Contemporary and historical human migration patterns shape hepatitis B virus diversity

Barney I Potter[1,*], Marijn Thijssen[1,*], Nídia Sequeira Trovão[2], Andrea Pineda-Peña[3,4], Marijke Reynders[5], Thomas Mina[6], Carolina Alvarez[1], Samad Amini-Bavil-Olyaee[7], Frederik Nevens[8], Piet Maes[1], Philippe Lemey[1], Marc Van Ranst[1], Guy Baele[1], Mahmoud Reza Pourkarim[1,]**

*contributed equally

[1] Department of Microbiology, Immunology and Transplantation, KU Leuven, Rega Institute, Laboratory for Clinical and Epidemiological Virology, Herestraat 49, Leuven BE-3000, Belgium
[2] Division of International Epidemiology and Population Studies, Fogarty International Center, National Institutes of Health, Bethesda, Maryland, United States 20892
[3] Global Health and Tropical Medicine, GHTM, Instituto de Higiene e Medicina Tropical, IHMT; Universidade Nova de Lisboa, UNL, Portugal Rua da Junqueira No 100, Lisbon 1349-008, Portugal
[4] Molecular Biology and Immunology Department, Fundacion Instituto de Inmunología de Colombia (FIDIC); Faculty of Animal Science, Universidad de Ciencias Aplicadas y Ambientales (U.D.C.A.), Avenida 50 No. 26-20, Bogota 0609, Colombia
[5] Department of Laboratory Medicine, Medical Microbiology, AZ Sint-Jan Brugge-Oostende AV, Ruddershove 10, B-8000 Bruges, Belgium
[6] Nonis Lab Microbiology - Virology Unit, Gregori Afxentiou 5, Limassol 4003, Cyprus
[7] Cellular Sciences Department, Biosafety Development Group, Amgen Inc., One Amgen Center Drive, Thousand Oaks, CA 91320, USA.
[8] Department of Gastroenterology and Hepatology, University Hospital Leuven, Leuven, Belgium

**Correspondence to:*
Dr. Mahmoud Reza Pourkarim
Laboratory for Clinical and Epidemiological Virology
Department of Microbiology, Immunology and Transplantation
KU Leuven
Herestraat 49, Post box 1040
BE-3000 Leuven, Belgium
Email: Mahmoudreza.pourkarim@kuleuven.be

ABSTRACT

Infection by hepatitis B virus (HBV) is responsible for approximately 296 million chronic cases of hepatitis B, and roughly 880,000 deaths annually. The global burden of HBV is distributed unevenly, largely owing to the heterogeneous geographic distribution of its subtypes, each of which demonstrates different severity and responsiveness to antiviral therapy. It is therefore crucial to the global public health response to HBV that the spatiotemporal spread of each genotype is well characterized. In this study, we describe a collection of 133 newly-sequenced HBV strains from recent African immigrants upon their arrival in Belgium. We incorporate these sequences – all of which we determine to come from genotypes A, D, and E – into a large-scale phylogeographic study with genomes sampled across the globe. We focus on investigating the spatio-temporal processes shaping the evolutionary history of the three genotypes we observe. We incorporate several recently published ancient HBV genomes for genotypes A and D to aid our analysis. We show that different spatio-temporal processes underlie the A, D and E genotypes with the former two having originated in southeastern Asia, after which they spread across the world. The HBV E genotype is estimated to have originated in Africa, after which it spread to Europe and the Americas. Our results highlight the use of phylogeographic reconstruction as a tool to understand the recent spatiotemporal dynamics of HBV, and highlight the importance of supporting vulnerable populations in accordance with the needs presented by specific HBV genotypes.

INTRODUCTION

Hepatitis B virus (HBV) imposes a significant global burden to public health (1,2), causing about 880,000 deaths per year (3) despite over two decades of efforts by the global community under the banner of the World Health Organization (4). The efficient implementation of prevention and control measures to combat the spread of HBV requires detailed insight into the epidemiology of the separate lineages circulating in a given region (5,6). Such information at both country and continent levels is pivotal for implementing effective intervention strategies and estimating the burden of disease accurately (1). Both effective antiviral therapies and a widely available vaccine have contributed to significant decreases in the risk of HBV infection globally (7), however their effectiveness can vary significantly by HBV genotype (8).

The global genetic diversity of HBV is classified into eight approved genotypes (A-H) and two tentative genotypes (I and J). These genotypes are further subdivided into more than 45 subgenotypes and quasi-subgenotypes (9,10). Increasing evidence suggests that its genotypes and subgenotypes are determinants for patients' disease progression and may prompt heterogeneous responses to antiviral therapy (11–16). The genotypes and subgenotypes of HBV show distinct geographical distributions. For example, genotype A (HBV-A) subgenotype A2 is dominant in Europe, whereas subgenotype A1 is mostly prevalent in Africa and East Asia. Genotypes B and C circulate with high frequency in Southeast Asia and the Pacific Islands (17). While genotype D (HBV-D) is distributed worldwide, subgenotype D1 is the most prevalent HBV subgenotype in western Asia and the Mediterranean Basin (18–20), and subgenotypes D2 and D3 are prevalent in eastern Europe (9,21). Genotype E (HBV-E) is mainly prevalent in West Africa (22), whereas genotype F is one of the most prevalent genotypes circulating in South and Central America (23). Genotype G is a rarely isolated genotype in France, the United States, and Belgium (24,25).

Genotype H has been occasionally reported in Japan but is commonly encountered in Central and South America. Recently, two new tentative genotypes (I and J) have been proposed, which are most likely recombinant strains of the previously identified genotypes (9).

The global epidemiology of HBV infection is characterized by hepatitis B surface Antigen (HBsAg) seroprevalence, which displays strong geographical variation. Regions may be classified as low seroprevalence (seroprevalence < 2%), such as European countries, intermediate seroprevalence (2% to 7%; e.g. Alaska, the Mediterranean basin, and India), or high seroprevalence (seroprevalence > 8%), such as the South Pacific and countries of Sub-Saharan Africa. In high seroprevalence regions, HBV is considered to be endemic. Viral transmission routes largely depend on regional prevalence (1). A study by the WHO in 2015 showed that global migration trends are crucial to the disease burden and gradual increase of the chronic reservoir in industrialized, low-endemicity countries receiving new waves of migrants from high- and intermediate-endemicity countries without obligatory vaccination (26). Consequently, large migration events have a considerable impact on the prevalence of communicable and contagious diseases like HBV, since there are significant differences of HBV prevalence in “source” and “sink” populations (27). It has been frequently reported that non-endemic HBV strains carried by streams of immigrants are changing the viral epidemiological profile in the destination countries (1,27–31). Such non-endemic strains present a long-lasting tolerance phase, with long-term viral shedding and hepatitis B e-Antigen (HBeAg) positive status (32), as well as high viral loads. This translates to different mutational patterns in different open reading frames (ORFs), which could potentially have an impact on diagnosis, prophylaxis, course of infection, and therapeutic approaches (33,34). Furthermore, mutations associated with antiviral resistance have been reported in HBV strains isolated from patients who had recently migrated into Europe, even

before exposure to treatment (8,35). European countries are now faced with new public health challenges to control HBV incidence and prevalence, necessitating updated policies for screening, monitoring, and preventing transmission from large reservoirs (5,31).

Historically, Belgium has witnessed high rates of immigration, particularly from African countries, with intermediate to high HBV seroprevalence (5). However, until now there has been a dearth of comprehensive evolutionary analyses of non-domestic HBV strains in Belgium. In this study, we describe a collection of HBV strains isolated from patients who had recently immigrated to Belgium from a number of African countries. Many of the viruses may potentially carry medically important genomic mutations, complicating public health interventions. Due to the limited global epidemiological insights and the potential impact of continued spread of such strains, we here conduct a molecular epidemiological study investigating the spatio-temporal processes shaping the HBV genotypes imported to Belgium, as well as their clinical characterization in a sample collection spanning half a decade. Our analyses differ from other recent studies such as that performed by Kocher et al. (36) in that we analyze each genotype independently in order to understand the spatiotemporal factors that have contributed to the global proliferation of each genotype.

MATERIALS AND METHODS

Study population & supplemental sequence selection

One hundred and thirty-three surface antigen (HBsAg) positive individuals, who immigrated to Belgium from Africa between 2005 and 2010 were enrolled in this study. Serum samples were collected from patients upon their first clinical visit during the study. Virological and serological

markers were assayed, and demographic information like sex, age, origin, possible transmission routes and antiviral therapy were collected. For our analyses, we supplemented these novel sequences with all HBV-A, HBV-D and HBV-E sequences available from the NCBI genome database (37) during our study period. Supplemental sequences were screened for quality and excluded if they showed signatures of recombination or gaps longer than 50bp in length. This study was approved by the Ethical Committee of University hospital of KU Leuven (S56121/ML10229).

Extraction, amplification, and complete genome sequencing

Viral DNA was extracted from sera using the QIAmp1 Viral DNA mini kit (Qiagen, Hilden, Germany). Complete HBV genomes were amplified and sequenced using a previously described method (1,38). To filter out possible recombinants, sequences were analyzed using SimPlot v.3.5.1 and BootScan (39).

Genotyping and subgenotyping

We aligned genomic sequences using MAFFT v7.0 (40), and used the same genome alignments for subsequent analyses. In order to determine genotype and subgenotype of sequenced strains, maximum-likelihood phylogenetic trees were constructed using IQ-TREE v1.6.12 (41) under an HKY nucleotide substitution model (42), accommodating among-site rate heterogeneity through a discretized gamma distribution (43). Clade support was verified using UFBoot2 with 1,000 replicates (44). We assigned new sequences genotype and subgenotype designations based on their phylogenetic placement.

Bayesian phylogenetic inference

We started by investigating the temporal signal of each genotype using TempEst (45). We first estimated an unrooted phylogeny for each genotype using maximum-likelihood inference produced in the previous step. Sequences that showed incongruent temporal patterns were excluded from further analyses. This resulted in a dataset with 583 HBV-A sequences (dated 1979–2014), 764 HBV-D sequences (dated 1975–2012), and 234 HBV-E sequences (dated 1994–2010). Phylogenetic relationships were inferred for each of the datasets separately by performing Bayesian phylogenetic inference using Markov chain Monte Carlo (MCMC), as available via the BEAST v1.10 package (46). We used an uncorrelated relaxed molecular clock with branch rates drawn from an underlying lognormal distribution to account for evolutionary rate variation among lineages, with a constant demographic model as the tree prior (47,48). To accommodate the impact of the overlapping regions in the HBV genome on the substitution process of each codon position, we assumed an HKY substitution model (42) without codon partitioning, with among-site rate heterogeneity through a discretized gamma distribution (43). We performed at least three independent replicates of our MCMC analysis to ensure proper convergence, with each replicate having a different starting seed to achieve proper statistical mixing and running for at least 500 million iterations to achieve convergence. Chains were sampled every 10,000 iterations. We assumed default priors for all other parameters in BEAST for these analyses and used the BEAGLE v3 (49) high-performance computational library to speed up the likelihood calculations. All relevant parameters exhibited proper statistical mixing, reporting ESS values over 200 as assessed using Tracer v.1.7 (50), with statistical uncertainty reflected by the 95% highest posterior density (HPD) interval.

Following the analysis of the contemporary genomic data sets, and to ensure sufficient temporal signal in the data, we repeated the analysis of HBV-A and HBV-D using available ancient genomes, collected from mummified tissue ranging in age from 450 years old to ~4,200 years old and spanning from Italy, across Eastern Europe, and into Central Asia (51,52). This resulted in the addition of four ancient genomes to the HBV-A dataset and five ancient genomes to the HBV-D dataset. No ancient sequences were available to add to the HBV-E dataset. Phylogenetic inference was performed under the same model parameters as analyses without ancient genomes, but used an updated lognormal evolutionary rate prior (μ=-11.3474, σ=1.22) derived from the HBV evolutionary rate estimated by Muhlemann et al. (2018) of 1.18×10^{-5} substitutions per site per year, based on the inclusion of ancient sequences. Default priors were used for all other parameters. For each subtype, MCMC chains were run for 500 million iterations and were sampled every 25,000th iteration. Three replicates of each subtype were run using different starting seeds to ensure proper convergence and statistical mixing. A single combined posterior distribution was formed by combining all replicates for each subtype, removing an appropriate proportion of each chain as burn-in. Maximum clade credibility (MCC) trees were summarized using TreeAnnotator v1.10 (46) and the resulting trees were visualized in FigTree v1.4.3 and baltic v.0.1.5 (https://github.com/evogytis/baltic accessed on 10 June 2020).

