dT primer according to the manufacturer's instructions. qRT-PCR was performed with the ABI Fast Advanced Master (Applied Biosystems) in an ABI Viia7 (Applied Biosystems). The following Taqman Gene Expression Assays from Applied Biosystems containing human specific primers and probe, which do not cross-react with murine signals, were used: hMxA (Hs00895608_m1), hISG15 (Hs00192713_m1), hCXCL10 (Hs00171042_m1), hOAS1 (Hs00973637_m1), hHLA-E (Hs03045171_m1, and hCASP8 (Hs01018151_m1). The human housekeeping genes hGAPDH (Hs99999905_m1) and hRPL30 (Hs00265497_m1) were used for normalization.

Immunofluorescence Stainings

Cryostat sections of chimeric mouse livers were stained as previously described (Lutgehetmann et al., 2012). Briefly, sections were fixed with acetone and incubated with mouse anti-CK18 (1:400, Dako, Glostrup, Denmark), rabbit anti-HBcAg (1:2000, Dako), mouse HLA-ABC (1:50, Antibodies-online, Aachen, Germany), and human anti-Delta (anti-HDAg-positive human serum, 1:8,000). Specific signals were visualized with Alexa 488-, 555-, or 633-labeled secondary antibodies (Invitrogen, Darmstadt, Germany). To enhance the HBcAg staining an anti-rabbit horseradish-peroxidase conjugated secondary antibody

(Jackson Immunoresearch, Suffolk, United Kingdom) and the TSA Fluorescein System (Perkin Elmer, Jügesheim, Germany) were used. Nuclear staining was achieved by Hoechst 33258 (1:20,000 diluted, Invitrogen, Waltham, MA, United States). Stained sections were then mounted with fluorescent mounting media (Dako) and analyzed with the fluorescence microscope BZ8710 (Keyence, Osaka, Japan) using the same settings for the different experimental groups. The percentages of HDAg-positive human hepatocytes were estimated as previously described (Lutgehetmann et al., 2012) and by using 2–5 visual fields (displaying an average of 500 human hepatocytes) per mouse liver.

RNA *in situ* Hybridization (RNAScope)

RNA *in situ* hybridization was performed on paraformaldehyde-fixed, cryo-preserved liver sections using the RNAScope Fluorescent Multiplex Kit (Advanced Cell Diagnostics, ACD, Hayward, CA, United States) according to the manufacturer's instructions and as previously described (Allweiss et al., 2016). Briefly, liver sections were fixed with 4% paraformaldehyde, dehydrated with ethanol and pretreated with Pretreat 4 (Pretreatment Kit, ACD) for 30 min. Liver sections were then incubated with RNAScope target probes, which specifically bind HBV pgRNA (assay number: 442741), HDV-1 (475311) or HDV-3 (478121-C2) AG HDV RNA, hISG15 (450921-C3), or hMxA (403831-C3) for 2 h at 40°C (HybEZ oven, ACD). DAPI staining was performed to visualize nuclei. Stained sections were analyzed by fluorescence microscopy (Biorevo BZ-9000, Keyence) using a 60 × /1.40 NA oil objective. Merged z stack images were prepared using the same settings for all groups. The percentages of HDV RNA positive human hepatocytes were estimated by using 3 visual fields (displaying an average of 500 human hepatocytes) per mouse liver.

Western Blot

Protein lysates were obtained by extracting liver tissue with T-Per Tissue Protein Extraction Reagent (Pierce, Rockford, United States) supplemented with NaF (10 mM), EDTA (2 mM), benzamidine (10 mM), PMSF (1 mM), leupeptin (1 μg/ml), Na3VO4 (2 mM), and aprotinin (1,5 μg/ml). Protein content was measured by Pierce BCA Protein Assay Kit (Thermo Scientific, Rockford, United States) following the manufacturer's instructions. For immunoblotting, 20 μg of protein extracts were denaturated at 95°C, separated on a 12% sodium dodecyl sulfate-polyacrylamide gel (NuView Precast gels, Peqlab, Erlangen, Germany) and blotted onto a nitrocellulose membrane (Hybond ECL Nitrocellulose Membrane, GE Healthcare, Buckinghamshire, United Kingdom). Small and large hepatitis delta antigens (S-HDAg, L-HDAg) were detected using a highly diluted human anti-Delta antibody (anti-HDAg-positive human serum (1:2,000). An anti-human horseradish-peroxidase conjugated secondary antibody (Jackson Immunoresearch, Suffolk, United Kingdom) was used in a dilution of 1:10,000. Signals were visualized with Super Signal West Dura Chemiluminescent Substrate (Pierce) and the Molecular Imager ChemiDoc XRS System (Bio-Rad Laboratories, Hercules, United States).

Statistics

Statistical analyses were performed with the GraphPad Prism 9 software. For group-wise comparisons, the non-parametric Mann-Whitney *U* test was applied. For correlations the Spearman test was used. P values < 0.05 were considered statistically significant.

RESULTS

HDV-3cc Infected Mice Show Higher Intrahepatic Amounts of HDV RNAs and HDAg Than Animals Infected With a HDV-1cc Strain

To comparatively assess infection and replication capacities of both cell culture (cc)-derived HDV strains, humanized USG mice were infected with HBV and either HDV-3cc (n = 7) or HDV-1cc (n = 5) and viral parameters were compared (**Figure 1A**). HDV viremia increased during the first 7 weeks of infection and reached stable levels of median 3×10^8 HDV RNA copies/ml in HBV/HDV-3cc co-infection and of median 8×10^7 HDV RNA copies/ml in HBV/HDV-1cc co-infected mice (**Figure 1B**). Nine weeks post infection (p.i.) median HDV levels were slightly higher in serum (0.4log, p = 0.0303) (**Figure 1B**) and clearly increased in the liver (1.1-log, p = 0.0025) of HBV/HDV-3cc-infected mice, compared to HBV/HDV-1cc-infected animals (**Figure 1C**). Consistent with the increased intrahepatic HDV RNA levels, the amount of HDAg positive human hepatocytes appeared higher in HBV/HDV-3cc co-infected mice (approximately 70 vs. 20% in HDV-1cc co-infection) (immunofluorescence staining, **Figures 1D,E**). Moreover, a clearly higher number of human cells was positive for genomic (G) and antigenomic (AG) HDV RNA in HDV-3cc infection compared to HDV-1cc infection (RNA *in situ* hybridization, **Figures 1F,G**). The increased levels of G and AG HDV RNA in HDV-3cc-infected mouse livers were also confirmed by performing a specific qRT-PCR assay using biotinylated reverse transcription primers and Dynabeads (see section "Materials and Methods"; **Supplementary Figure 2A**). RNA Sanger sequencing of the amber/W site revealed that the HDV RNA encoding for the S-HDAg (ATC) tends to be dominant in serum of HDV-3cc-infected mice, while human hepatocytes infected with HDV-1cc preferentially generated HDV RNA encoding for the L-HDAg (ACC) (**Supplementary Figure 2B**). However, the protein expression of intrahepatic S- and L-HDAg determined by western blot was very similar in HBV/HDV-1cc- and HBV/HDV-3cc-infected animals (**Supplementary Figure 2C**). The slightly different migration properties of HDV-1 and HDV-3 HDAgs likely occur due to different amino acid sequences and were also observed in an *in vitro* study comparing these two genotypes (Casey and Gerin, 1998). In line with previous findings (Lutgehetmann et al., 2012), the increased efficacy of HDV-3cc to infect humanized livers was associated with certain levels of HBV suppression, as indicated by the lower HBV viremia (1.0-log, p = 0.0051, **Figure 2A**) and intrahepatic HBV