Phylogeographic inference

From the combined posterior distribution of each genotype's phylogenetic reconstruction, 1,000 trees were sampled and used to perform subsequent Bayesian phylogeographic reconstruction in BEAST using empirical tree distributions (53). This analysis modeled geographic locations as

discrete traits, and allowed for different migration rates to and from each location (54). This approach conditions on the geographic locations recorded at the tips of the trees and models the transition history among those locations as a continuous-time Markov chain (CTMC) process to infer the unobserved locations at the internal nodes of each tree. BEAST default priors were used for each parameter. For these reconstructions, MCMC chains were run for 10 million iterations and sampled every 10,000th iteration. The links between locations that contribute significantly to explaining the migration history were identified using a Bayesian stochastic search variable selection (BSSVS) procedure (54), and SpreaD3 (55) was used to calculate the Bayes factor (BF) support for these links. We quantify geographic transitions by logging Markov jumps and rewards through the phylogeographic analysis. We note that the data consist of a heterogeneous number of sequences for each location (Supplementary Tables 1, 2, 3) which show strong sampling bias between individual countries that can strongly impact the outcome of a discrete phylogeography study (56). To mitigate this issue, we opted to group sequences into continental regions, aiming to contain a number of sequences of the same order of magnitude where possible: Africa, the Americas, East & South Asia, Europe, and West & Central Asia. We note that for HBV-E, no samples from East & South Asia nor West & Central Asia were available.

Statistical analysis

Statistical analyses were performed using Stata software version 16.1 for Windows. Categorical variables are expressed as frequencies; continuous variables are expressed as either means and standard deviations or medians and interquartile ranges, according to their distributions. Chi-squared test and Fisher's exact test were applied to assess the association between

categorical variables. T-test and Mann-Whitney-Wilcoxon test were used to assess continuous variables. The significance level was set to 0.05.

RESULTS

Lineages circulating in Belgium and new isolates

We originally investigated 133 individuals coming from 15 countries in Sub-Saharan Africa. The majority are originally from the Republic of the Congo (n = 44; 33%), Rwanda (n = 23; 17.25%) and Guinea (n = 21; 15.75%) . The country of origin for one individual was unknown. Most of the individuals were male (n = 96; 72%) and the median age was 30 years [interquartile range (IQR) = 25-37]. Nineteen individuals (14%) were co-infected with human immunodeficiency virus (HIV), while three (2%) were hepatitis C virus (HCV) carriers. Positive e-antigen (eAg) was present in 78 cases (58.65%) (Suppl. Table 1).

Phylogenetic analysis and nucleotide divergence analysis of sequences with full-length genomes (n=121) and large S genes (n=12) revealed three distinct HBV genotypes circulating in African immigrants in Belgium. The most frequently identified genotypes were HBV-E (n = 67, 50.4%) and HBV-A (n = 55, 41.4%). Recombination was explored using SimPlot and BootScan. (Suppl. Table 2 and Suppl. Figure 1). The genomes of three isolates (ID: MB-58, MB-56 and MB-92), from one patient of Guinea and two patients of Rwanda, showed recombination: between HBV-A and HBV-E in MB-58, and between HBV-D and HBV-E in both MB-56 and MB-92 (Suppl. Figure 1). MB-58 harbors 900 base pairs (bp) of HBV-E (1100 – 2000 bp) in the context of HBV-A. MB-92 carries 600 bp (1800 – 2400 bp) of HBV-E within HBV-D, with the

wild type length of 3182 bp. However, MB-56 harbors the same piece of HBV-E in the context of HBV-D with 3212 bp of wild type genome length of HBV-E.

HBV Mutational Patterns.

For our new sequences, single nucleotide polymorphisms and amino acid polymorphisms were assessed throughout the genome. We further evaluated if the genotypes identified were associated with the presence of specific drug resistance mutations in the *pol* gene and medically important mutations at *core*, and *X* genes. HBV-A was associated with the presence of *core* mutations at positions 1762, 1896, 1899 and double mutations 1653+1762, 1764+1766 and 1896+1899, and with *X* mutations at position 130. HBV-D was positively associated with the presence of a mutation at position 180 of the *pol* gene. We found that HBV-E was associated with single mutations in the *core* gene at positions 1653, 1762, 1858, 1896, 1899, and double mutations at positions 1653+1762, 1653+1764 and 1764+1766 (Suppl. Table 3).

We quantified strains with substitutions in different S ORFs for all three genotypes (Suppl. Table 4). Clinically important mutations which were mostly located at the major hydrophilic region (MHR) of HBsAg including sP120T, sT131K, sM133T, and sG145A were detected in all strains. Additionally, strains that harbored antiviral resistant mutations in the *pol* ORF, showed corresponding mutations E164D/G/K in the MHR region. We found no strains with evidence of genomic insertions. In 31 strains, we found evidence of deletions, primarily in the HBV pre-S1 and Pre-S2 region.

Phylogenetic inference and subgenotyping

After filtering recombinants from our dataset, 118 complete genome sequences across HBV-A (n=47), HBV-D (n=7), and HBV-E (n=64) were retained and combined with sequences from across the world to construct time-stamped phylogenies for each of the genotypes. For the genotypes analyzed, 54 belong to HBV-A and HBV-D, the only genotypes we consider for which subgenotype definitions exist. The remaining 64 new sequences belong to HBV-E.

Using maximum-likelihood phylogenetic inference, we inferred subgenotypes for novel HBV-A and HBV-D sequences. Among HBV-A, the identified subgenotypes were A1 (n=29, 61.7%), A2 (n=1, 2.1%), A3 (n=9, 19.1%), A5 (n=2, 4.2%), and A6 (n=7, 12.72%) (Figure 1). Among HBV-D, we identified the subgenotypes of seven sequences; these belong to D1 (n=4, 57.1%), D2 (n=2, 28.6%), and D7 (n=1, 14.3%) (Figure 2). While subgenotype definitions do not exist for HBV-E, we note that our sequences span the entire diversity of the genotype (Figure 3).

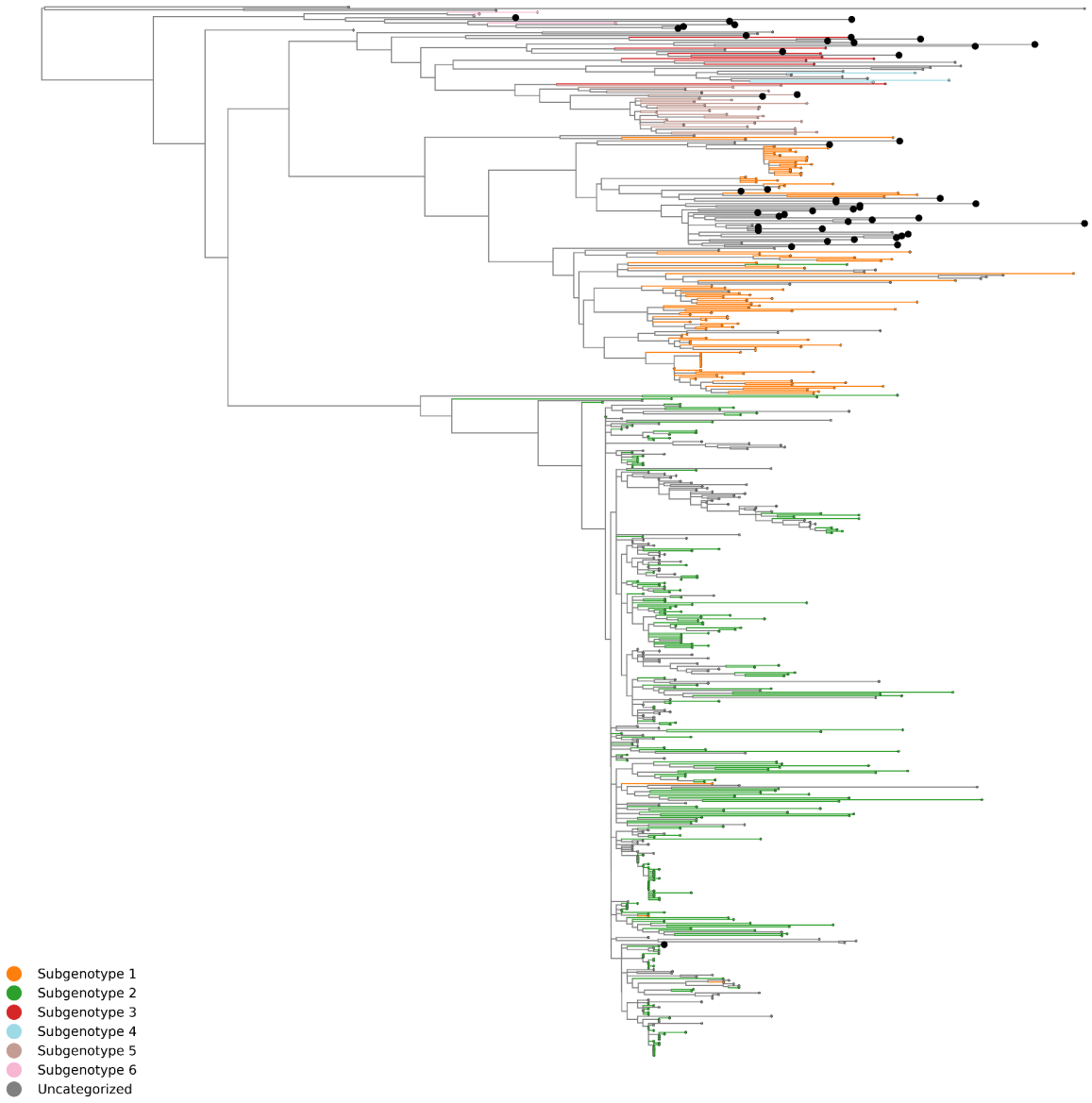


Figure 1. Newly sequenced HBV-A genomes in this study. Midpoint rooted phylogenetic tree (ancient sequences removed and ancient branch lengths rescaled for visibility) representing the diversity of the novel HBV-A sequences from this study. The sequences from our study are shown as enlarged tips on the phylogeny. Branches leading to taxa with a known subgenotype are colored by their subgenotype. Out of a total of 47 new genomes, we introduce 29 genomes that cluster most closely with subgenotype 1. Additionally, we introduce one genome that falls within subgenotype 2, nine new genomes that lie within the diversity of subgenotype 3, two genomes that cluster with subgenotype 5, and six genomes that cluster most closely with

subgenotype 6.

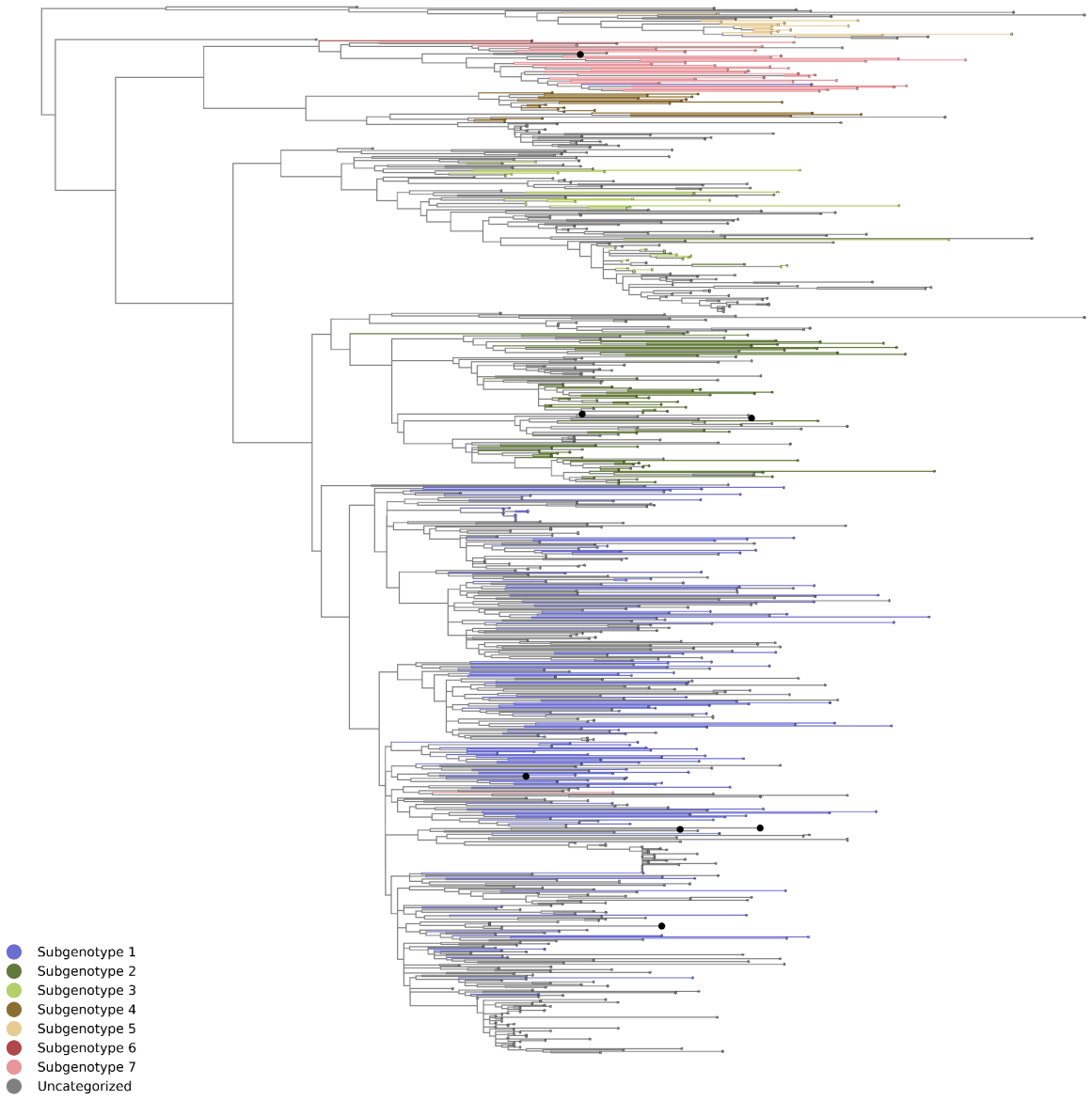


Figure 2. Newly sequenced HBV-D genomes in this study. Midpoint rooted phylogenetic tree (ancient sequences removed, ancient branch lengths rescaled for visibility) representing the diversity of the novel HBV-D sequences from this study. The sequences from our study are shown as enlarged tips on the phylogeny. Branches leading to taxa with a known subgenotype are colored by their subgenotype. Here, we introduce four novel genomes that fall within the known diversity of subgenotype 1, two genomes that fall within subgenotype 2, and one genome in subgenotype 7.

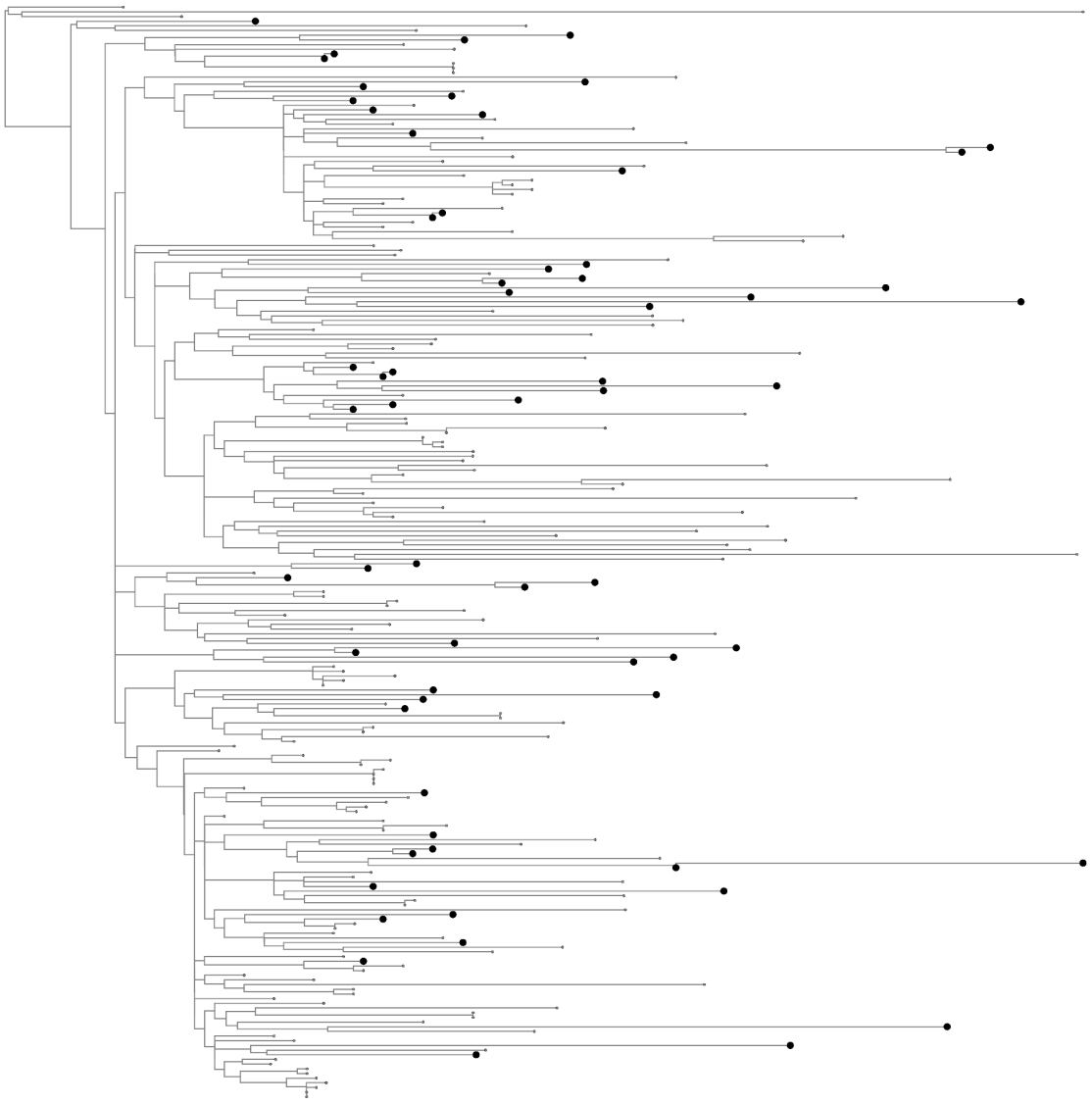


Figure 3. Newly sequenced HBV-E genomes in this study. Midpoint rooted phylogenetic tree representing the diversity of the novel HBV-E sequences from this study. The sequences from our study are shown as enlarged tips on the phylogeny. Novel genomes introduced here represent over a quarter (64/234) of the total number of available HBV-E sequences.

Following maximum-likelihood inference, we performed Bayesian phylogenetic inference on all three datasets. The mean evolutionary rate and time to the most recent common ancestor (tMRCA) were first estimated for all data sets without including any ancient sequences

(and without any informative prior information). We estimated HBV-A to evolve at a mean rate of 5.75 x 10^{-4} [95% HPD: 3.81 x 10^{-4} – 7.90 x 10^{-4}] substitutions(subst)/site/year, HBV-D at a mean rate of 1.27 x 10^{-3} [95% HPD: 1.12 x 10^{-3} – 1.43 x 10^{-3}] subst/site/year, and HBV-E at a mean rate of 6.84 x 10^{-4} [95% HPD: 2.08 x 10^{-4} – 1.16 x 10^{-3}] subst/site/year. However, the lack of temporal signal in HBV data sets is often problematic, making it difficult to accurately estimate the time scale onto an HBV phylogeny and to hence reconstruct accurate dates for historical migrations of the virus. Mühlemann et al. (2018) have shown that including ancient genomic sequences can provide the required additional information to warrant the use of molecular clock models to reconstruct time-stamped phylogenetic trees for HBV. For HBV-A and HBV-D, we created new datasets that include all previous data supplemented by the relevant ancient sequences. For these new datasets containing ancient genomes, we set an informative prior distribution on the mean of the underlying lognormal distribution for the uncorrelated relaxed clock model, based on the reported estimate from Mühlemann et al. (2018). For HBV-A, this leads us to a somewhat increased mean rate of 7.32 x 10^{-4} [95% HPD: 4.14 x 10^{-4} – 1.06 x 10^{-3}] subst/site/year, while for HBV-D we obtained a much decreased mean rate of 9.65 x 10^{-5} [95% HPD: 5.97 x 10^{-5} – 1.36 x 10^{-4}] subst/site/year.

Discrete phylogeography

We performed discrete phylogeographic reconstruction using MCMC, conditional on the posterior tree distributions from our Bayesian phylogenetic inference with ancient genomes. Figure 4 shows the discrete phylogeographic reconstruction for HBV-A, including the relevant ancient genomic sequences from Mühlemann et al. (2018). We estimate HBV-A to have originated in East/South Asia around 2500 BCE, after which it spread to Europe circa 2000 BCE,

and onwards to Africa between 500-1500 CE. Following the jump to Africa, the topology of trees generated both with and without ancient genomes is well conserved. Including the ancient sequences resulted in a more recent estimate of the first occurrence of HBV-A in Africa by about 50 years (from the 1740s to the 1790s). We observed at least 5 different introductions from Africa to the Americas taking place after the year 1700, consistent with the timing of the transatlantic slave trade which took place between 1520s-1850s. Following 1960, we observed a marked increase in the number of viral transmission events between the Americas, Europe, and Western Asia. During this time period, we infer the majority of viral migration to be of European origin, which seeded introductions primarily to the Americas and East/South Asia (avg. 13 and 12 Markov jumps, respectively).