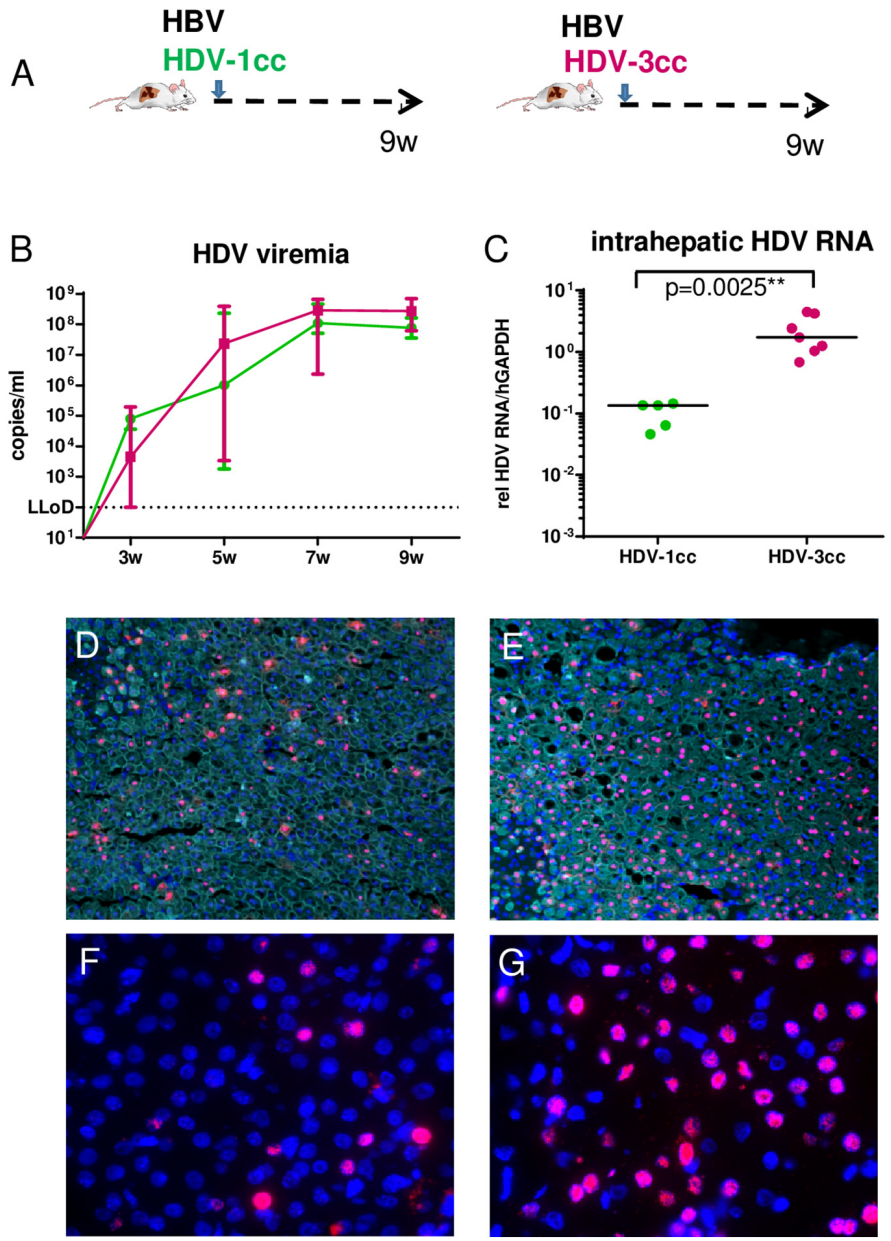

FIGURE 1 | HDV in mice co-infected with HBV and HDV-1cc or HDV-3cc. **(A)** Experimental setting. qRT-PCR measurements (HDV primers/probe recognize all genotypes) of serum HDV RNA (quantification with plasmid standard) **(B)** and liver HDV RNA (normalized to housekeeping gene hGAPDH) **(C)** in HBV/HDV-1cc (green) and HBV/HDV-3cc (red) co-infected mice 9 weeks post infection. Results are expressed as median ± range **(B)**, the bar shows median levels **(C)**. $^{**}p < 0.01$. Immunofluorescence staining of HDAg (red) and CK18 (detecting human hepatocytes, aqua) in HBV/HDV-1cc **(D)** and HBV/HDV-3cc **(E)** co-infected mice at the end of the experiment. Nuclei are stained with Hoechst 33258 (blue). RNA *in situ* hybridization staining of AG HDV RNA (red) in HBV/HDV-1cc **(F)** and HBV/HDV-3cc **(G)** co-infected mice at the end of the experiment. Nuclei are stained with dapi (blue).

pgRNA levels (0.6-log, $p = 0.0480$, **Figure 2B**) determined in mice co-infected with HDV-3, compared to HBV-mono-infected and HBV/HDV-1cc co-infected mice. The development of HBV viremia was also slightly delayed in HBV/HDV-1 co-infected mice. However, intrahepatic HDV RNA levels remained lower in these mice and liver HBV pgRNA levels appeared comparable in HBV mono-infected and HBV/HDV-1 infected mice at 9 weeks p.i. (**Figures 2A,B**). Likewise, RNA *in situ* hybridization staining demonstrated substantially lower amounts of HBV pgRNA in livers of HBV/HDV-3cc-infected than in HBV/HDV-1cc-infected animals (**Figures 2C,D**). Circulating HBsAg levels (**Figure 2E**) and intrahepatic amounts of covalently closed circular (ccc) DNA (**Figure 2F**) did not differ significantly between HDV-1 and HDV-3 infected animals, although levels appeared lower in mice receiving HDV-3, further indicating HDV-mediated suppression of HBV. Furthermore, we investigated whether

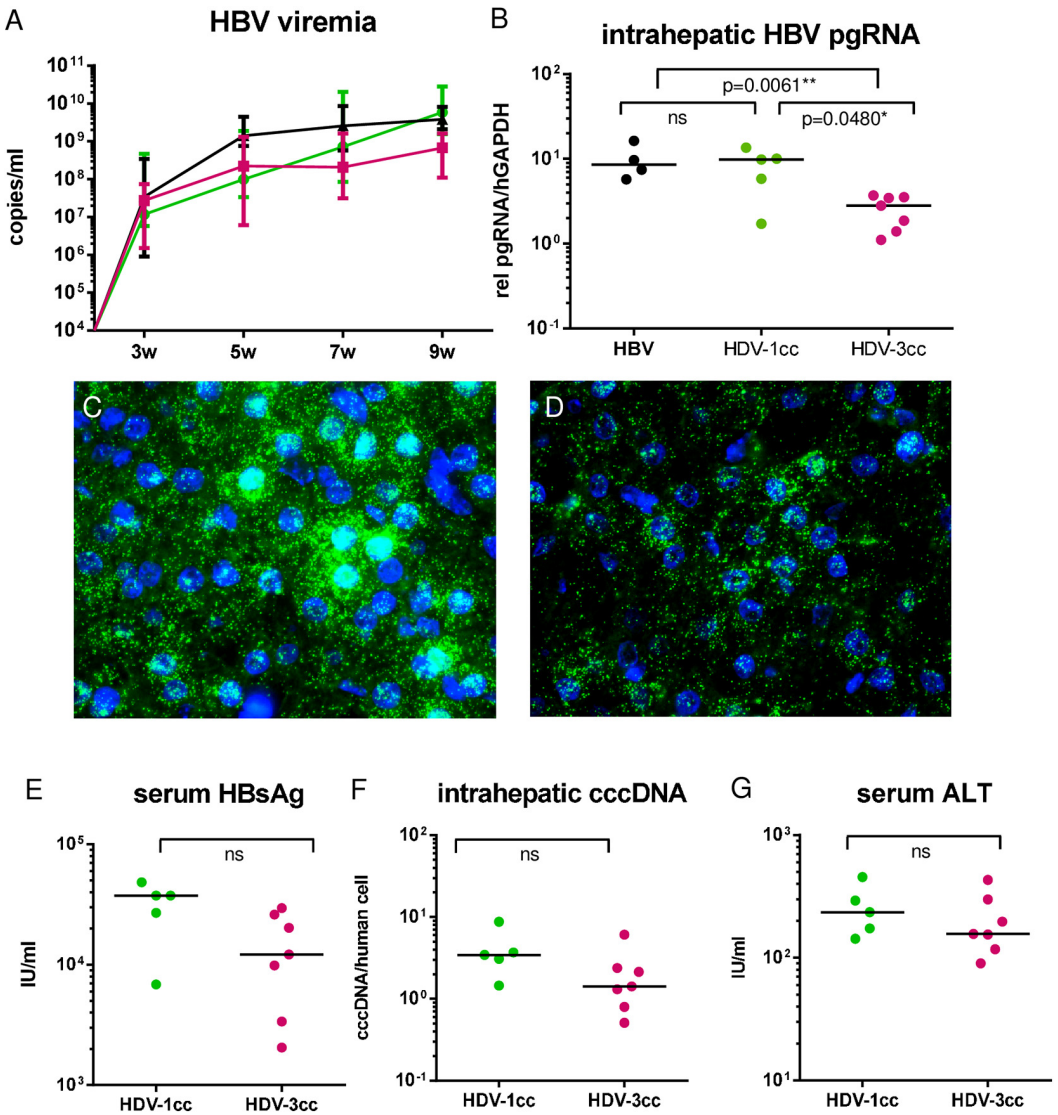

FIGURE 2 | HBV in mice co-infected with HBV and HDV-1cc or HDV-3cc. qRT-PCR measurements of serum HBV DNA (quantification with plasmid standard) **(A)** and liver pregenomic HBV RNA (normalized to housekeeping gene hGAPDH) **(B)** in HBV/HDV-1cc (green line or dots) and HBV/HDV-3cc (red line or dots) co-infected mice 9 weeks post infection compared to stable HBV mono-infected mice (black line or dots). Results are expressed as median ± range **(A)**, the bar shows median levels **(B)**. $^{*}p < 0.05$, $^{**}p < 0.01$. RNA *in situ* hybridization (RNAScope) staining of pregenomic HBV RNA (green) in HBV/HDV-1cc **(C)** and HBV/HDV-3cc **(D)** co-infected mice at the end of the experiment. Nuclei are stained with dapi (blue). Serum HBsAg levels **(E)**, intrahepatic cccDNA levels (normalized to human beta globin) **(F)** and serum ALT levels **(G)** in HBV/HDV-1cc and HBV/HDV-3cc co-infected mice at the end of the experiment. The bar shows median levels **(E–G)**.

HDV-1cc and HDV-3cc infection leads to different cytotoxic effects on human hepatocytes of USG mice. In livers of both HDV-1cc and HDV-3cc-infected mice the apoptosis marker human caspase 8 was not significantly induced compared to uninfected controls (**Supplementary Figure 3A**, qRT-PCR). Immunofluorescence staining demonstrated very low and similar numbers of caspase 3 positive human hepatocytes in both co-infection settings (data not shown). Levels of serum ALT, which increase during liver injury, were comparable in HBV/HDV-3cc and HBV/HDV-1cc-infected mice (**Figure 2G**). In line, human serum albumin (HSA) levels, which reflect the amount of human hepatocytes in chimeric mouse livers, remained stable throughout the course of infection with both HDV-1cc and HDV-3cc (**Supplementary Figure 3B**), indicating that neither HBV/HDV-1cc nor HBV/HDV-3cc co-infections promote detectable cell death in USG mice.