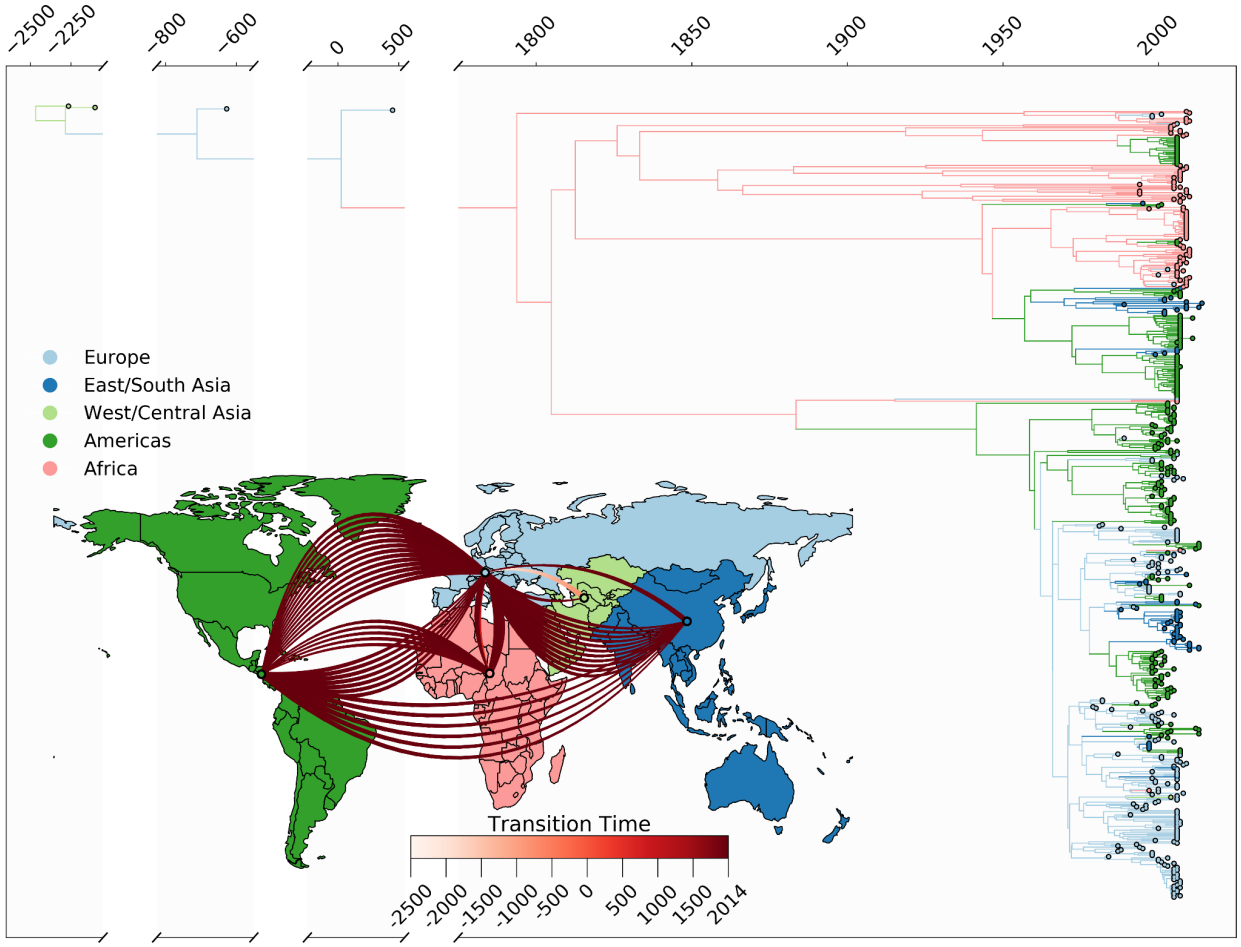


Figure 4. Spatio-temporal transmission dynamics of HBV-A. Time-scaled maximum clade credibility tree representing the spatio-temporal history of HBV genotype A. The time of origin is inferred to be at 2504 BCE. Tips and branches are colored according to inferred location. Breaks in the time axis represent long periods of time covered by individual branches. (inset) World map colored according to the geographic regions used for the discrete phylogeographic analysis. Each line represents an inferred transmission event between two regions (55 total). Line colors denote timing intervals of each introduction. Concave up lines represent west to east movement, concave down lines represent east to west movement, concave right lines represent north to south movement, and concave left lines represent south to north movement. Europe shows the greatest number of outgoing introductions (29). West/Central Asia is the inferred root location of the tree, with one ancient transmission to Europe inferred to have taken place between 2243-876 BCE.

Using the same methods as for HBV-A, we performed discrete phylogeographic inference for HBV-D. We estimate HBV-D to have originated in East/South Asia around 1779

BCE, after which it rapidly spread to West/Central Asia. We observe much more of HBV-D's evolutionary history taking place in East/South Asia and West/Central Asia, with the former spawning introductions into Africa, the Americas, and Europe. Most of the introductions into the Americas occur from Europe however, with Europe in turn being mostly seeded from East/South Asia (Figure 5). For HBV-D, we observe a large impact of including the ancient sequences into our data set. While we have shown that this leads to a large decrease in evolutionary rate, Figure 5 shows much older divergence times being estimated compared to a data set with only modern-day sequences (data not shown). The tMRCA of the largest predominantly African lineage is estimated to be around 550 CE, with the oldest tMRCA of a mixed European/American clade estimated to be around 1100 CE. We infer the MRCA of this clade to have existed within the Americas, however this timing does not correspond to any know human movement patterns, so we believe this inference to be an artifact of either heterogeneous sampling or the weak temporal signal of HBV-D. As with genotype A, individual clades show a strong local geographic structuring, with few individual location changes occurring within the most recent 50 years.

We also performed similar phylogeographic reconstruction for HBV-E, however without any ancient genomes, as none were publicly available (Figure 6). We estimated the tMRCA of HBV-E to 1825 CE, with the introductions into Europe and the Americas estimated to have come directly from the African continent. Given that nearly all European sequences in the tree appear as singletons, dating these introductions would be accompanied with much uncertainty.

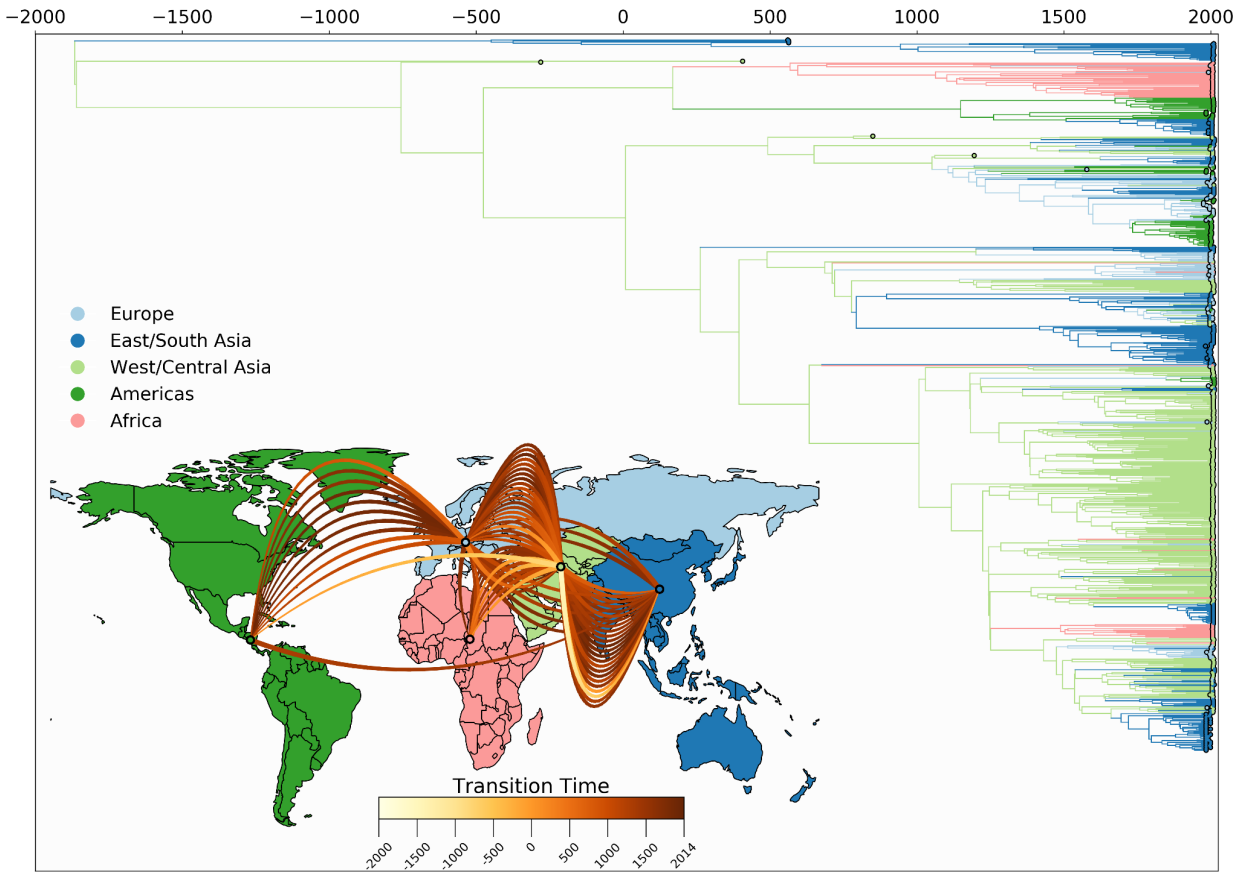


Figure 5. Spatio-temporal transmission dynamics of HBV-D. Time-scaled maximum clade credibility tree representing the evolutionary history of HBV genotype D. The time of origin is inferred to be at 1779 BCE. Tips and branches are colored according to inferred locations. (inset) World map colored according to the geographic regions used for the discrete phylogeographic analysis. Each curved line represents an inferred transmission event between two regions (82 total). Line colors denote timing intervals of each introduction. Concave up lines represent west to east movement, concave down lines represent east to west movement, concave right lines represent north to south movement, and concave left lines represent south to north movement. West/Central Asia shows the greatest number of outgoing transmission events (55), as well as being the most likely location of the most recent common ancestor to sampled modern sequences.

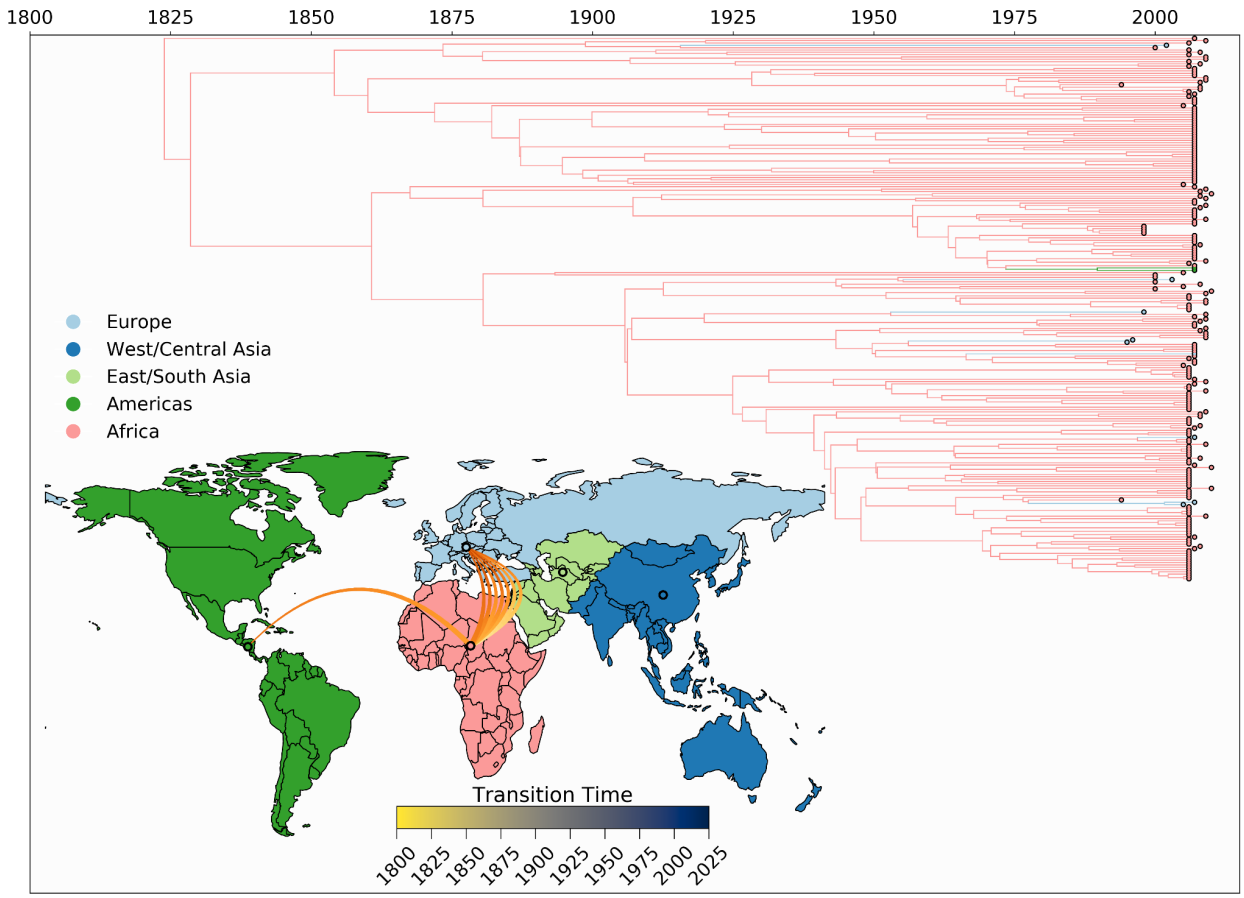


Figure 6. Spatio-temporal transmission dynamics of HBV-E. Time-scaled maximum clade credibility tree representing the evolutionary history of HBV genotype E. The time of origin is inferred to be at 1824 CE. Tip and branches are colored according to inferred location. *(inset)* World map colored according to the geographic regions used for the discrete phylogeographic analysis. Each counterclockwise arrow represents an inferred transmission event between two regions. Arrow colors denote timing intervals of each introduction. All migration events (8 total) are inferred to have originated in Africa; Europe was the destination of seven migrations, and the Americas were the destination of one.

We summarized the strongest supported rates of discrete location transitions between pairs of regions for each dataset (Figure 7). For HBV-A, we find very strong Bayes factor support (57) for movements from Europe to East/South Asia and the Americas, as well as strong support for movements from the Americas and Africa to Europe. For HBV-D, we observe very strong support for outgoing movements from West/Central Asia to Europe, Africa and

East/South Asia. We also find very strong support for outgoing movements from Europe to East/South Asia, West/Central Asia and the Americas. Finally, for HBV-E, we only find very strong support for movements from Africa into Europe and strong support for movements from Africa to the Americas. For each genotype, we estimated Markov jump counts to quantify total geographic transitions. We estimated 62 [95% HPD: 51–72] transitions in the evolutionary history of HBV-A, 92 [95% HPD: 85–100] transitions in the evolutionary history of HBV-D, and 8 [95% HPD: 8–10] transitions in the evolutionary history of HBV-E.

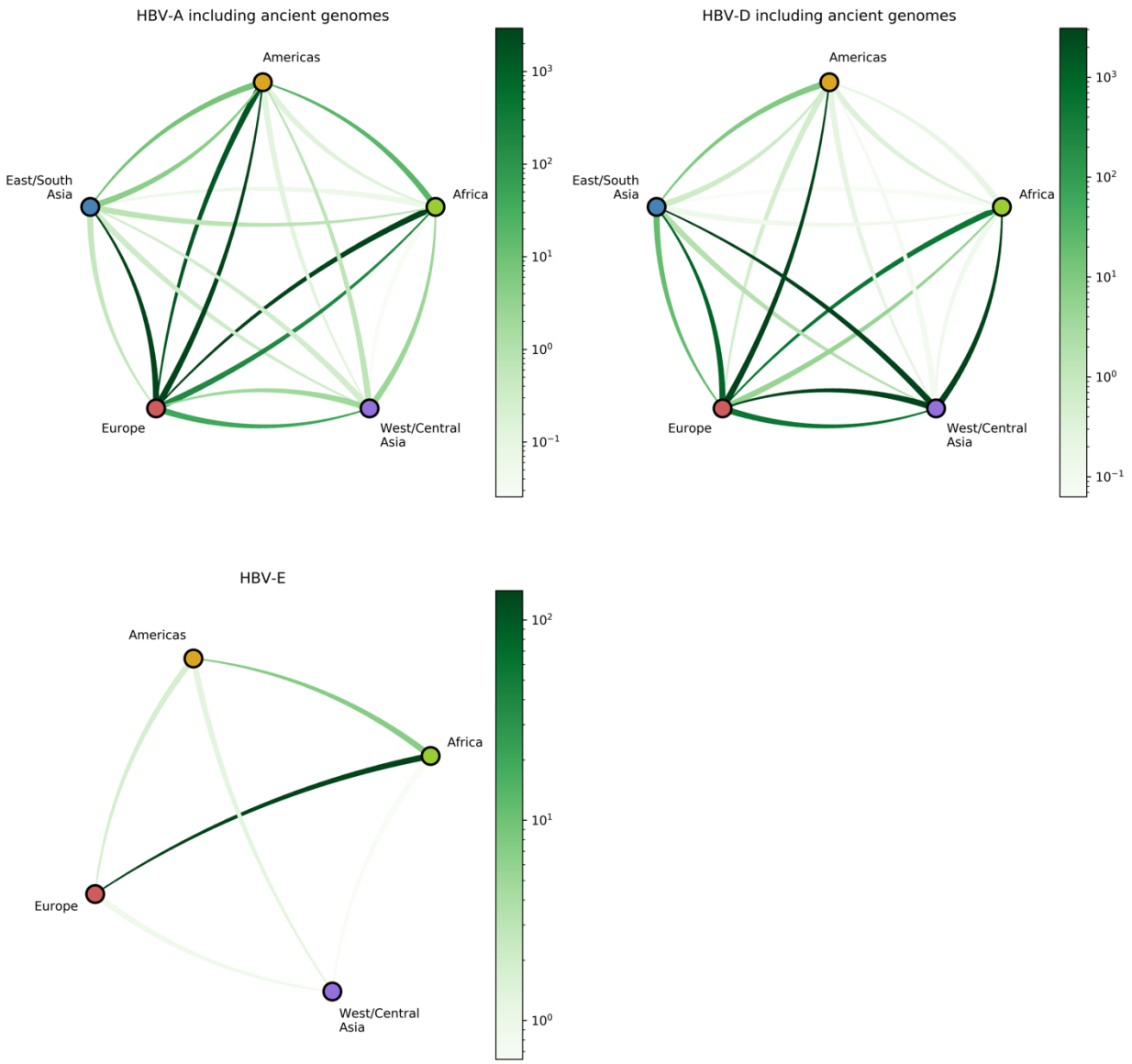


Figure 7. Bayes factor support for HBV's geographic transitions. Networks representing posterior support for migrations between geographic regions as determined by BSSVS analysis. Inferred migrations are represented as curved lines. Bayes factors are represented by color intensity; darker lines depict higher Bayes factors, therefore higher posterior support. Direction of movement is represented by anticlockwise curvature of each line. In HBV-A, we estimate strongly supported migrations out of Europe to all other regions, as well as high support for inferred migrations into Europe from Africa and the Americas. In HBV-D, we observe well-supported migration history from Africa to Europe, and from Europe to the Americas and Asia. Finally, in HBV-E, we observe very strong support for migration from Africa to Europe, as well as relatively high support for migration from Africa to the Americas.

DISCUSSION AND CONCLUSION

Human mobility exerts an immense pressure on the infectious disease profile and public health landscape of countries who receive significant numbers of immigrants (31). The prevalence and diversity of HBV is disproportionately affected by immigration when compared with other bloodborne viruses such as HIV and HCV. A recent large study showed that the frequency of HBV infection in immigrants in parts of Europe is three times greater than for HIV and four times greater than for HCV (58). Specific virological characteristics of HBV such as a long window of infectiousness, possibilities of occult infection promoted by mutations (59–61), and undetectable levels of genome amplification without antigen production have raised concerns about cryptic movement of specific HBV genotypes from countries of high endemicity to those of low endemicity (62–64). Systemic inequities in access to health services faced by immigrants in higher income countries (65) compound with these virological characteristics to impair the implementation of effective public health responses to HBV.

Despite the significant public health burden of HBV worldwide, there is limited insight into the evolutionary history of its different genotypes. To the best of our knowledge, we present the most comprehensive investigation to date on the origins of HBV genotypes A, D, and E, owing to the large number of full-length African HBV genomes and broad spatiotemporal distribution of sequences we consider. In this study, we focused on HBV genotypes obtained from African immigrants in Belgium upon their arrival from 2005 to 2010, along with a broad selection of other publicly available sequences, both contemporary and ancient. Almost 16% of HBV carriers from our study cohort showed co-infection with two other bloodborne viruses (HIV and HCV) which makes the design of antiviral therapies more complicated in destination countries (66,67). Three distinct genotypes were detected in immigrants from 15 African

countries: HBV-E (50.4%), HBV-A (41.4%), and HBV-D (8.6%). This finding agrees with previous reports of HBV genotype diversity in sub-Saharan countries (68,69). The diversity of imported genotypes underscores the potential for changes in the epidemiological landscape of HBV in recipient countries (29,70). The HBV-A strains that we generated belong to the subgenotypes A1, A3, and quasi-subgenotypes A4, A5, and A7 which have already been isolated from African HBV carriers living in different parts of the world (69,71–73). The recent discovery of novel subgenotype A8 from our study cohort (10) highlights the lack of comprehensive information about HBV in African immigrants in Belgium.

There is a well-documented association between HBV genotype and clinical outcome; this relation is further documented for individual HBV ORF mutations (16,29,74–76). These associations between genotype-specific mutations in circulating African strains and clinical severity have been demonstrated, for instance with particular medically important mutations being mostly detected in the *core* gene of HBV strains isolated from patients with end stage liver diseases such as hepatocellular carcinoma (HCC) (77–79). The association between A1762T, G1764A, and A1762T/G1764A, G1896A, and an increased risk of HCC has been previously shown (80). High frequency of pre-core and core mutations detected in HBV-E and HBV-A in our study highlights the importance of investigating individual origins for HBV cases, so that clinical interventions can be targeted appropriately. Importantly, we identified clinically and para-clinically related mutations at ORF S, as previously described (60,79,81).

Excepting their most ancient origins, our results agree with the work of others, with HBV-A demonstrating an African origin (1,73,82,83). Our results underscore the African continent as the evolutionary cradle of HBV-A diversification (69,83), and we date the proliferation and worldwide dispersal of this genotype to coincide with the transatlantic slave

trade (1520s–1850s). Similar to HBV-A, we find the most ancient origins of HBV-D to be in Central Asia, however we also infer the majority of its evolutionary history to remain in the region, with much later introductions into other global regions (9). For both HBV-A and HBV-D, the location that we infer for the MRCA of the genotype corresponds to the location of the most ancient genomes that we included in the study. As more ancient HBV genomes from different locations and times become available, we propose incorporating them into a study such as ours to reduce the uncertainty of MRCA location and time inference.

Almost half of the HBV strains that we generated in this study were classified as HBV-E (Supplementary Table 2). This high proportion is supported by other studies from West African countries (69,84). Because of low diversity along its entire length of genome, HBV-E does not classify into subgenotypes (85) which may point to HBV-E having evolved more recently than other genotypes (22). In this study dataset, three strains with recombination signature were detected (Supplementary Figure S1) as previously described (16,82,85,86). Recombination can happen in over-populated locales with high prevalence of HBV, mostly occurring in East Asia and Africa (9,86).

We observed that HBV-D was the fastest evolving among the genotypes we encountered. HBV-E is the most recently diverged genotype, however both its divergence and evolutionary rate estimates have large 95% highest posterior density intervals, indicating that these estimates must be cautiously interpreted. This is possibly explained by the smaller dataset for HBV-E (234 sequences), compared to the dataset sizes for the other genotypes (583 HBV-A and 764 HBV-D sequences).

There is recent evidence that HBV has limited temporal signal warranting temporal and evolutionary estimation not reliable, particularly for very long timeframes, with the risk of

underestimating when the MRCA of each genotype diverged from the ancestral HBV lineage (51). To avoid the inherent biases of estimating these measures for viruses that have an ancient time of divergence, we opted for analyzing each genotype separately. Additionally, we include ancient HBV genomes when possible to provide increased temporal signal in our dataset, thus making the estimation of evolutionary rate and tMRCA more reliable. A recent study by Kocher et al. (36) performed similar temporal phylogenetic analysis of HBV and found MRCA estimates for each genotype different from ours by approximately a factor of two. This discrepancy may be explained by their use of a epoch-based time-dependent rate model, while we use a relaxed clock model for each genotype that we consider. Additionally, differences may arise as a result of our analysis of each HBV genotype separately, whereas Kocher et al. consider the entire unified HBV phylogeny. We find different mean evolutionary rates for each genotype, suggesting that a single estimate for all genotypes may not fully capture the different evolutionary dynamics affecting each, not even when using a time-dependent rate model.

Rather than employing each individual country as a different location for our discrete phylogeographic reconstruction, we have opted to perform these analyses on the regional level in order to deal with sampling bias among the different countries in our data sets. Unfortunately, to maintain consistent geographic subdivisions across datasets there remains significant heterogeneity in the number of sequences between regions for some analyses. Phylogeographic inference using structured coalescent models (56,87) offers a potential solution to deal with sampling bias, at the expense of much increased computational demands, however we have instead opted to use continent-level analysis here to reduce computational complexity (88).

This study highlights the importance of genotype-aware treatment of HBV interventions, particularly in the case of individuals who have recently migrated from regions where HBV

diversity is high. Because HBV genotype and specific mutational profile can have a significant impact on clinical outcomes, we find it imperative that these characteristics be taken into account so that individuals receive appropriate clinical and public health support for their infecting HBV genotype, and so that local diversity of HBV is well characterized so that policymakers and public health institutions can create the most effective intervention strategies possible. We believe this effort can be significantly assisted by both genetic analyses for individual level interventions, and by large-scale phylogenetic and phylogeographic analyses to inform population level interventions that can lessen the overall burden of hepatitis B.