Interestingly, mRNA expression levels of human specific interferon simulated genes (hISGs) such as hMxA, hOAS1 and hHLA-E did not differ significantly between HBV/HDV-3cc and HBV/HDV-1cc co-infected mice, but clearly increased compared to uninfected animals on a global level (qRT-PCR) and on a single cell level (RNA *in situ* hybridization) (**Figure 3**). The expression of hISG15 and hCXCL10 tended to be more enhanced in HDV-3cc-infected mice compared to HDV-1cc-infected animals

(**Figure 3**) and such ISG increase correlated with intrahepatic HDV RNA levels (**Supplementary Figure 4**). Of note, an induction of hISGs was not exclusively observed in HDV AG RNA-positive or neighboring human cells, but also in further distant HDV-negative bystander human hepatocytes (**Figure 3**). Triple staining of HLA-ABC, HBcAg and HDAg indicated that the HDV-negative cells that show an induction of ISGs were either HBcAg-positive or negative for both of these HDV and HBV infection markers (**Supplementary Figure 5**). The underlying mechanisms mediating this rather "global sensing" of the infection needs to be investigated in future studies. Overall, this HDV-3cc strain demonstrated a higher infection efficiency in USG mice, as displayed by increased HDV viremia, intrahepatic HDV RNA levels and increased HBV suppression capacities compared to HDV-1cc, while levels of apoptosis and liver injury markers were comparable between HDV genotypes in human liver chimeric mice that lack functional NK, B and T cells.

HDV-3cc/HDV-1cc Co-infection Occurs in USG Mice but Not at the Single Cell Level

To explore the ability of these two HDV genotypes to co-infect human hepatocytes *in vivo*, stably HBV-infected humanized mice (n = 7, median HBV viremia: 5 × 10^9 copies/ml) were simultaneously infected with HDV-3cc and HDV-1cc (**Figure 4A**). Four weeks p.i., both HDV genotypes were detected in serum and livers of USG mice by using HDV-1 and HDV-3 specific qRT-PCRs, which did not detect the respective other HDV genotype (**Figures 4B,C**). Interestingly, HDV-3cc viremia reached maximum levels already at 4 weeks p.i. and remained stable until the end of the experiment at 10 weeks p.i. (**Figure 4B**). In contrast, HDV-1cc RNA in serum was slightly lower (0.9-log) compared to HDV-3cc levels at 4 weeks p.i., but strongly decreased (more than 3-log) in the following weeks (**Figure 4B**). In line with the viremia levels, intrahepatic HDV-3cc RNA was already at high levels at 4 weeks p.i. and at similar levels in mice sacrificed at 10 weeks p.i., while the HDV-1cc RNA levels within these HBV/HDV-1cc/HDV-3cc co-infected mouse livers at 10 weeks p.i. appeared clearly lower than in livers of mice sacrificed at 4 weeks p.i. (**Figure 4C**). Strain-specific RNA *in situ* hybridization indicated a similar number of HDV-1cc and HDV-3cc AG RNA-positive human hepatocytes at week 4 p.i. in livers of HBV/HDV-1cc/HDV-3cc-infected mice (**Figure 4D**). At 10 weeks p.i. clearly lower amounts of HDV-1cc AG RNA-positive human hepatocytes (1%) were detected compared to 4 weeks p.i. (**Figure 4E**), indicating that HDV-3cc became the predominant strain in HBV/HDV-1cc/HDV-3cc-infected mice (66.7% HDV-3cc AG RNA-positive). Strikingly, AG HDV RNAs of HDV-1cc and HDV-3cc could very rarely (less than 0.1%) be detected in the same hepatocyte at any of the time-points observed (**Figure 4F**), suggesting that these HDV genotypes

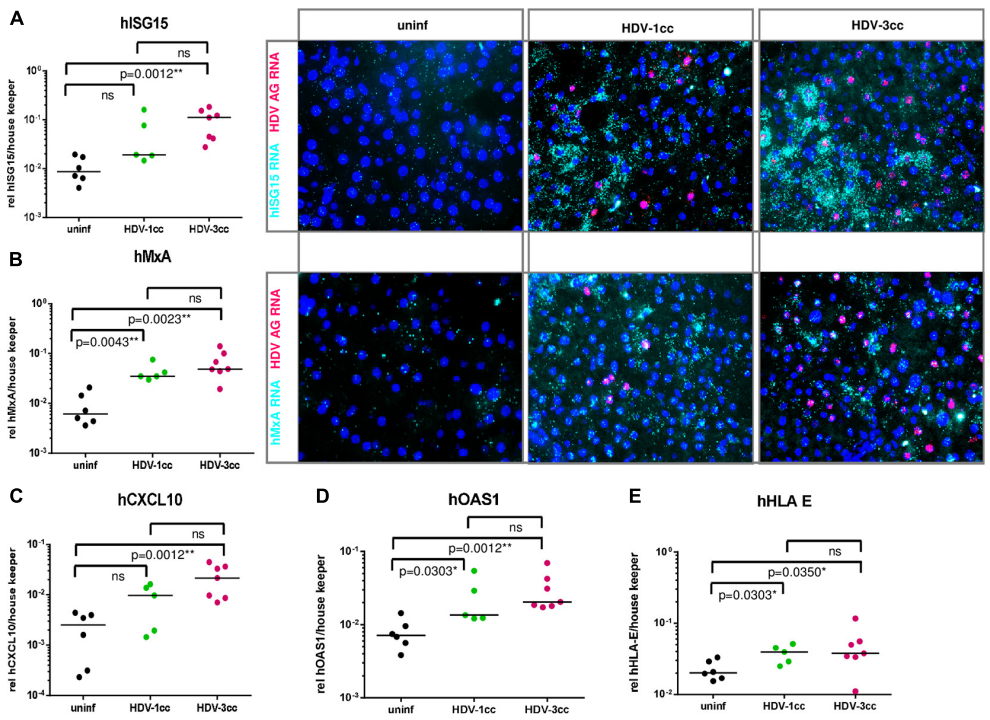

FIGURE 3 | Human specific ISGs and chemokines in mice co-infected with HBV and HDV-1cc or HDV-3cc. **(A–E)** qRT-PCR measurements (normalized to median of housekeeping genes hGAPDH and hRPL30) **(A left, B left, and C–E)** and RNA *in situ* hybridization staining **(A right, B right)** of human ISG15 **(A)**, MxA **(B)**, CXCL10 **(C)**, OAS1 **(D)**, and HLA-E **(E)** in livers of HBV/HDV-1cc and HBV/HDV-3cc co-infected mice 9 weeks post infection compared to uninfected mice. The bar shows median levels. *$p < 0.05$, **$p < 0.01$. RNA *in situ* hybridization stainings show hISG15 RNA and hMxA RNA in aqua and HDV AG RNA in red. Nuclei are stained with dapi (blue).

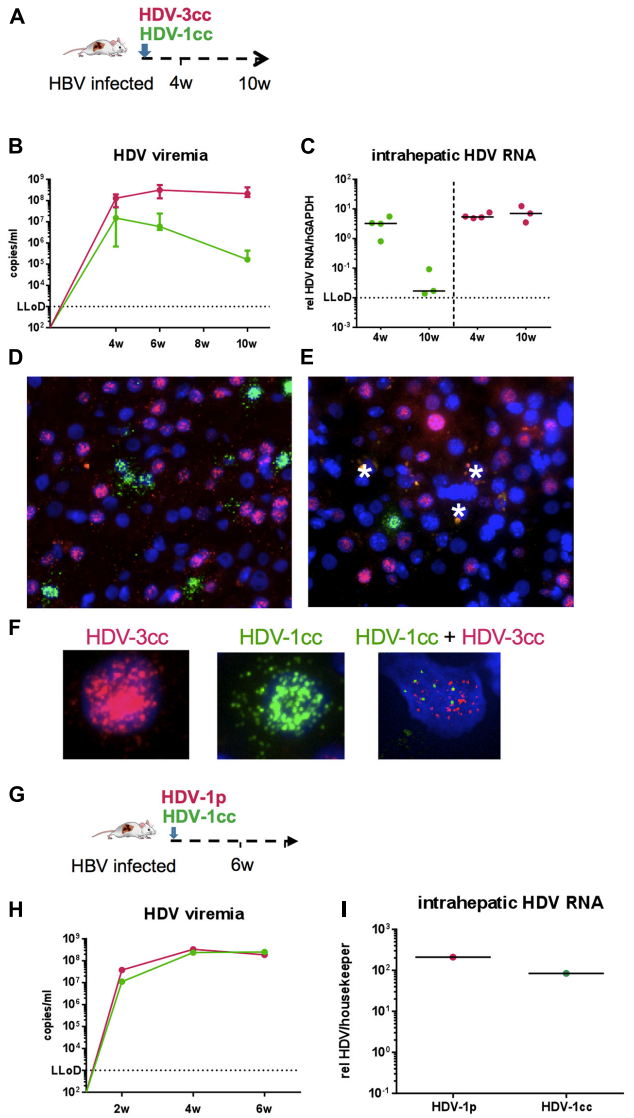

FIGURE 4 | HDV in chronic HBV-infected mice simultaneously super-infected with HDV-1cc and HDV-3cc or with HDV-1cc and HDV-1p. **(A)** Experimental setting of HDV-1/HDV-3 double infection. **(B,C)** qRT-PCR measurements (genotype specific HDV primers/probe for HDV-1 and HDV-3 were used) of serum HDV RNA (quantification with cell culture derived RNA standard) at different time-points **(B)** and liver HDV RNA (normalized to housekeeping gene hGAPDH) 4 and 10 weeks post super-infection with HDV **(C)**. Results are expressed as median ± range **(B)**, the bar shows median levels **(C)**. HDV-1cc RNA levels are depicted in green (line or dots) and HDV-3cc RNA levels in red (line or dots). **(D,E)** RNA *in situ* hybridization (RNAScope) staining of HDV-1cc (green) and HDV-3cc AG HDV RNA (red) in HDV-1cc/HDV-3cc-infected mice 4 weeks **(D)** and 10 weeks **(E)** post super-infection. Phagocytic remnants have an autofluorescence and are marked with a white star. **(F)** Examples of HDV-3cc AG (red, left), HDV-1cc AG (green, middle), and HDV1/3 AG double positive human hepatocytes in mice infected with both HDV genotypes. Nuclei are stained with dapi (blue). **(G)** Experimental setting of HDV-1cc/HDV-1p double infection. **(H,I)** qRT-PCR measurements (strain specific HDV primers/probe for HDV-1cc and HDV-1p were used) of serum HDV RNA (quantification with cell culture derived RNA standard) at different time-points **(H)** and liver HDV RNA (normalized to housekeeping gene hGAPDH) 6 weeks post super-infection with HDV **(I)**. HDV-1cc RNA levels are depicted in green (line or dots) and HDV-1p RNA levels in red (line or dots).

do not replicate efficiently within the same cell. HBV viremia remained stable for the first 4 weeks after simultaneous super-infection with HDV-1cc and HDV-3cc but decreased in the following 6 weeks (0.8-log, $p = 0.0167$), while HBV mono-infected mice maintained stable HBV DNA viremia levels until the end of the experiment (**Supplementary Figure 6A**). Also intrahepatic HBV pgRNA levels were still high - and comparable to stable HBV mono-infected mice - at 4 weeks post HDV super-infection but appeared lower at 10 weeks post HDV super-infection (1.2-log decrease, **Supplementary Figure 6B**).

To comparatively assess the capacity of two different HDV strains with less divergent genome sequences to infect the livers of humanized mice, we super-infected one stable HBV-infected mice with HDV-1cc and another HDV-1 isolate derived from a chronic HBV/HDV infected patient (HDV-1p) (**Figure 4G**). 6 weeks after simultaneous HDV-1cc/HDV-1p super-infection, high and comparable levels of HDV RNA were measured in serum and liver by employing a strain specific qRT-PCR (which did not cross-react with the respective other HDV-1 isolate) (**Figures 3H,I**). Due to the high sequence similarity between HDV-1cc and HDV-1p, the RNA *in situ* hybridization assay could not be performed, thus hindering analyses of co-infection events at single-cell level. However, NGS analysis revealed that within the observation time, no recombination events were detected between HDV-1cc and HDV-3cc, or HDV-1cc and HDV1p *in vivo* (data not shown). These data suggest that concomitant infection of the same hepatocyte with two actively replicating HDV strains may not be a common event *in vivo*.

Sequential Super-Infection With HDV-3cc and HDV-1cc (and Vice Versa) Was Strongly Impaired *in vivo*

We also investigated whether mice already chronically infected with HBV and HDV could be super-infected by another HDV genotype. We used stably HBV-infected USG mice and sequentially super-infected them with HDV-1cc for 9 weeks (first HDV super-infection) and then subsequently with HDV-3cc for additional 9 weeks (second HDV super-infection) ($n = 2$) (**Figure 5A**). After infection with HDV-1cc, mice developed viremia levels of around 1×10^8 HDV-1 RNA copies/ml, which remained stable after super-infection with the second HDV strain (HDV-3cc) (**Figure 5B**). Surprisingly, genotype specific qRT-PCR showed that serum HDV-3cc RNA levels remained around the lower limit of detection (LLoD = 1×10^3 copies/ml) for the entire observation time (**Figure 5B**). In contrast, control mice that had been infected only with HBV and super-infected with HDV-3cc ($n = 2$), developed as expected HDV-3cc viremia up to levels of 6×10^7 copies/ml within 9 weeks (**Figure 5B**). In line with the HDV viremia, intrahepatic HDV-3cc RNA levels were clearly lower or below quantitative detection in mice that were previously infected with HDV-1cc, compared to mice that were only super-infected with HDV-3cc (**Figure 5C**). Using RNA *in situ* hybridization, we detected HDV-1cc AG RNA in approximately 20% of human hepatocytes at the end of experiment, while we were not able to detect HDV-3cc AG RNA-positive human hepatocytes in livers of these

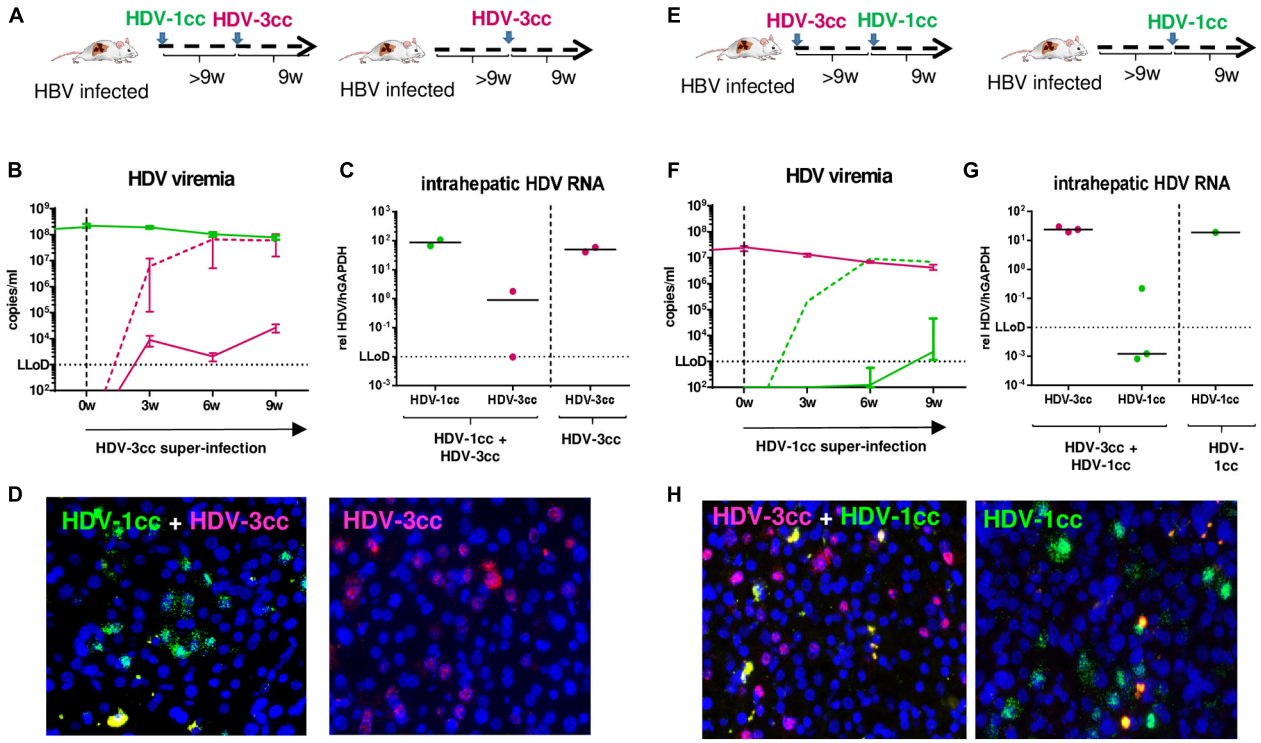

FIGURE 5 | HDV in chronic HBV-infected mice sequentially super-infected with HDV-1cc and HDV-3cc. **(A)** Experimental setting (first HDV-1cc, then HDV-3cc; control mice were super-infected with HDV-3cc only). **(B,C)** qRT-PCR measurements (genotype specific HDV primers/probe for HDV-1 and HDV-3 were used) of serum HDV RNA (quantification with cell culture derived RNA standard) at different time-points **(B)** and liver HDV RNA (normalized to housekeeping gene hGAPDH) at the end of experiment **(C)**. HDV-1cc RNA levels are depicted in green (line or dots) and HDV-3cc RNA levels in red (line or dots). The dashed red line shows HDV viremia in control mice (HDV-3cc only). Results are expressed as median ± range **(B)**, the bar shows median levels **(C)**. **(D)** RNA *in situ* hybridization (RNAScope) staining of HDV-1cc (green) and HDV-3cc AG HDV RNA (red) in HDV-1cc/HDV-3cc sequential super-infected (**left**) and HDV-3cc control mice (**right**) at the end of the experiment. Nuclei are stained with dapi (blue). **(E)** Experimental setting (first HDV-3cc, then HDV-1cc; control mice were super-infected with HDV-1cc only). **(F,G)** qRT-PCR measurements (genotype specific HDV primers/probe for HDV-1 and HDV-3 were used) of serum HDV RNA (quantification with cell culture derived RNA standard) at different time-points **(F)** and liver HDV RNA (normalized to housekeeping gene hGAPDH) at the end of experiment **(G)**. HDV-1cc RNA levels are depicted in green (line or dots) and HDV-3cc RNA levels in red (line or dots). The dashed green line shows HDV viremia in control mice (HDV-1cc only). Results are expressed as median ± range **(F)**, the bar shows median levels **(G)**. **(H)** RNA *in situ* hybridization (RNAScope) staining of HDV-1cc (green) and HDV-3cc AG HDV RNA (red) in HDV-3cc/HDV-1cc sequential super-infected (**left**) and HDV-1cc control mice (**right**) at the end of the experiment. Nuclei are stained with dapi (blue).

mice (**Figure 5D**). HBV-infected USG mice, which were super-infected with HDV-3cc alone, showed the presence of HDV-3cc antigenomic RNA-positive liver cells (**Figure 5D**). Conclusively, after an infection with HDV-1cc was established in HBV-infected USG mice, a second super-infection with HDV-3cc appeared strongly impaired.