ACKNOWLEDGEMENTS

BP and GB acknowledge support from the Internal Funds KU Leuven (Grant No. C14/18/094). GB acknowledges support from the Research Foundation - Flanders ("Fonds voor Wetenschappelijk Onderzoek - Vlaanderen," G0E1420N, G098321N). PL acknowledges support from the Research Foundation - Flanders ("Fonds voor Wetenschappelijk Onderzoek - Vlaanderen," G0D5117N, G0B9317N, G051322N). The opinions expressed in this article are those of the authors and do not reflect the view of the National Institutes of Health, the Department of Health and Human Services, or the United States government.

REFERENCES

1. Pourkarim MR, Lemey P, Amini-Bavil-Olyaee S, Houspie L, Verbeeck J, Rahman M, et al. Molecular characterization of hepatitis B virus strains circulating in Belgian patients co-infected with HIV and HBV: Overt and occult infection. J Med Virol. 2011;83(11):1876–84.
2. Malik GF, Zakaria N, Majeed MI, Ismail FW. Viral Hepatitis - The Road Traveled and the

Journey Remaining. Hepatic Med Evid Res. 2022;14:13–26.
3. Revill PA, Tu T, Netter HJ, Yuen LKW, Locarnini SA, Littlejohn M. The evolution and clinical impact of hepatitis B virus genome diversity. Nat Rev Gastroenterol Hepatol. 2020 Oct;17(10):618–34.
4. Mahmoud Reza Pourkarim, Philippe Lemey, Marc Van Ranst. Iran's hepatitis elimination programme is under threat. Lancet. 2018 Sep 12;392(10152).
5. Schweitzer A, Horn J, Mikolajczyk RT, Krause G, Ott JJ. Estimations of worldwide prevalence of chronic hepatitis B virus infection: a systematic review of data published between 1965 and 2013. The Lancet. 2015 Oct 17;386(10003):1546–55.
6. Pourkarim MR, Van Ranst M. Guidelines for the detection of a common source of hepatitis B virus infections. Hepat Mon. 2011 Oct;11(10):783–5.
7. Lu Y, Clark-Deener S, Gillam F, Heffron CL, Tian D, Sooryanarain H, et al. Virus-like particle vaccine with B-cell epitope from porcine epidemic diarrhea virus (PEDV) incorporated into hepatitis B virus core capsid provides clinical alleviation against PEDV in neonatal piglets through lactogenic immunity. Vaccine. 2020 Jul 14;38(33):5212–8.
8. Bottecchia M, Madejón A, Puente S, García-Samaniego J, Rivas P, Herrero D, et al. Detection of hepatitis B virus genotype A3 and primary drug resistance mutations in African immigrants with chronic hepatitis B in Spain. J Antimicrob Chemother. 2011 Mar 1;66(3):641–4.
9. Pourkarim MR, Amini-Bavil-Olyaee S, Kurbanov F, Van Ranst M, Tacke F. Molecular identification of hepatitis B virus genotypes/subgenotypes: Revised classification hurdles and updated resolutions. World J Gastroenterol WJG. 2014 Jun 21;20(23):7152–68.
10. Thijssen M, Trovão NS, Mina T, Maes P, Pourkarim MR. Novel hepatitis B virus subgenotype A8 and quasi-subgenotype D12 in African–Belgian chronic carriers. Int J Infect Dis. 2020 Apr 1;93:98–101.
11. McMahon BJ. The influence of hepatitis B virus genotype and subgenotype on the natural history of chronic hepatitis B. Hepatol Int. 2009 Jun 1;3(2):334–42.
12. Croagh CM, Desmond PV, Bell SJ. Genotypes and viral variants in chronic hepatitis B: A review of epidemiology and clinical relevance. World J Hepatol. 2015 Mar 27;7(3):289–303.
13. Shen T, Yan XM. Hepatitis B virus genetic mutations and evolution in liver diseases. World J Gastroenterol WJG. 2014 May 14;20(18):5435–41.
14. Pourkarim MR, Vergote V, Amini-Bavil-Olyaee S, Sharifi Z, Sijmons S, Lemey P, et al. Molecular characterization of hepatitis B virus (HBV) strains circulating in the northern coast of the Persian Gulf and its comparison with worldwide distribution of HBV subgenotype D1. J Med Virol. 2014 May;86(5):745–57.
15. Zhang M, Li G, Shang J, Pan C, Zhang M, Yin Z, et al. Rapidly decreased HBV RNA predicts responses of pegylated interferons in HBeAg-positive patients: a longitudinal cohort study. Hepatol Int. 2020 Mar;14(2):212–24.
16. Mina T, Amini-Bavil-Olyaee S, Tacke F, Maes P, Van Ranst M, Pourkarim MR. Genomic Diversity of Hepatitis B Virus Infection Associated With Fulminant Hepatitis B Development. Hepat Mon. 2015 Jun 23;15(6):e29477.
17. Norder H, Couroucé AM, Coursaget P, Echevarria JM, Lee SD, Mushahwar IK, et al. Genetic Diversity of Hepatitis B Virus Strains Derived Worldwide: Genotypes, Subgenotypes, and HBsAg Subtypes. Intervirology. 2004;47(6):289–309.
18. Zehender G, Ebranati E, Gabanelli E, Shkjezi R, Lai A, Sorrentino C, et al. Spatial and

Temporal Dynamics of Hepatitis B Virus D Genotype in Europe and the Mediterranean Basin. PLOS ONE. 2012 May 25;7(5):e37198.
19. Al-Qahtani AA, Pourkarim MR, Trovão NS, Vergote V, Li G, Thijssen M, et al. Molecular epidemiology, phylogenetic analysis and genotype distribution of hepatitis B virus in Saudi Arabia: Predominance of genotype D1. Infect Genet Evol J Mol Epidemiol Evol Genet Infect Dis. 2020 Jan;77:104051.
20. Trovão NS, Thijssen M, Vrancken B, Pineda-Peña AC, Mina T, Amini-Bavil-Olyaee S, et al. Reconstruction of the origin and dispersal of the worldwide dominant Hepatitis B Virus subgenotype D1. Virus Evol. 2022 Jan 1;8(1):veac028.
21. Pineda-Peña AC, Faria NR, Mina T, Amini-Bavil-Olyaee S, Alavian SM, Lemey P, et al. Epidemiological history and genomic characterization of non-D1 HBV strains identified in Iran. J Clin Virol. 2015 Feb 1;63:38–41.
22. Ingasia LAO, Kostaki EG, Paraskevis D, Kramvis A. Global and regional dispersal patterns of hepatitis B virus genotype E from and in Africa: A full-genome molecular analysis. PLOS ONE. 2020 Oct 8;15(10):e0240375.
23. Pujol F, Jaspe RC, Loureiro CL, Chemin I. Hepatitis B virus American genotypes: Pathogenic variants ? Clin Res Hepatol Gastroenterol. 2020 Nov 1;44(6):825–35.
24. Kay A, Zoulim F. Hepatitis B virus genetic variability and evolution. Virus Res. 2007 Aug;127(2):164–76.
25. Mina T, Amini-Bavil-Olyaee S, Dekervel J, Verslype C, Nevens F, Maes P, et al. A rare case of HBV genotype fluctuation (shifting and reversion) after liver transplantation. J Clin Virol Off Publ Pan Am Soc Clin Virol. 2015 Oct;71:93–7.
26. Sharma S, Carballo M, Feld JJ, Janssen HLA. Immigration and viral hepatitis. J Hepatol. 2015 Aug 1;63(2):515–22.
27. Chu JJ, Wörmann T, Popp J, Pätzelt G, Akmatov MK, Krämer A, et al. Changing epidemiology of Hepatitis B and migration—a comparison of six Northern and North-Western European countries. Eur J Public Health. 2013 Aug 1;23(4):642–7.
28. Khan A, Tanaka Y, Saito H, Ebinuma H, Sekiguchi H, Iwama H, et al. Transmission of hepatitis B virus (HBV) genotypes among Japanese immigrants and natives in Bolivia. Virus Res. 2008 Mar;132(1–2):174–80.
29. Mina T, Amini-Bavil-Olyaee S, Shirvani-Dastgerdi E, Trovão NS, Van Ranst M, Pourkarim MR. 15year fulminant hepatitis B follow-up in Belgium: Viral evolution and signature of demographic change. Infect Genet Evol. 2017 Apr;49:221–5.
30. Coppola N, Alessio L, Gualdieri L, Pisaturo M, Sagnelli C, Minichini C, et al. Hepatitis B virus infection in undocumented immigrants and refugees in Southern Italy: demographic, virological, and clinical features. Infect Dis Poverty. 2017 Feb 9;6(1):33.
31. Thijssen M, Lemey P, Amini-Bavil-Olyaee S, Dellicour S, Alavian SM, Tacke F, et al. Mass migration to Europe: an opportunity for elimination of hepatitis B virus? Lancet Gastroenterol Hepatol. 2019 Apr 1;4(4):315–23.
32. Lampertico P, Maini M, Papatheodoridis G. Optimal management of hepatitis B virus infection – EASL Special Conference. J Hepatol. 2015 Nov 1;63(5):1238–53.
33. Velkov S, Protzer U, Michler T. Global Occurrence of Clinically Relevant Hepatitis B Virus Variants as Found by Analysis of Publicly Available Sequencing Data. Viruses. 2020 Nov;12(11):1344.
34. Limeres MJ, Gomez ER, Noseda DG, Cerrudo CS, Ghiringhelli PD, Nusblat AD, et al. Impact of hepatitis B virus genotype F on in vitro diagnosis: detection efficiency of HBsAg

from Amerindian subgenotypes F1b and F4. Arch Virol. 2019 Sep 1;164(9):2297–307.
35. Selabe SG, Lukhwareni A, Song E, Leeuw YGM, Burnett RJ, Mphahlele MJ. Mutations associated with lamivudine-resistance in therapy-naïve hepatitis B virus (HBV) infected patients with and without HIV co-infection: Implications for antiretroviral therapy in HBV and HIV co-infected South African patients. J Med Virol. 2007;79(11):1650–4.
36. Kocher A, Papac L, Barquera R, Key FM, Spyrou MA, Hübler R, et al. Ten millennia of hepatitis B virus evolution. 2021;8.
37. Sayers EW, Bolton EE, Brister JR, Canese K, Chan J, Comeau DC, et al. Database resources of the national center for biotechnology information. Nucleic Acids Res. 2022 Jan 7;50(D1):D20–6.
38. Pourkarim MR, Verbeeck J, Rahman M, Amini-Bavil-Olyaee S, Forier AM, Lemey P, et al. Phylogenetic analysis of hepatitis B virus full-length genomes reveals evidence for a large nosocomial outbreak in Belgium. J Clin Virol. 2009 Sep 1;46(1):61–8.
39. Lole KS, Bollinger RC, Paranjape RS, Gadkari D, Kulkarni SS, Novak NG, et al. Full-Length Human Immunodeficiency Virus Type 1 Genomes from Subtype C-Infected Seroconverters in India, with Evidence of Intersubtype Recombination. J Virol. 1999 Jan 1;73(1):152–60.
40. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol Biol Evol. 2013 Apr 1;30(4):772–80.
41. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Mol Biol Evol. 2015 Jan 1;32(1):268–74.
42. Hasegawa M, Kishino H, Yano T aki. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol. 1985 Oct 1;22(2):160–74.
43. Yang Z. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: Approximate methods. J Mol Evol. 1994 Sep 1;39(3):306–14.
44. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: Improving the Ultrafast Bootstrap Approximation. Mol Biol Evol. 2018 Feb 1;35(2):518–22.
45. Rambaut A, Lam TT, Max Carvalho L, Pybus OG. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evol [Internet]. 2016 Jan 1 [cited 2021 Sep 16];2(1). Available from: https://doi.org/10.1093/ve/vew007
46. Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evol [Internet]. 2018 Jan 1 [cited 2021 Sep 16];4(1). Available from: https://academic.oup.com/ve/article/doi/10.1093/ve/vey016/5035211
47. Kingman JFC. The coalescent. Stoch Process Their Appl. 1982 Sep;13(3):235–48.
48. Drummond AJ, Nicholls GK, Rodrigo AG, Solomon W. Estimating Mutation Parameters, Population History and Genealogy Simultaneously From Temporally Spaced Sequence Data. Genetics. 2002 Jul 1;161(3):1307–20.
49. Suchard MA, Rambaut A. Many-core algorithms for statistical phylogenetics. Bioinformatics. 2009 Jun 1;25(11):1370–6.
50. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. Syst Biol. 2018 Sep;67(5):901–4.
51. Ross ZP, Klunk J, Fornaciari G, Giuffra V, Duchêne S, Duggan AT, et al. The paradox of HBV evolution as revealed from a 16th century mummy. PLOS Pathog. 2018 Jan 4;14(1):e1006750.

52. Mühlemann B, Jones TC, Damgaard P de B, Allentoft ME, Shevnina I, Logvin A, et al. Ancient hepatitis B viruses from the Bronze Age to the Medieval period. Nature. 2018 May;557(7705):418–23.
53. Pagel M, Meade A, Barker D. Bayesian Estimation of Ancestral Character States on Phylogenies. Syst Biol. 2004 Oct 1;53(5):673–84.
54. Lemey P, Rambaut A, Drummond AJ, Suchard MA. Bayesian Phylogeography Finds Its Roots. PLOS Comput Biol. 2009 Sep 25;5(9):e1000520.
55. Bielejec F, Baele G, Vrancken B, Suchard MA, Rambaut A, Lemey P. SpreaD3: Interactive Visualization of Spatiotemporal History and Trait Evolutionary Processes. Mol Biol Evol. 2016 Aug 1;33(8):2167–9.
56. Maio ND, Wu CH, O'Reilly KM, Wilson D. New Routes to Phylogeography: A Bayesian Structured Coalescent Approximation. PLOS Genet. 2015 Aug 12;11(8):e1005421.
57. Kass RE, Raftery AE. Bayes Factors. J Am Stat Assoc. 1995 Jun 1;90(430):773–95.
58. González R, Barea L, Arruga A, Richart A, Soriano V. Overt and occult hepatitis B among immigrants and native blood donors in Madrid, Spain. Ther Adv Infect Dis. 2020 Jan 1;7:2049936120982122.
59. Pronier C, Candotti D, Boizeau L, Bomo J, Laperche S, Thibault V. The contribution of more sensitive hepatitis B surface antigen assays to detecting and monitoring hepatitis B infection. J Clin Virol. 2020 Aug;129:104507.
60. Olusola BA, Faneye AO, Oluwasemowo OO, Motayo BO, Adebayo S, Oludiran-Ayoade AE, et al. Profiles of mutations in hepatitis B virus surface and polymerase genes isolated from treatment-naïve Nigerians infected with genotype E. J Med Microbiol. 2021;70(3):001338.
61. Raimondo G, Locarnini S, Pollicino T, Levrero M, Zoulim F, Lok AS, et al. Update of the statements on biology and clinical impact of occult hepatitis B virus infection. J Hepatol. 2019 Aug;71(2):397–408.
62. Azarkar Z, Ziaee M, Ebrahimzadeh A, Sharifzadeh G, Javanmard D. Epidemiology, risk factors, and molecular characterization of occult hepatitis B infection among anti-hepatitis B core antigen alone subjects. J Med Virol. 2019;91(4):615–22.
63. Zhu HL, Li X, Li J, Zhang ZH. Genetic variation of occult hepatitis B virus infection. World J Gastroenterol. 2016 Apr 7;22(13):3531–46.
64. Bedi HK, Chahal D, Lowe CF, Ritchie G, Hussaini T, Marquez V, et al. Occult Hepatitis B Reactivation after Liver Transplant: The Role of a Novel Mutation in the Surface Antigen. J Clin Transl Hepatol. 2021 Feb 28;9(1):136–8.
65. Smith J. Migrant health is public health, and public health needs to be political. Lancet Public Health. 2018 Sep 1;3(9):e418.
66. Arora U, Garg P, Agarwal S, Nischal N, Shalimar, Wig N. Complexities in the treatment of coinfection with HIV, hepatitis B, hepatitis C, and tuberculosis. Lancet Infect Dis [Internet]. 2021 May 21 [cited 2021 Sep 16]; Available from: https://www.sciencedirect.com/science/article/pii/S1473309920307659
67. Torimiro JN, Nanfack A, Takang W, Keou CK, Joyce AN, Njefi K, et al. Rates of HBV, HCV, HDV and HIV type 1 among pregnant women and HIV type 1 drug resistance-associated mutations in breastfeeding women on antiretroviral therapy. BMC Pregnancy Childbirth. 2018 Dec 22;18(1):504.
68. Forbi JC, Ben-Ayed Y, Xia G liang, Vaughan G, Drobeniuc J, Switzer WM, et al. Disparate distribution of hepatitis B virus genotypes in four sub-Saharan African countries. J Clin

Virol. 2013 Sep 1;58(1):59–66.
69. Andernach IE, Nolte C, Pape JW, Muller CP. Slave Trade and Hepatitis B Virus Genotypes and Subgenotypes in Haiti and Africa. Emerg Infect Dis. 2009 Aug;15(8):1222–8.
70. Aguilera A, Trastoy R, Rodríguez-Frias F, Muñoz-Bellido JL, Melón S, Suárez A, et al. GEHEP 010 study: Prevalence and distribution of hepatitis B virus genotypes in Spain (2000–2016). J Infect. 2020 Oct 1;81(4):600–6.
71. Kurbanov F, Tanaka Y, Fujiwara K, Sugauchi F, Mbanya D, Zekeng L, et al. A new subtype (subgenotype) Ac (A3) of hepatitis B virus and recombination between genotypes A and E in Cameroon. J Gen Virol. 2005;86(7):2047–56.
72. Olinger CM, Venard V, Njayou M, Oyefolu AOB, Maïga I, Kemp AJ, et al. Phylogenetic analysis of the precore/core gene of hepatitis B virus genotypes E and A in West Africa: new subtypes, mixed infections and recombinations. J Gen Virol. 2006;87(5):1163–73.
73. Pourkarim MR, Lemey P, Amini-Bavil-Olyaee S, Maes P, Van Ranst M. Novel hepatitis B virus subgenotype A6 in African-Belgian patients. J Clin Virol. 2010 Jan;47(1):93–6.
74. Liu S, Zhang H, Gu C, Yin J, He Y, Xie J, et al. Associations Between Hepatitis B Virus Mutations and the Risk of Hepatocellular Carcinoma: A Meta-Analysis. JNCI J Natl Cancer Inst. 2009 Aug 5;101(15):1066–82.
75. Mello FMMA de, Kuniyoshi ASO, Lopes AF, Gomes-Gouvêa MS, Bertolini DA. Hepatitis B virus genotypes and mutations in the basal core promoter and pre-core/core in chronically infected patients in southern Brazil: a cross-sectional study of HBV genotypes and mutations in chronic carriers. Rev Soc Bras Med Trop. 2014 Dec;47:701–8.
76. Hou W. HCC-Associated Viral Mutations in Patients with HBV Genotype F1b Infection. Clin Lab [Internet]. 2019 Jun 1 [cited 2021 Nov 17];65(6). Available from: https://doi.org/10.7754/Clin.Lab.2018.181121
77. Amougou MA, Marchio A, Bivigou-Mboumba B, Noah DN, Banai R, Atangana PJA, et al. Enrichment in selected genotypes, basal core and precore mutations of hepatitis B virus in patients with hepatocellular carcinoma in Cameroon. J Viral Hepat. 2019;26(9):1086–93.
78. Mbamalu C, Ekejindu I, Enweani I, Kalu S, Igwe D, Akaeze G. Hepatitis B virus precore/core region mutations and genotypes among hepatitis B virus chronic carriers in South-Eastern, Nigeria. Int J Health Sci. 2021;15(2):26–38.
79. Mak D, Kramvis A. Molecular characterization of hepatitis B virus isolated from Black South African cancer patients, with and without hepatocellular carcinoma. Arch Virol. 2020 Aug 1;165(8):1815–25.
80. Wei F, Zheng Q, Li M, Wu M. The association between hepatitis B mutants and hepatocellular carcinoma. Medicine (Baltimore). 2017 May 12;96(19):e6835.
81. Lin YT, Jeng LB, Chan WL, Su IJ, Teng CF. Hepatitis B Virus Pre-S Gene Deletions and Pre-S Deleted Proteins: Clinical and Molecular Implications in Hepatocellular Carcinoma. Viruses. 2021 May;13(5):862.
82. Pourkarim MR, Amini-Bavil-Olyaee S, Lemey P, Maes P, Van Ranst M. Are hepatitis B virus “subgenotypes” defined accurately? J Clin Virol. 2010 Apr 1;47(4):356–60.
83. Toyé RM, Cohen D, Pujol FH, Sow-Sall A, Lô G, Hoshino K, et al. Hepatitis B Virus Genotype Study in West Africa Reveals an Expanding Clade of Subgenotype A4. Microorganisms. 2021 Mar 17;9(3):623.
84. Assih M, Ouattara AK, Diarra B, Yonli AT, Compaore TR, Obiri-Yeboah D, et al. Genetic diversity of hepatitis viruses in West-African countries from 1996 to 2018. World J Hepatol. 2018 Nov 27;10(11):807–21.

85. Forbi JC, Vaughan G, Purdy MA, Campo DS, Xia G liang, Ganova-Raeva LM, et al. Epidemic History and Evolutionary Dynamics of Hepatitis B Virus Infection in Two Remote Communities in Rural Nigeria. PLOS ONE. 2010 Jul 19;5(7):e11615.
86. Liu H, Shen L, Zhang S, Wang F, Zhang G, Yin Z, et al. Complete genome analysis of hepatitis B virus in Qinghai-Tibet plateau: the geographical distribution, genetic diversity, and co-existence of HBsAg and anti-HBs antibodies. Virol J. 2020 Jun 12;17(1):75.
87. Müller NF, Rasmussen D, Stadler T. MASCOT: parameter and state inference under the marginal structured coalescent approximation. Kelso J, editor. Bioinformatics [Internet]. 2018 May 22 [cited 2021 Sep 16]; Available from: https://academic.oup.com/bioinformatics/advance-article/doi/10.1093/bioinformatics/bty406/5001387
88. Hong SL, Dellicour S, Vrancken B, Suchard MA, Pyne MT, Hillyard DR, et al. In Search of Covariates of HIV-1 Subtype B Spread in the United States—A Cautionary Tale of Large-Scale Bayesian Phylogeography. Viruses. 2020 Feb;12(2):182.

SUPPLEMENTARY FIGURES AND TABLES

Supplementary Table 1. Socio-demographic and clinical characteristics of the 133 HBV-infected individuals in our study.

Socio-demographic characteristics	Men (%)	96 (72%)
	Age (years) median [IQR]	30 [25-37]
Serology	HIV co-infection (%)*	19 (14%)
	HCV co-infection (%)*	3 (2%)
	Positive eAg (%)	78 (58.65%)
	Positive eAb (%)	63 (47.37%)

	Positive Anti-HBs (%)**	3 (2.34%)	
Genotypes (n, %)	A (55, 41.35%)	A1	33 (24.7%)
		A2	1 (0.75%)
		A3 Quasi-subgenotype	14 (10.52%)
		A6	7 (5.26%)
	D (8, 6.0%)	D1	5 (3.75%)
		D2	2 (1.5%)
		D7	1 (0.75%)
	E (67, 50.25%)	E	67 (50.25%)
	Recombinant forms (3, 2.25%)	A/E	1 (0.75%)
		D/E	2 (1.5%)

*One study participant was not tested for both HIV and HCV

** Based on 128 cases, data not available for 5 individuals

Supplementary Table 2. Frequencies of HBV genotypes and subgenotypes according to the countries of origin.

Country of origin	HBV Genotype and subgenotypes (%)										Total (%)
	A (n=55, 41.4%)					D (n= 8, 6%)			E (n=67, 50.4%)	Recombinant forms (n=3, 2.2%)	
	A1	A2	A3*	A5*	A6	D1	D2	D7			
Angola	0	0	1	0	0	0	0	0	0	0	1 (0.75)
Burkina Faso/Togo	0	0	0	0	0	0	0	0	1	0	1 (0.75)
Burundi	3	0	0	0	0	0	0	0	1	0	4 (3.0)
Cameroon	0	0	2	0	0	0	0	0	8	0	10 (7.52)
Congo	15	1	3	2	7	2	1	1	12	0	44 (33.0)
Ghana	0	0	0	0	0	1	0	0	3	0	4 (3.0)
Guinea	0	0	1	0	0	0	0	0	19	1	21 (15.75)
Ivory Coast	0	0	1	0	0	0	0	0	6	0	7 (5.26)

Mauritania	0	0	1	0	0	0	0	0	5	0	6 (4.51)
Mozambique	0	0	0	0	0	0	0	0	1	0	1 (0.75)
Niger	0	0	0	0	0	1	0	0	4	0	5 (3.75)
Nigeria	0	0	1	0	0	0	0	0	2	0	3 (2.26)
Rwanda	15	0	1	0	0	1	1	0	3	2	23 (17.25)
Senegal	0	0	0	0	0	0	0	0	1	0	1 (0.75)
Zimbabwe	0	0	0	0	0	0	0	0	1	0	1 (0.75)
Unknown	0	0	1	0	0	0	0	0	0	0	1 (0.75)
Total (%)	33 (24.7)	1 (0.75)	12 (9.0)	2 (1.5)	7 (5.26)	5 (3.75)	2 (1.5)	1 (0.75)	67 (50.25)	3 (2.25)	133 (100%)

*Cluster as Quasi-subgenotypes A3

Supplementary Table 3. Frequency of drug resistance mutations in *core*, *pol* and *X* genes and association with genotypes.

Gene	Mutation	n (%)	Genotype* OR [95% CI]		
			A	D	E
core	1653	15 (11.28)	0.317 [0.085-1.183]	-	4.581 [1.229-17.08]*
	1762	40 (30.08)	0.422 [0.188-0.944]*	1.427 [0.324-6.281	2.355 [1.092-5.076]*
	1764	49 (36.84)	0.965 [0.471-1.975]	1.777 [0.424-7.452]	0.915 [0.452-1.852]
	1766	18 (13.53)	2.535 [0.914-7.028]	-	0.442 [0.155-1.260]
	1858	7 (5.26)	3.8 [0.709-20.352]	-	0.151 [0.017-1.295]*
	1896	34 (25.56)	0.087 [0.025-0.304]*	1.819 [0.410-8.054]	5.689 [2.26-14.321]*
	1899	20 (15.04)	0.125 [0.027-0.567]*	1.981 [0.37-10.593]	4.862 [1.529-15.459]*
	1653+1762	12 (9.02)	0.112 [0.014-0.901]*	-	12.767 [1.598-102-003]*
	1653+1764	12 (9.02)	0.112 [0.014-0.901]	-	12.767 [1.598-102.003]*
	1762+1764	38 (28.57)	0.472 [0.210-1.060]	1.542 [0.349-6.802]	2.073 [0.956-4.491]
	1762+1766	5 (3.75)	0.943 [0.152-5.842]	-	1.5 [0.242-9.28]
	1764+1766	10 (7.51)	6.468 [1.316-31.768]*	-	0.095 [0.011-0.78]*
	1858+1762	3 (2.26)	2.905 [0.256-32.867]	-	0.484 [0.042-5.479]
	1858+1764	4	4.442	-	-

		(3.01)	[0.449-43.882]		
	1858+1766	4 (3.01)	1.433 [0.195-10.501]	-	0.318 [0.032-3.139]
	1896+1899	14 (10.53)	0.092 [0.011-0.730]*	1.230 [0.14-10.807]	6.981 [1.487-32.557]*
	1653+1762+1764	12 (9.02)	0.112 [0.014-0.901]*	-	12.767 [1.598-102.003]
	1762+1764+1766	3 (2.26)	2.905 [0.256-32.867]	-	0.484 [0.042-5.479]
	1762+1764+1858	2 (1.50)	-	-	-
pol	173	2 (1.50)	-	-	-
	180	7 (5.26)	3.8 [0.709-20.352]	8 [1.278-50.039]*	-
	204	17 (12.78)	0.745 [0.258-2.154]	2.444 [0.451-13.231]	1.125 [0.405-3.118]
	180+204	7 (5.26)	3.8 [0.709-20.352]	8 [1.278-50.039]*	-
	180+204+173	2 (1.50)	-	-	-
X	127	20 (15.04)	0.729 [0.27-1.965]	3.811 [0.833-17.427]	0.982 [0.379-2.542]
	130	43 (32.33)	0.423 [0.193-0.926]*	2.205 [0.524-9.275]	1.827 [0.873-3.826]
	131	47 (35.34)	0.942 [0.457-1.942]	1.906 [0.454-8.002]	1.043 [0.512-2.124]
	127+130	17 (12.78)	0.991 [0.352-2.789]	4.757 [1.024-22.09]	0.653 [0.232-1.834]
	127+131	17 (12.78)	0.991 [0.352-2.789]	4.757 [1.024-22.09]	0.653 [0.232-1.834]
	130+131	39 (29.32)	0.446 [0.199-1.0]*	2.571 [0.609-10.851]	1.897 [0.885-4.066]

	127+130+131	17 (12.78)	0.991 [0.352-2.789]	4.757 [1.024-22.09]	0.653 [0.232-1.834]

Supplementary Table 4. Frequency of point mutations at different ORFs of Large S genes of different genotypes.

Genotypes		**PreS1**				**PreS2**				***MHR***			
		No.	**%**	**Del**	**Stop**	**No.**	**%**	**Del**	**Stop**	**No.**	**%**	**Del**	**Stop***
A	(55, 42.5%)	37	67.2	4	1	40	72	2	1	40	73	-	3
D	(8, 6.0%)	4	50	0	-	7	87.5	-	-	7	87.5	1	3
E	(67,51.5%)	47	70	4	1	49	73	15	2	59	88	5	5
Total	(130,100%)	**88**		**8**	**2**	**96**		**17**	**3**	**106**		**6**	**11**

Del: the number of deletions detected in different ORFs

*: detection of stop codon at downstream of MHR

Supplementary Table 5. Distribution of sequences used in this study by continent.

	HBV-A	HBV-D	HBV-E
Africa	54	53	159
Asia	63	449	-
Europe	234	118	73
North America	187	56	-
Oceania	-	64	-
South America	49	29	2

Supplementary Table 6. Distribution of sequences used in this study by country.

	HBV-A	HBV-D	HBV-E
Angola			16
Argentina	19	14	
Belarus	3	9	
Belgium	158	55	73
Brazil	26	15	
Cameroon	17		5
Canada		6	
China		22	
Colombia	4		2
Democratic Republic of the Congo			4
Fiji		4	
France	2		
Gabon	3		
Germany	10		
Ghana		1	5
Greenland		10	

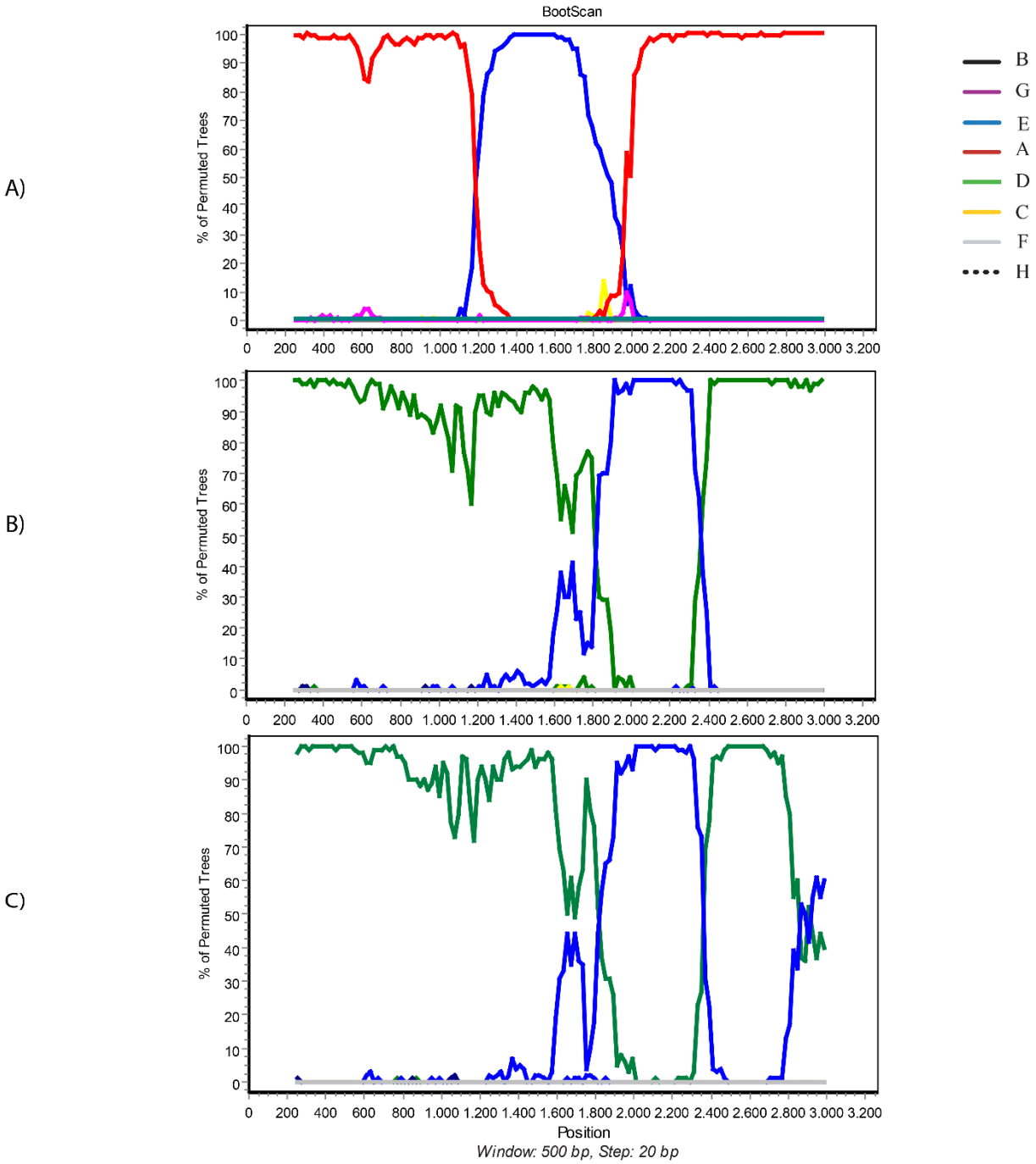
Guinea			77
Haiti	57	5	
Hungary	1		
India	10	84	
Indonesia		8	
Iran		165	
Italy		1	
Japan	50	25	
Kazakhstan		5	
Kenya	4		
Kiribati		8	
Kyrgyzstan		1	
Latvia	4	5	
Lebanon		42	
Malawi	1		
Malaysia	1	1	
Mongolia	1	14	
New Caledonia		5	
New Zealand		42	

Niger	3		
Nigeria	3		50
Pakistan		4	
Panama	1		
Papua New Guinea		2	
Poland	45	8	
Russia	8	16	
Samoa		2	
Serbia	2	5	
Slovakia	1		
South Africa	21		
Sudan		2	2
Sweden		14	
Syria		69	
Tonga		1	
Tunisia	2	50	
United Kingdom		1	
United States	129	35	

Uzbekistan	1	5	

Supplementary Table 7. Distribution of sequences used in this study by region. These geographic divisions were used for phylogeographic inference.

	HBV-A	HBV-D	HBV-E
Africa	54	53	159
Americas	236	287	2
East/South Asia	62	226	-
Europe	234	118	73
West/Central Asia	1	85	-



Supplementary Figure S1. BootScan analysis of HBV isolates. We analyzed the complete genomes of the MB-58, MB-92, and MB-56 isolates of one patient from Guinea and two patients from Rwanda. The BootScan results reveal A) recombination between HBV genotypes A and E in one strain (MB-58), and in B) and C), recombination between HBV genotypes A and E in two other strains (MB-92 and MB-56, respectively).