Similarly, predominance of the HDV genotype that infected mouse livers first was determined also when stable HBV-infected mice were first super-infected with HDV-3cc and 9 weeks later with HDV-1cc ($n = 3$) (**Figures 5E–H**). Even in these mice, serum and liver HDV-1cc RNA levels hardly exceeded the LLoD, while HDV-3cc viremia (HDV-3cc 1×10^7 copies/ml) and intrahepatic levels remained at high levels throughout the course of the experiment (**Figures 5F,G**). As expected, a control mouse that was infected with HBV and HDV-1cc only, developed serum HDV-1cc RNA levels of 8×10^6 copies/ml (**Figure 5F**). In line with the previous experimental setting, HDV-1cc antigenomic RNA-positive human hepatocytes could only be detected in mice that were infected with HBV and HDV-1cc, but not in mice that already displayed a productive HDV-3cc infection (**Figure 5H**).

To exclude that the observed viral interference of HDV-1cc and HDV-3cc occurs exclusively among these two particular HDV strains, we also used stably HBV-infected mice and super-infected them first with a patient derived HDV-1 strain (HDV-1p) and then with HDV-3cc ($n = 3$) (**Figure 6A**). Even in this sequential super-infection setting, the presence of HDV-1p (serum HDV-1p 2×10^5 copies/ml) substantially hindered the establishment of HDV-3cc infection, as demonstrated by the low levels (around the LLoD) of HDV-3cc RNA in serum and liver (**Figures 6B,C**) and by the absence of HDV-3cc antigenomic RNA-positive human hepatocytes in HBV/HDV-1p/HDV-3cc super-infected mice (RNA *in situ* hybridization, **Figure 6D**). HBV-infected control animals, which were super-infected with HDV-3cc ($n = 2$) only, developed HDV-3cc RNA viremia levels of 6×10^7 copies/ml (**Figure 6B**) and showed the presence of HDV-3cc antigenomic RNA-positive liver cells (**Figure 6D**).

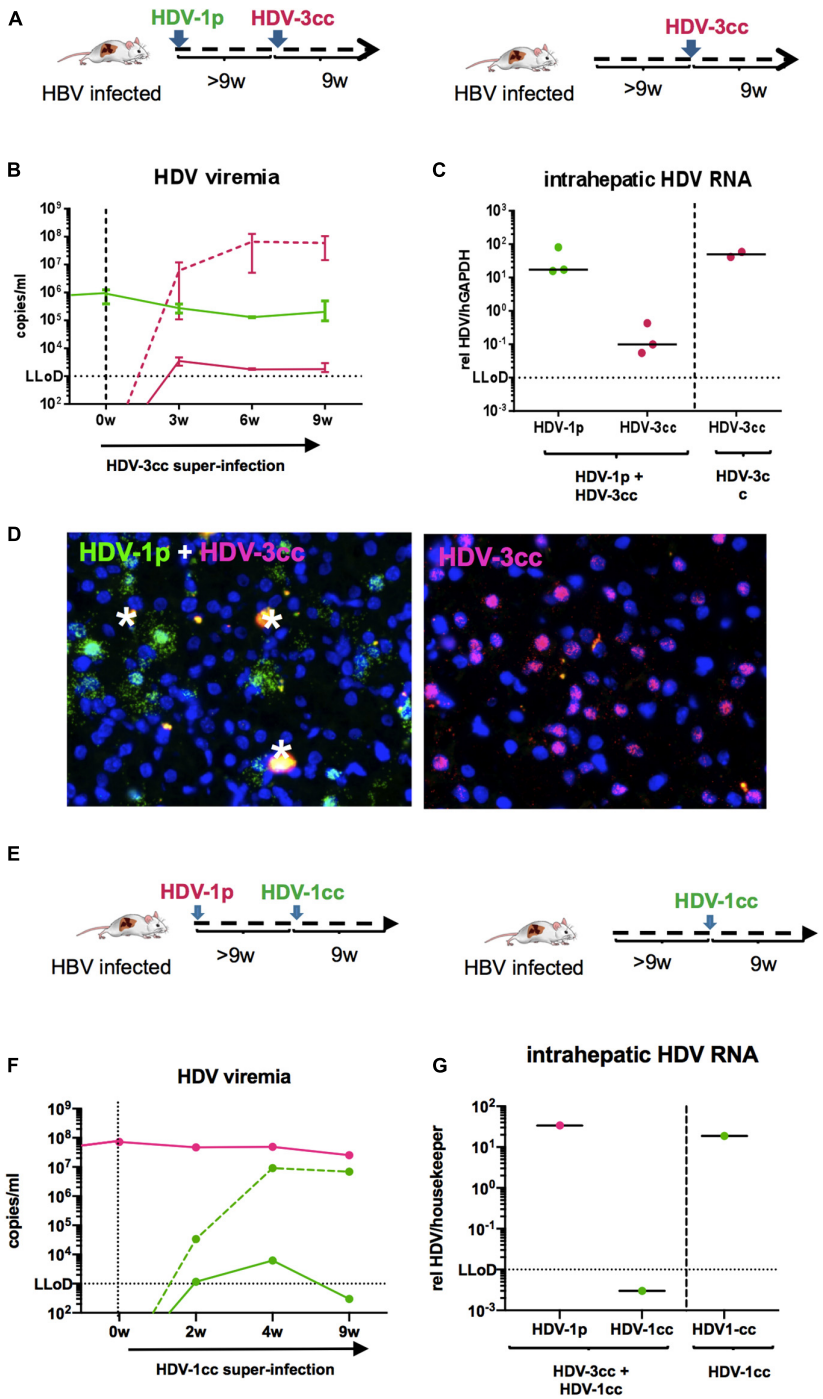

FIGURE 6 | HDV in chronic HBV-infected mice sequentially super-infected with HDV-1p and HDV-3cc, or HDV-1p and HDV-1cc. **(A)** Experimental setting (first HDV-1p, then HDV-3cc; control mice were super-infected with HDV-3cc only). **(B,C)** qRT-PCR measurements (genotype specific HDV primers/probe for HDV-1 and HDV-3 were used) of serum HDV RNA (quantification with cell culture derived RNA standard) at different time-points **(B)** and liver HDV RNA (normalized to housekeeping gene hGAPDH) at the end of experiment **(C)**. HDV-1p RNA levels are depicted in green (line or dots) and HDV-3cc RNA levels in red (line or dots). The dashed red line shows HDV viremia in control mice (HDV-3cc only). Results are expressed as median ± range **(B)**, the bar shows median levels **(C)**. **(D)** RNA *in situ* hybridization (RNAScope) staining of HDV-1p (green) and HDV-3cc AG HDV RNA (red) in HDV-1p/HDV-3cc sequential super-infected (**left**) and HDV-3cc control mice (**right**) at the end of the experiment. Nuclei are stained with dapi (blue). Phagocytic remnants have an autofluorescence and are marked with a white star. **(E)** Experimental setting (first HDV-1p, then HDV-1cc; control mouse was super-infected with HDV-1cc only). **(F,G)** qRT-PCR measurements (strain specific HDV primers/probe for HDV-1cc and HDV-1p were used) of serum HDV RNA (quantification with cell culture derived RNA standard) at different time-points **(F)** and liver HDV RNA (normalized to housekeeping gene hGAPDH) at the end of experiment **(G)**. HDV-1p RNA levels are depicted in red (line or dots) and HDV-1cc RNA levels in green (line or dots). The dashed green line shows HDV viremia in control mouse (HDV-1cc only).

Finally, and in an attempt to investigate whether viral interference mechanisms may hinder the establishment of productive super-infections also among HDV strains of the same genotype, one HBV/HDV-1p infected humanized mouse was super-infected with HDV-1cc (**Figure 6E**). Interestingly, even after 9 weeks of super-infection with the second HDV-1 isolate no serum and liver HDV-1cc RNAs were detected, while HDV-1p viremia remained unchanged, whereas one mouse that was infected in parallel with HBV and HDV-1cc only, - as expected - developed high HDV-1cc serum RNA levels (**Figures 6F,G**).

Overall, our sequential super-infection experiments in USG mice revealed that an already established HDV infection strongly impairs the establishment of a productive subsequent HDV infection not only with a genetically distant genotype, but also with a similar isolate.

DISCUSSION

Hepatitis D Virus is classified into eight different genotypes with a sequence divergence from 15% up to 40% according to partial or full-length genome sequencing (Deny, 2006; Le Gal et al., 2017). Despite the presence of conserved nucleotides, amino acid motifs and positions across all genotypes, indicating the existence of conservatory constraints both at genomic and protein level, inter-genotype differences are also linked to distinct geographical distributions, and different treatment responses and pathogenicity. HDV-1 infection is distributed around the world, shows diverse clinical outcomes and was studied the most *in vitro*, *in vivo,* and in patients. HDV-3 is geographically isolated in the Amazonas region, is associated with the most severe disease outcomes among all HDV genotypes (Casey et al., 1996), but seems to respond better to treatment with PEGylated interferon alpha than HDV-1 (Borzacov et al., 2016). Nevertheless, only very few HDV strains have been studied *in vivo* to date.

Taking advantage of the availability of a commonly used HDV-1 clone (Kuo et al., 1988, 1989) and an HDV-3 clone isolated and generated from an infected Peruvian man with acute severe hepatitis (Casey et al., 1993), in this study we first investigated virological differences between HDV-1 and HDV-3 in primary human hepatocytes co-infected with HBV *in vivo*. By using USG mice that were repopulated with human hepatocytes we observed that HDV-3 establishes higher rates of HDV viremia and intrahepatic HDV RNA levels than this HDV-1 strain in HBV-infected chimeric mice. Despite the use of inocula containing comparable levels of HDV genome equivalents, immunofluorescence staining and RNA *in situ* hybridization revealed that around 70% of human hepatocytes were infected with HDV-3, while only 20% of human cells were HDV-1-positive. In the absence of HBV, HDV-1, and HDV-3 were able to initially infect comparable amounts of human cells *in vitro* (data not shown) and *in vivo* (Giersch et al., 2021), indicating that HDV1 and HDV3 mainly differ in their spreading and assembly capacities. HDV has been observed to suppress HBV replication (Niro et al., 2001; Giersch and Dandri, 2015) and indeed the higher infection levels of HDV-3 determined in this study were accompanied by a stronger suppression of serum and intrahepatic HBV markers compared to HBV/HDV-1 infection.

An *in vitro* study suggested that the severity of hepatic inflammation might be associated with distinct efficacies of HDV genotypes to generate L-HDAg, package and finally secrete viral particles (Hsu et al., 2002). The authors observed that HDV-2 secretes fewer viral particles than HDV-1, which may reduce the extent of liver infection and result in less severe hepatic inflammation (Hsu et al., 2002). While genome sequencing in our study revealed that HDV editing at the amber/W site and thus the production of genomes encoding for the L-HDAg was comparatively lower in HDV-3 infection, the amounts of S- and L-HDAg determined by western blot analysis were similar in HDV-1- and HDV-3-infected animals. However, we observed a higher intracellular productivity of HDV-3, which may explain the faster spreading kinetics and increased numbers of infected cells achieved in the same time frame in comparison to HDV-1 and therefore contribute to the more severe clinical course of HDV-3 infections. Moreover, the degree of apoptosis and liver injury appeared low and comparable with both strains, while the induction of some human specific interferon stimulated genes (hISG15, hCXCL10) tended to be slightly higher in mice harboring HDV-3, suggesting that an enhancement of the intrinsic innate responses may contribute to augment the severity of HDV-3-associated liver disease. However, these mice lack B, T, and NK cells and it is also plausible that the severity of HDV-3 in patients may rather be caused by responses of the adaptive immune system or by so far unknown genotype specific events.

Secondly, we investigated whether HDV-1 and HDV-3 could co-infect the same liver, a setting prone to genetic recombination. To date, co-infections of HDV-1 and HDV-2 were observed *in vitro* and in patients (Wu et al., 1999; Wang and Chao, 2005; Sy et al., 2015), but not between HDV-1 and HDV-3 (Cicero et al., 2016).

Here, we observed for the first time that co-infections of HDV-1 and HDV-3 are possible in human liver chimeric mice already infected with HBV. However, one strain (HDV-3) became dominant within 10 weeks after infection, while viremia and intrahepatic viral loads of the other HDV-1 strain used in these experiments were clearly decreased. Interestingly, even at an early time-point of HDV-1/HDV-3 co-infection, when both HDV genotypes were clearly detected in the liver, we could rarely find human hepatocytes harboring HDV-1 and HDV-3 AG RNA at the same time, indicating that the two genotypes are not able to replicate efficiently within the same cell. Furthermore, we inoculated chronic HBV-infected mice first with HDV-1 and several weeks later with HDV-3 (and vice versa) and demonstrated that when an infection with one HDV genotype is already established in humanized livers, the super-infection with another genotype was strongly impaired. Remarkably, also when one mouse was stably infected with one HDV-1 isolate, a super-infection with another HDV-1 strain could not be established. Taken together, these results indicate the existence of a strong viral inter- and intragenotypic interference and may allow HDV to evade excessive recombination by incoming, divergent viruses, which is believed to occur through RNA template switching of the host RNA polymerase during replication of G HDV RNA (Chao,

2007; Lin et al., 2015). Recombination of different genotypes can occur when at least two viral genomes co-infect the same host cell and is observed among RNA viruses such as hepatitis C virus (HCV), influenza A virus and human immunodeficiency virus (HIV) (Perez-Losada et al., 2015). In general, inter-genomic recombination is more likely to occur in plus-strand viruses (McVean et al., 2002), but was also described for HDV-1 and HDV-2 *in vitro* and in patients (Wang and Chao, 2005; Sy et al., 2015). Lin et al. (2015) observed *in vitro* that intra-genotypic recombination rates (between two different HDV-1 isolates) were higher compared to recombination rates that were obtained between different genotypes (HDV-1 and HDV-2 or HDV-4) and proposed that a decreased genetic distance favors HDV recombination frequency. Although, in this study simultaneous super-infection with two different HDV-1 isolates, led to comparable levels of both HDV-1 strains in serum and liver, NGS analysis of HDV1/3 and HDV1/1 infected mice failed to detect inter- or intra-genotypic recombination within the observed time. These results indicate that simultaneous replication of different HDV genotypes and strains in the same human hepatocyte do not occur frequently *in vivo*, not even in the absence of an adaptive immune system, which could affect HDV spreading capacities. HDV recombination events described in patients may require a longer-term coexistence in the same liver and perhaps even immune mediated selective pressures favoring co-infection and recombination events.

Casey and Gerin (1998) performed co-transfection experiments using HuH7 cells and observed that expression of HDV-3 S-HDAg strongly inhibited HDV-1 RNA replication, while transfection of HDV-1 S-HDAg resulted in a two- to threefold inhibition of HDV-3 RNA replication. The natural function of the L-HDAg is to suppress HDV replication (and support virus assembly) and interestingly, the L-HDAg of HDV-3 was shown to decrease not only HDV-3 but also HDV-1 replication. Vice versa, the L-HDAg of HDV-1 was also able to suppress both HDV-1 and HDV-3 replication, even though the sequence of the C-terminal extension of the HDAg responsible for this inhibitory activity (Chao et al., 1990) is highly different between the two genotypes (Casey and Gerin, 1998). Therefore, it is conceivable that also in our study the presence of S- and L-HDAgs from HDV-1 hindered the establishment and in particular the replication capacities of HDV-3 (and vice versa) in the same human hepatocyte. Future studies are needed to proof whether the inability of HDAg to support replication of another more distant HDV genotype is indeed the main factor responsible for the observed interference between HDV-1 and HDV-3 *in vivo*.

It is striking that even though only about 20% of human hepatocytes were infected with HDV-1, a sequential super-infection with HDV-3 was not achievable in the "remaining" 80% of the human hepatocytes. We showed that the induction of innate immune responses is significantly higher in HBV/HDV co-infected than in HBV mono-infected human liver chimeric mice (Giersch et al., 2015). Interestingly, this study revealed that the expression of human ISGs (i.e., hISG15, hMxA) was not only elevated in HDV-infected cells but also in the surrounding HDV-negative hepatocytes, suggesting that HDV infection induces a broad antiviral state of the host, which may also hinder super-infection with another HDV genotype and even with another HDV isolate of the same genotype. This hypothesis is in line with a publication by Han et al. (2011) who showed that the host antiviral response induced by interferon treatment in primary human hepatocytes, was largely mediated at the level of HDV entry. The same study also observed that hepatocytes already infected with HDV were resistant to a second infection with vesicular stomatitis virus (VSV).

Overall, our experiments in human liver chimeric mice clearly demonstrated that simultaneous infection with the two most distinct HDV genotypes 1 and 3 is possible within the same liver at least for a short period of time, while replication of two different HDV isolates within the same human hepatocyte almost never occurred in the setting of co- or sequential super-infection. The observed viral interference between HDV genotypes would hamper their co-existence in the same individual and thus their recombination. Such co-infection interferences might also contribute to the geographical isolation of HDV-3 and to a further divergent development of these already genetically distinct genotypes.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: EBI, PRJEB44735.

ETHICS STATEMENT

All animal experiments were conducted in accordance with the European Communities Council Directive (86/609/EEC) and were approved by the City of Hamburg, Germany. Written informed consent was obtained from the owners for the participation of their animals in this study.

AUTHOR CONTRIBUTIONS

ML and MD initiated and supervised the study. ML, MD, KG, and LH designed experiments. LA and TV generated chimeric mice. KG, LH, and AV performed analyses and generated data. JH and NF performed NGS analysis. CS provided infectious HDV-1 and HDV-3 cell culture supernatant. JC provided the HDV-3 pCMV3-Peru-1.2 plasmid. KG, LH, and MD wrote the manuscript. LA, TV, AV, ML, CS, and JC discussed the data and corrected the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the German Research Foundation (DFG) by a grant to MD and ML (SFB 841, A8). MD also received funding from the German Center for Infection Research (DZIF-BMBF; TTU-hepatitis 05.816; 05.822; 05.714). All funding

sources supporting the work are acknowledged and authors have nothing to disclose.

ACKNOWLEDGMENTS

We are grateful to R. Reusch and N. Jaeger for excellent assistance with the mouse colony and C. Dettmer and C. Eggers for their great technical help. We thank John Taylor (Philadelphia, PA, United States) for providing the HDV recombinant plasmid pSVL(D3) for HDV-1 and Dieter Glebe (Gießen, Germany) for providing infectious HBV particles.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.671466/full#supplementary-material

Supplementary Figure 1 | HDV RNA amplification and standard curves. HDV RNA was extracted from HuH7 cell culture supernatant, which contained either HDV-1cc, HDV-1p, or HDV-3cc virions, as described in Materials and Methods, and HDV RNA standard curves using primers and probes specific for HDV-1 (all isolates), HDV-1cc, HDV1-p, or HDV-3 were prepared in 1:5 dilutions.

Supplementary Figure 2 | G/AG HDV RNAs, genome sequence and HDAgs in mice co-infected with HBV and HDV-1cc or HDV-3cc. **(A)** qRT-PCR measurements (dynabead based assay, see Materials and Methods) of liver G and AG HDV RNA (normalized to housekeeping gene hGAPDH) in HBV/HDV-1cc (green) and HBV/HDV-3cc (red) co-infected mice 9 weeks post infection. The bar shows median levels. **(B)** Sanger sequencing of genomic HDV RNA in serum of HBV/HDV-1cc and HBV/HDV-3cc co-infected mice 9 weeks post infection. Displayed is the HDV amber/W site. The sequence ATG represents a stop codon and thus encodes for the S-HDAg, while RNA editing leads to the sequence ACC, which encodes for the L-HDAg. **(C)** Western blot analysis of intrahepatic S- (24 kDa) and L-HDAg (27 kDa) in two HBV/HDV-3 (left two lanes) and HBV/HDV-1 (right two lanes) infected mice.

Supplementary Figure 3 | Cell death markers in mice co-infected with HBV and HDV-1cc or HDV-3cc. **(A)** qRT-PCR measurement of intrahepatic human caspase 8 (normalized to median of housekeeping genes hGAPDH and hRPL30) in HBV/HDV-1cc and HBV/HDV-3cc co-infected mice 9 weeks post infection compared to uninfected mice. The bar shows median levels. **(B)** Human serum albumin (ELISA) in HBV/HDV-1cc and HBV/HDV-3cc co-infected mice at different time-points of the experiment.

Supplementary Figure 4 | Correlation of liver HDV RNA and hISG mRNA. qRT-PCR measurements of liver HDV RNA (normalized to housekeeping gene hGAPDH) and hISG15 or hCXCL10 (normalized to median of housekeeping genes hGAPDH and hRPL30) in HBV/HDV-3cc (red dots) and HBV/HDV-1cc infected mice (green dots) 9 weeks post infection. Correlations between HDV RNA and hISG15 or hCXCL10 show a p-value of $p = 0.0323^*$ (spearman r = 0.63) and $p = 0.0591$ (spearman r = 0.57), respectively, when HBV/HDV-1cc and HBV/HDV-3cc infected mice are analyzed together.

Supplementary Figure 5 | Intrahepatic HLA ABC expression. Immunofluorescence staining of human HLA ABC (red), HBcAg (green), HDAg (turquoise) and overlays of HLA ABC with HBcAg or with HBcAg and HDAg in HBV/HDV-1cc (left) and HBV/HDV-3cc (right) co-infected mice at the end of the experiment. Nuclei are stained with Hoechst 33258 (blue).

Supplementary Figure 6 | HBV in chronic HBV-infected mice simultaneously super-infected with HDV-1cc and HDV-3cc. qRT-PCR measurements of serum HBV DNA (quantification with plasmid standard) **(A)** and liver pregenomic HBV RNA (normalized to housekeeping gene hGAPDH) **(B)** in chronic HBV-infected mice simultaneously super-infected with HDV-1cc and HDV-3cc (red line or dots) compared to stable HBV mono-infected mice (black line or dots) at indicated time-points post HDV super-infection. Results are expressed as median ± range **(A)**, the bar shows median levels **(B)**.

REFERENCES

Allweiss, L., Gass, S., Giersch, K., Groth, A., Kah, J., Volz, T., et al. (2016). Human liver chimeric mice as a new model of chronic hepatitis E virus infection and preclinical drug evaluation. *J. Hepatol.* 64, 1033–1040. doi: 10.1016/j.jhep.2016.01.011

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Borzacov, L. M., de Figueiredo Nicolete, L. D., Souza, L. F., Dos Santos, A. O., Vieira, D. S., and Salcedo, J. M. (2016). Treatment of hepatitis delta virus genotype 3 infection with peg-interferon and entecavir. *Int. J. Infect. Dis.* 46, 82–88. doi: 10.1016/j.ijid.2016.03.017

Botelho-Souza, L. F., Vasconcelos, M. P. A., Dos Santos, A. O., Salcedo, J. M. V., and Vieira, D. S. (2017). Hepatitis delta: virological and clinical aspects. *Virol. J.* 14:177.

Cappy, P., Lucas, Q., Kankarafou, N., Sureau, C., and Laperche, S. (2020). No evidence of HCV-assisted HDV propagation in a large cohort of hepatitis C positive blood donors. *J. Infect. Dis.* 223, 1376–1380. doi: 10.1093/infdis/jiaa517

Casey, J. L., and Gerin, J. L. (1998). Genotype-specific complementation of hepatitis delta virus RNA replication by hepatitis delta antigen. *J. Virol.* 72, 2806–2814. doi: 10.1128/jvi.72.4.2806-2814.1998

Casey, J. L., Brown, T. L., Colan, E. J., Wignall, F. S., and Gerin, J. L. (1993). A genotype of hepatitis D virus that occurs in northern South America. *Proc. Natl. Acad. Sci. U.S.A.* 90, 9016–9020. doi: 10.1073/pnas.90.19.9016

Casey, J. L., Niro, G. A., Engle, R. E., Vega, A., Gomez, H., McCarthy, M., et al. (1996). Hepatitis B virus (HBV)/hepatitis D virus (HDV) coinfection in outbreaks of acute hepatitis in the Peruvian Amazon basin: the roles of HDV genotype III and HBV genotype F. *J. Infect. Dis.* 174, 920–926. doi: 10.1093/infdis/174.5.920

Chang, F. L., Chen, P. J., Tu, S. J., Wang, C. J., and Chen, D. S. (1991). The large form of hepatitis delta antigen is crucial for assembly of hepatitis delta virus. *Proc. Natl. Acad. Sci. U.S.A.* 88, 8490–8494. doi: 10.1073/pnas.88.19.8490

Chao, M. (2007). RNA recombination in hepatitis delta virus: implications regarding the abilities of mammalian RNA polymerases. *Virus Res.* 127, 208–215. doi: 10.1016/j.virusres.2007.01.003

Chao, M., Hsieh, S. Y., and Taylor, J. (1990). Role of two forms of hepatitis delta virus antigen: evidence for a mechanism of self-limiting genome replication. *J. Virol.* 64, 5066–5069. doi: 10.1128/jvi.64.10.5066-5069.1990

Chemin, I., Pujol, F. H., Scholtes, C., Loureiro, C. L., Amirache, F., Levrero, M., et al. (2020). Preliminary evidence for hepatitis delta virus exposure in patients who are apparently not infected with hepatitis B virus. *Hepatology*. 73, 861–864. doi: 10.1002/hep.31453

Chen, H. Y., Shen, D. T., Ji, D. Z., Han, P. C., Zhang, W. M., Ma, J. F., et al. (2018). Prevalence and burden of hepatitis D virus infection in the global population: a systematic review and meta-analysis. *Gut*. 68, 512–521. doi: 10.1136/gutjnl-2018-316601

Cicero, M. F., Pena, N. M., Santana, L. C., Arnold, R., Azevedo, R. G., Leal, E. S., et al. (2016). Is hepatitis delta infections important in Brazil? *BMC Infect. Dis.* 16:525. doi: 10.1186/s12879-016-1856-9

Dandri, M., Murray, J. M., Lutgehetmann, M., Volz, T., Lohse, A. W., and Petersen, J. (2008). Virion half-life in chronic hepatitis B infection is strongly correlated with levels of viremia. *Hepatology* 48, 1079–1086. doi: 10.1002/hep.22469

Delfino, C. M., Cerrudo, C. S., Biglione, M., Oubina, J. R., Ghiringhelli, P. D., and Mathet, V. L. (2018). A comprehensive bioinformatic analysis of hepatitis

D virus full-length genomes. *J. Viral Hepat.* 25, 860–869. doi: 10.1111/jvh.12876

Deny, P. (2006). Hepatitis delta virus genetic variability: from genotypes I, II, III to eight major clades? *Curr. Top. Microbiol. Immunol.* 307, 151–171. doi: 10.1007/3-540-29802-9_8

Ferns, R. B., Nastouli, E., and Garson, J. A. (2011). Quantitation of hepatitis delta virus using a single-step internally controlled real-time RT-qPCR and a full-length genomic RNA calibration standard. *J. Virol. Methods* 179, 189–194. doi: 10.1016/j.jviromet.2011.11.001

Freitas, N., Cunha, C., Menne, S., and Gudima, S. O. (2014). Envelope proteins derived from naturally integrated hepatitis B virus DNA support assembly and release of infectious hepatitis delta virus particles. *J. Virol.* 88, 5742–5754. doi: 10.1128/jvi.00430-14

Giersch, K., Allweiss, L., Volz, T., Helbig, M., Bierwolf, J., Lohse, A. W., et al. (2015). Hepatitis Delta co-infection in humanized mice leads to pronounced induction of innate immune responses in comparison to HBV mono-infection. *J. Hepatol.* 63, 346–353. doi: 10.1016/j.jhep.2015.03.011

Giersch, K., and Dandri, M. (2015). Hepatitis B and Delta Virus: advances on studies about interactions between the two viruses and the infected hepatocyte. *J. Clin. Transl. Hepatol.* 3, 220–229. doi: 10.14218/jcth.2015.00018

Giersch, K., Bhadra, O. D., Volz, T., Allweiss, L., Riecken, K., Fehse, B., et al. (2019). Hepatitis delta virus persists during liver regeneration and is amplified through cell division both in vitro and in vivo. *Gut* 68, 150–157. doi: 10.1136/gutjnl-2017-314713

Giersch, K., Hermanussen, L., Volz, T., Kah, J., Allweiss, L., Casey, J., et al. (2021). Murine hepatocytes do not support persistence of Hepatitis D virus mono-infection in vivo. *Liver Int.* 41, 410–419. doi: 10.1111/liv.14677

Gudima, S., He, Y., Meier, A., Chang, J., Chen, R., Jarnik, M., et al. (2007). Assembly of hepatitis delta virus: particle characterization, including the ability to infect primary human hepatocytes. *J. Virol.* 81, 3608–3617. doi: 10.1128/jvi.02277-06

Han, Z., Nogusa, S., Nicolas, E., Balachandran, S., and Taylor, J. (2011). Interferon impedes an early step of hepatitis delta virus infection. *PLoS One* 6:e22415. doi: 10.1371/journal.pone.0022415

Hsu, S. C., Syu, W. J., Sheen, I. J., Liu, H. T., Jeng, K. S., and Wu, J. C. (2002). Varied assembly and RNA editing efficiencies between genotypes I and II hepatitis D virus and their implications. *Hepatology* 35, 665–672. doi: 10.1053/jhep.2002.31777

Ivaniushina, V., Radjef, N., Alexeeva, M., Gault, E., Semenov, S., Salhi, M., et al. (2001). Hepatitis delta virus genotypes I and II cocirculate in an endemic area of Yakutia, Russia. *J. Gen. Virol.* 82, 2709–2718. doi: 10.1099/0022-1317-82-11-2709

Koh, C., Heller, T., and Glenn, J. S. (2019). Pathogenesis of and new therapies for hepatitis D. *Gastroenterology* 156, 461-476.e1.

Kuo, M. Y., Chao, M., and Taylor, J. (1989). Initiation of replication of the human hepatitis delta virus genome from cloned DNA: role of delta antigen. *J. Virol.* 63, 1945–1950. doi: 10.1128/jvi.63.5.1945-1950.1989

Kuo, M. Y., Goldberg, J., Coates, L., Mason, W., Gerin, J., and Taylor, J. (1988). Molecular cloning of hepatitis delta virus RNA from an infected woodchuck liver: sequence, structure, and applications. *J. Virol.* 62, 1855–1861. doi: 10.1128/jvi.62.6.1855-1861.1988

Lai, M. M. (2005). RNA replication without RNA-dependent RNA polymerase: surprises from hepatitis delta virus. *J. Virol.* 79, 7951–7958. doi: 10.1128/jvi.79.13.7951-7958.2005

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

Le Gal, F., Brichler, S., Drugan, T., Alloui, C., Roulot, D., Pawlotsky, J. M., et al. (2017). Genetic diversity and worldwide distribution of the deltavirus genus: a study of 2,152 clinical strains. *Hepatology* 66, 1826–1841. doi: 10.1002/hep.29574

Le Gal, F., Gault, E., Ripault, M. P., Serpaggi, J., Trinchet, J. C., Gordien, E., et al. (2006). Eighth major clade for hepatitis delta virus. *Emerg. Infect. Dis.* 12, 1447–1450. doi: 10.3201/eid1209.060112

Lin, C. C., Yang, Z. W., Iang, S. B., and Chao, M. (2015). Reduced genetic distance and high replication levels increase the RNA recombination rate of hepatitis delta virus. *Virus Res.* 195, 79–85. doi: 10.1016/j.virusres.2014.08.011

Lutgehetmann, M., Mancke, L. V., Volz, T., Helbig, M., Allweiss, L., Bornscheuer, T., et al. (2012). Humanized chimeric uPA mouse model for the study of hepatitis B and D virus interactions and preclinical drug evaluation. *Hepatology* 55, 685–694. doi: 10.1002/hep.24758

Magoc, T., and Salzberg, S. L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27, 2957–2963. doi: 10.1093/bioinformatics/btr507

Malmstrom, S., Larsson, S. B., Hannoun, C., and Lindh, M. (2012). Hepatitis B viral DNA decline at loss of HBeAg is mainly explained by reduced cccDNA load–down-regulated transcription of PgRNA has limited impact. *PLoS One* 7:e36349. doi: 10.1371/journal.pone.0036349

McVean, G., Awadalla, P., and Fearnhead, P. (2002). A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics* 160, 1231–1241. doi: 10.1093/genetics/160.3.1231

Mederacke, I., Bremer, B., Heidrich, B., Kirschner, J., Deterding, K., Bock, T., et al. (2010). Establishment of a novel quantitative hepatitis D virus (HDV) RNA assay using the Cobas TaqMan platform to study HDV RNA kinetics. *J. Clin. Microbiol.* 48, 2022–2029. doi: 10.1128/jcm.00084-10

Niro, G. A., Gravinese, E., Martini, E., Garrubba, M., Facciorusso, D., Conoscitore, P., et al. (2001). Clearance of hepatitis B surface antigen in chronic carriers of hepatitis delta antibodies. *Liver* 21, 254–259. doi: 10.1034/j.1600-0676.2001.021004254.x

Perez-Losada, M., Arenas, M., Galan, J. C., Palero, F., and Gonzalez-Candelas, F. (2015). Recombination in viruses: mechanisms, methods of study, and evolutionary consequences. *Infect. Genet. Evol.* 30, 296–307. doi: 10.1016/j.meegid.2014.12.022

Perez-Vargas, J., Amirache, F., Boson, B., Mialon, C., Freitas, N., Sureau, C., et al. (2019). Enveloped viruses distinct from HBV induce dissemination of hepatitis D virus in vivo. *Nat. Commun.* 10:2098.

Pfluger, L. S., Schulze Zur Wiesch, J., Polywka, S., and Lutgehetmann, M. (2021). Hepatitis delta virus propagation enabled by hepatitis C virus-Scientifically intriguing, but is it relevant to clinical practice? *J. Viral Hepat.* 28, 213–216. doi: 10.1111/jvh.13385

Rizzetto, M., Hoyer, B., Canese, M. G., Shih, J. W., Purcell, R. H., and Gerin, J. L. (1980). delta Agent: association of delta antigen with hepatitis B surface antigen and RNA in serum of delta-infected chimpanzees. *Proc. Natl. Acad. Sci. U.S.A.*77, 6124–6128. doi: 10.1073/pnas.77.10.6124

Sy, B. T., Nguyen, H. M., Toan, N. L., Song, L. H., Tong, H. V., Wolboldt, C., et al. (2015). Identification of a natural intergenotypic recombinant hepatitis delta virus genotype 1 and 2 in Vietnamese HBsAg-positive patients. *J. Viral Hepat.* 22, 55–63. doi: 10.1111/jvh.12228

Wang, C. J., Chen, P. J., Wu, J. C., Patel, D., and Chen, D. S. (1991). Small-form hepatitis B surface antigen is sufficient to help in the assembly of hepatitis delta virus-like particles. *J. Virol.* 65, 6630–6636. doi: 10.1128/jvi.65.12.6630-6636.1991

Wang, K. S., Choo, Q. L., Weiner, A. J., Ou, J. H., Najarian, R. C., Thayer, R. M., et al. (1986). Structure, sequence and expression of the hepatitis delta (delta) viral genome. *Nature* 323, 508–514. doi: 10.1038/323508a0

Wang, T. C., and Chao, M. (2005). RNA recombination of hepatitis delta virus in natural mixed-genotype infection and transfected cultured cells. *J. Virol.* 79, 2221–2229. doi: 10.1128/jvi.79.4.2221-2229.2005

Wu, J. C., Huang, I. A., Huang, Y. H., Chen, J. Y., and Sheen, I. J. (1999). Mixed genotypes infection with hepatitis D virus. *J. Med. Virol.* 57, 64–67. doi: 10.1002/(sici)1096-9071(199901)57:1<64::aid-jmv9>3.0.co;2-w

Yan, H., Zhong, G., Xu, G., He, W., Jing, Z., Gao, Z., et al. (2012). Sodium taurocholate cotransporting polypeptide is a functional receptor for human hepatitis B and D virus. *Elife* 1:e00049.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership