UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

HISTÓRIA EVOLUTIVA DO VÍRUS DA HEPATITE B EM POPULAÇÕES NATIVAS AMERICANAS

BIBIANE ARMILIATO DE GODOY

Dissertação submetida ao Programa de Pós-Graduação em Genética e Biologia Molecular da UFRGS como requisito parcial para a obtenção do grau de Mestre em Genética e Biologia Molecular.

Orientador: Prof. Dr.Nelson Jurandi Rosa Fagundes

Porto Alegre

Março de 2014

INSTITUIÇÕES FINANCIADORAS

Este trabalho teve como fontes financiadoras: o Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), a Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) e o Instituto Nacional de Genética Médica Populacional (INAGEMP).

Todo o procedimento laboratorial foi realizado em colaboração com o Laboratório de Gastroenterologia e Hepatologia Tropical, coordenado pelo Dr. João Renato Rebello Pinho, no Instituto de Medicina Tropical de São Paulo, Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo. A análise dos resultados foi desenvolvida no Laboratório de Genética Humana e Evolução, Departamento de Genética da Universidade Federal do Rio Grande do Sul.

O aluno recebeu bolsa de estudos concedida pelo CNPq, vinculada ao PPGBM/UFRGS.

AGRADECIMENTOS

Ao meu orientador Nelson, por ter aberto as portas deste projeto para mim, trazendo o ‘mundo’ intrigante dos vírus e, com este, muitos novos desafios e possibilidades; também, por todo o conhecimento e motivação constantemente irradiados e pela incessante forma de questionar os fatos e gerar novas ideias. É, com certeza, um exemplo de profissional que levarei comigo sempre.

Ao Prof. Salzano, responsável pelo desenvolvimento de uma bela história de pesquisa com Nativos Americanos, pela contínua cordialidade, pelo exemplo inspirador de pesquisador que é e por permitir que esse trabalho se concretizasse.

A todos os colegas do Laboratório de Genética Humana e Evolução, pelo intercâmbio de conhecimento e por tornarem todos esses anos de convívio muito mais divertidos. Em especial à Dana, pelos inúmeros ensinamentos e por representar meu primeiro vínculo com o grupo, e ao querido Lucas pela amizade, parceria intelectual e auxílio em diversos momentos.

Ao grupo de pesquisa do Laboratório de Gastroenterologia e Hepatologia Tropical, em especial à Michele, pela ótima recepção que me deram e pela paciência ao me apresentar todo o trabalho de excelência realizado pela equipe.

Às amigas de ‘Brasiliano’ e à Karina, pelo apoio e paciência em todos os momentos e por tornarem meus dias muito mais coloridos, mostrando que o caminho se torna mais fácil quando temos pessoas especiais por perto.

À minha família, meus incansáveis incentivadores, pelo exemplo de caráter e persistência que norteiam cada um dos meus passos e por serem, antes de qualquer coisa, os melhores amigos que eu poderia ter. Nenhuma conquista seria possível sem seu o apoio e carinho.

SUMÁRIO

LISTA DE ABREVIATURAS

HBV: Vírus de Hepatite B

ORF: fase aberta de leitura

HBcAg: antígeno core

HBsAg: antígeno S de superfície

HBeAg: antígeno e

HBx: proteína regulatória X

cccDNA: molecular circular fechada covalentemente

ALT: alanina aminotransferase

HDV: Vírus de Hepatite D

RESUMO

Introdução: O Vírus da Hepatite B (HBV) é um vírus de DNA com tropismo por hepatócitos, que apresenta um genoma circular parcialmente dupla fita. Baseado na divergência de sequência do genoma completo, dez linhagens evolutivas, denominadas "genótipos" de HBV, foram descritas (A-J), sendo F e H considerados como "indígenas" da América. Os genótipos de HBV apresentam uma forte estruturação geográfica, o que pode refletir padrões das migrações humanas. Na América do Sul, áreas de alto endemismo incluem a região amazônica, e as maiores taxas de infecção têm sido observadas em populações Nativas Americanas. Embora a forte estruturação geográfica seja indicativa de uma origem antiga, a maioria das análises visando datar a origem dos genótipos "americanos" F e H resulta em datações extremamente recentes que não condizem com eventos históricos relacionados ao HBV.

Objetivo: Os objetivos desse trabalho compreendem avaliar o impacto de diferentes taxas evolutivas e da seleção purificadora sobre as estimativas de datação molecular a fim de inferir quais taxas são mais condizentes com a origem do HBV na América; caracterizar o HBV circulante em uma amostra histórica de Nativos Americanos, e discutir os processos históricos que possam ser relevantes para entender os padrões observados.

Material e Métodos: Nós realizamos análise Bayesiana utilizando sequências disponíveis dos genótipos F e H e diferentes taxas evolutivas previamente reportadas, e comparamos a ocorrência de mutações sinônimas e não-sinônimas em ramos da filogenia classificados como "antigos" ou "recentes" a fim de inferir a atuação da seleção purificadora ao longo do tempo. Para caracterização do HBV presente nas populações Nativas Americanas, detecção e amplificação do DNA viral foi obtida através de PCR seguido de sequenciamento e análise filogenética. Análise Bayesiana de *Skyline Plot* foi realizada para comparar a dinâmica populacional do subgenótipo A1 presente na amostra de Nativos Americanos e em outras cepas isoladas no Brasil.

Resultados e conclusão: Nossos resultados mostram que as estimativas de datação molecular são fortemente influenciadas pelas taxas evolutivas utilizadas na análise. Além

disso, foi observado excesso de mutações não-sinônimas nos ramos recentes da filogenia, o que é compatível com a ocorrência de seleção purificadora, e pode gerar um viés sobre as estimativas, produzindo datações recentes demais. Na amostra de Nativos Americanos, nós constatamos o predomínio do subgenótipo A1, relacionado com populações africanas. Análise de *Skyline Plot* mostrou que a expansão populacional nas cepas isoladas de Nativos Americanos é mais recente que aquela inferida para outras cepas brasileiras, sugerindo que os processos históricos que contribuíram para a formação do subgenótipo A1 dos Nativos Americanos são relacionados com ondas migratórias mais recentes em direção à região amazônica.

ABSTRACT

Introduction: Hepatitis B virus (HBV) is a hepatotropic DNA virus that presents a partially double-stranded circular genome. Based on sequence divergence of the complete genome, ten HBV evolutionary lineages, called "genotypes" have been described (A-J), with F and H being considered as indigenous from the Americas. HBV genotypes present a remarkable geographic structure which may reflect historic patterns of human migrations. In South America, areas of high endemism include the Amazon basin, and high prevalence rates have been observed in Native American populations. Although the strong geographical structure indicates an ancient origin, most analysis trying to date the origin of the "American" genotypes F and H result in extremely recent dates that disagree with historical events related with HBV.

Objective: The aims of this study comprise evaluate the impact of different evolutionary rates and of the purifying selection on molecular dating estimates in order to infer which rates are in better agreement with the origin of HBV in the Americas; to characterize the HBV circulating in a historical sample of Native Americans, and discussing the historical processes that might be relevant to understand the observed patterns.

Materials and Methods: We performed a Bayesian analysis using the available sequences of F and H genotypes and different evolutionary rates previously reported, and compared the occurrence of synonymous and non-synonymous mutations in branches of phylogeny classified as "old" or "young" in order to infer the effects of purifying selection over time. For the characterization of HBV from Native American populations, detection and amplification of viral DNA were obtained through PCR followed by sequencing and phylogenetic analysis. Bayesian Skyline Plot analysis was performed to compare the population dynamics of the A1 subgenotype present in the sample of Native American and in other strains isolated from Brazil.

Results and Conclusion: Our results show that molecular dating estimates are strongly influenced by the evolutionary rate assumed in the analysis. In addition, we observed an excess of non-synonymous mutations in recent branches of phylogeny, which is

compatible with the occurrence of purifying selection and may create a bias on the estimates, producing too recent datings. In the sample of Native Americans, we observed a predominance of the A1 subgenotype, related with African populations. Skyline Plot analysis showed that population expansion in strains isolated from Native Americans is more recent than that inferred from other Brazilian strains, suggesting that the historical processes that contributed to the presence of A1 subgenotype A1 Native Americans are related with more recent migratory waves towards the Amazon region.

CAPÍTULO I - INTRODUÇÃO

1. INTRODUÇÃO

1.1. Hepatite B

A hepatite B é uma infecção hepática causada pelo Vírus de Hepatite B (HBV). Sua transmissão pode ocorrer de forma horizontal – pelas vias parenteral e sexual – e de forma vertical – da mãe para o filho (Alvarado-Mora *et al*., 2013a). De modo geral, a disseminação do vírus ocorre através do contato com sangue ou outros fluidos corporais, sendo as transmissões via mãe para filho e entre irmãos as maiores responsáveis pelos casos da doença em países em desenvolvimento (revisado por Franco *et al*., 2012). As consequências do contágio envolvem desde uma infecção assintomática até infecção crônica que pode evoluir para complicações como cirrose hepática e carcinoma hepatocelular (Dandri e Locarnini, 2012). A história natural da hepatite B crônica geralmente consiste em quatro fases: 1) imunotolerante, que ocorre após contágio perinatal e se caracteriza pela replicação ativa do vírus e aumento nos níveis séricos de HBsAg, HBeAg e HBV DNA, podendo perdurar por mais de 20 anos; após esta fase se inicia a fase 2) imunoativa, na qual o sistema imune começa a reconhecer os epitopos virais expressos pelos hepatócitos, ativando a resposta imune celular citotóxica e, posteriormente, a humoral, promovendo intensa ação inflamatória no fígado, o que ocasiona aumento da alanina aminotransferase (ALT) e diminuição nos níveis séricos de HBV DNA; a transição desta fase para a fase 3) não replicativa é denominada soroconversão, onde ocorre a diminuição dos níveis de HBeAg, HBV DNA e ALT. Muitos portadores permanecem nesta fase durante toda a vida, entretanto alguns indivíduos passam por fases de 4) reativação na qual a replicação viral se reinicia podendo gerar consequências mais graves (Danri e Locarnini, 2012). A probabilidade de progressão para hepatite crônica está relacionada com a idade no período da infecção, sendo mais frequente em indivíduos infectados no período perinatal (90%), quando comparados a infecções ocorridas entre 1 e 4 anos (30%) ou na idade adulta (5%) (Ott *et al*., 2012).

A disponibilidade de uma vacina efetiva, desenvolvida em 1982, produziu uma significativa redução nas taxas globais de infecção, mortalidade de crianças com hepatite fulminante, taxas de portadores crônicos e incidência de carcinoma hepatocelular na infância (Chang *et al*., 1997; Kao *et al*., 2001; Chang *et al*., 2009). Entretanto, mais de 360 milhões de pessoas estão cronicamente infectadas por HBV e cerca de um milhão de

pessoas morre por ano em decorrência de patologias hepáticas associadas à infecção, o que mantém a hepatite B como um problema de saúde importante em nível global (Block *et al*., 2007).

1.2. Vírus da hepatite B

O vírus da Hepatite B pertence à família *Hepadnaviridae*, cujos membros são caracterizados por acentuado tropismo por hepatócitos (Ganem e Price, 2004). Sua estrutura esférica de 42-44nm é composta por um envelope lipoprotéico, onde estão localizados os antígenos de superfície (Dane *et al*., 1970; Ganem, 1991). Dentro do envelope existe o nucleocapsídeo, ou core, que abriga o DNA do vírus. Este, por sua vez, compreende um DNA circular relaxado e parcialmente dupla-fita, cujas fitas apresentam comprimentos diferentes: a maior, de sentido negativo, possui cerca de 3.200 pb, enquanto sua complementar possui comprimento variável abrangendo de 50 a 100% da primeira. A estrutura circular do DNA é mantida através das extremidades coesivas e complementaridade existente entre ambas as fitas (Wei *et al*., 2010).

O genoma do HBV é inteiramente codificante e possui quatro fases abertas de leitura (ORFs) parcialmente sobrepostas (Block *et al*., 2007) (Figura 1):

- Pré-S/S que codifica três proteínas de superfície: L, cujo papel é importante para montagem e liberação do vírion, além de fazer a ligação do vírus aos receptores, sendo requerida para a entrada do HBV e também, em caso de coinfecção, do Vírus de Hepatite Delta (HDV) nas células do hospedeiro e (Bruss e Ganem, 1991; Klingmüller *et al*., 1993; Alvarado-Mora *et al.*, 2013c; M, de função ainda desconhecida; e S (HBsAg), a mais abundante, que constitui o envelope do nucleocapsídeo, sendo necessária para a montagem e liberação também do HDV (Alvarado-Mora *et al.*, 2013c);

- Pré-Core/Core que codifica o antígeno core (HBcAg), constituinte do nucleocapsídeo, além do antígeno secretado "e" (HBeAg);

- P que codifica a proteína terminal TP que atuará como primer para a replicação do DNA do HBV, além da polimerase viral, que possui atividade de transcriptase reversa, DNA polimerase e RNaseH (Dandri e Locarnini, 2012);

- X que codifica a proteína regulatória X (HBx), cuja função está relacionada com a expressão gênica vírus-hospedeiro, sendo requerida para a replicação viral *in vivo* (Zoulim *et al*., 1994; Lucifora *et al*., 2011).

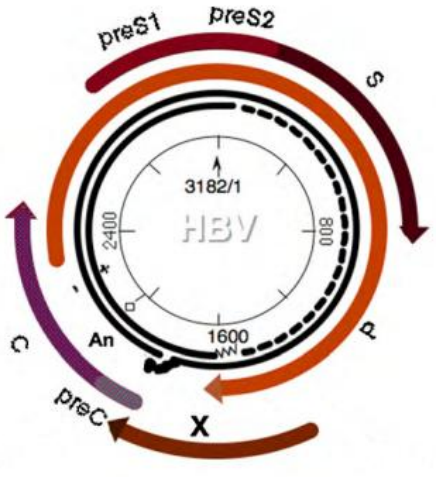


Figura 1: A organização genômica do HBV. As linhas pretas (contínuas e pontilhadas) correspondem ao DNA parcialmente dupla-fita. As setas coloridas correspondem às quatro ORFs do genoma viral que originarão todas as proteínas virais descritas anteriormente (modificado de Wei et al., 2010).

1.3. Replicação do HBV

No processo de infecção (Figura 2), o HBV se liga ao receptor específico de hepatócito e penetra na célula através de endocitose. No citoplasma, o vírus libera seu nucleocapsídeo e tem seu genoma transferido para o núcleo, onde é convertido a uma molécula circular fechada covalentemente (*covalently closed circular molecule* - cccDNA) pela maquinaria enzimática do hospedeiro. O cccDNA interage com diversas proteínas, incluindo histonas, mantendo assim uma estrutura de "minicromossomo" (Newbold *et al*., 1995; Dandri e Locarnini, 2012). Essa estrutura servirá como molde para a transcrição de todos os RNAs virais, que incluem um RNA maior que serve de molde para a replicação a partir de transcrição reversa realizada pela polimerase viral, além de dois transcritos menores que serão transportados para o citoplasma e posteriormente traduzidos, originando as demais proteínas envolvidas na estrutura do HBV completo (Wei *et al*., 2010).

A montagem das novas partículas virais se inicia com a organização do nucleocapsídeo no citosol. Posteriormente, a molécula do RNA maior é incorporada à estrutura, originando duas fitas de DNA após a transcrição reversa. A conversão do RNA longo em DNA ocorre de modo sequencial: a primeira fita (-) é feita a partir do RNA encapsidado que é degradado em seguida, enquanto a segunda fita (+) utilizará a fita (-)

recém sintetizada como molde (Pollack *et al*., 1994; Ganem e Prince, 2004). Alguns dos nucleocapsídeos contendo DNA viral maduro serão transportados de volta para o núcleo com a finalidade de manter uma quantidade suficiente de moldes transcricionais intranucleares, entretanto, a maior parte deles receberá as proteínas do envelope viral no retículo endoplasmático, formando vírions completos que serão liberados da célula (Ganem e Prince, 2004; Wei *et al*., 2010; Dandri e Locarnini, 2012). Parte destas proteínas produzidas (HBsAg e HBeAg) são utilizadas como marcadores sorológicos da doença, uma vez que a presença delas no sangue circulante é indicativa de replicação ativa do vírus. Outra maneira utilizada para diagnóstico e acompanhamento da doença é a realização de PCR, o qual permite a detecção específica de DNA viral bem como sua quantificação (carga viral). As consequências do processo de infecção vão depender fortemente da interação entre diversos fatores do vírus e do hospedeiro, originando desfechos clínicos que podem variar entre a hepatite assintomática até a insuficiência hepática (Tanwar and Dusheiko, 2012).

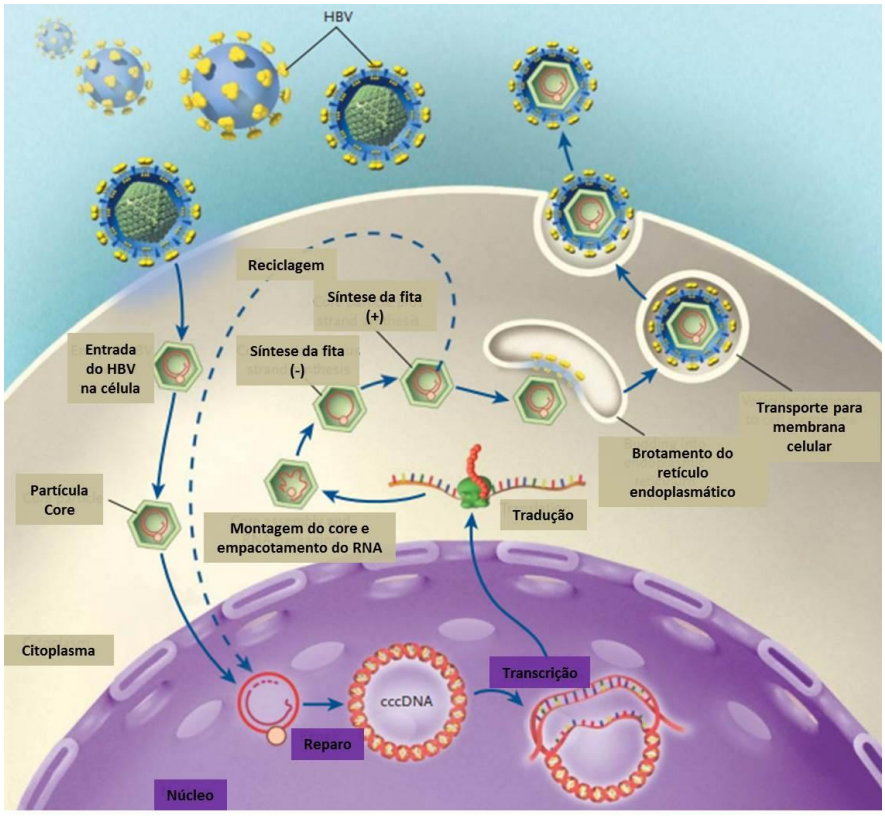


Figura 2: Representação do ciclo de replicação do HBV (Modificado de Ganem e Prince, 2004).

1.4. Epidemiologia do HBV

Levando em consideração o número de portadores de HBV, o mundo pode ser dividido em regiões de alta, intermediária e baixa prevalência (Figura 3). Áreas de alta endemia abrangem regiões onde a taxa de portadores é maior do que 8%, como ocorre no sudeste da Ásia, África subsaariana e bacia Amazônica; países do leste europeu e da bacia do Mediterrâneo são considerados como de endemia intermediária, apresentado entre 2% e 8% de prevalência para HBV; e as regiões de baixa endemia, cuja prevalência é menor que 2% e incluem os Estados Unidos, norte da Europa, Austrália e parte da América do Sul (revisado por Franco *et al*., 2012). Na América Latina, as regiões de maior prevalência abrangem o Caribe e a bacia amazônica, onde a porcentagem de portadores crônicos da doença ultrapassa os 2%. No sul da América do Sul (Chile, Uruguai, Paraguai e Argentina) e na América Central a prevalência de HBV é baixa, e a transmissão sexual representa a forma mais comum de contágio da doença (Tanaka, 2000; Parana and Almeida, 2005; Franco *et al*., 2012).

Considerando todo o continente americano, as maiores taxas de infecção têm sido observadas em populações Nativas Americanas. Estudos prévios mostram que a prevalência de HBV em Ameríndios pode ultrapassar os 30% (Torres, 1996; Echevarria e Leon, 2003; Braga, 2004). Além disso, tem sido documentada frequente ocorrência de infecção oculta (Roman *et al*., 2010; Cardona *et al*., 2011), bem como aumento na coinfecção de HBV com hepatite Delta e na incidência de hepatite fulminante na região da bacia Amazônica, sendo proposta uma possível associação entre o perfil genético presente nessas populações e a maior patogenicidade dessas infecções (Casey *et al*., 1996; Quintero *et al*., 2001; Devesa *et al*., 2007).

O sequenciamento do genoma completo de HBV levou à identificação de 10 linhagens evolutivas (A-J), denominadas como "genótipos", e assim classificados de acordo com divergências nucleotídicas maiores que 8% (Yu *et al*., 2010). Divergências intragenotípicas maiores que 4% levam a uma classificação posterior em "subgenótipos" (Alvarado-Mora e Pinho, 2013b). Com exceção dos dois mais recentemente descritos (I e J), a distribuição geográfica dos outros oito genótipos de HBV está bem estabelecida (Figura 3) (Kurbanov et *al*., 2010). Os genótipos A e D são predominantes no Velho Mundo, mas ocorrem também nos demais continentes, sendo observados em altas frequências no Brasil (Ribeiro *et al*., 2006; Devesa *et al*., 2007). Genótipos B e C são

encontrados principalmente no sudeste e extremo leste asiático, enquanto o genótipo E circula no oeste da África Subsaariana (Pujol e Devesa, 2005). O genótipo G ainda não tem uma distribuição completamente esclarecida, mas sua ocorrência foi observada nos Estados Unidos, México e Europa (Stuyver *et al*., 2000; Sánchez-Tapias *et al*., 2002). O genótipo F, considerado autóctone da América do Sul, é altamente predominante entre Nativos Americanos (Gaspar e Yoshida, 1987; Blitz *et al*., 1998; Viana *et al*., 2005). Ele é proximamente relacionado ao genótipo H, que parece restrito à América do Norte e Central, especialmente México (Norder *et al*., 1993; Arauz-Ruiz *et al*., 2002). O clado que contém esses dois genótipos (F+H) representa a linhagem mais divergente na filogenia do HBV.

A distribuição dos genótipos HBV é resultado de muitos aspectos históricos das regiões, refletindo padrões de migração das populações humanas (Campos *et al*., 2005). Por exemplo, na América do Norte, Europa e Austrália os genótipos A, B, C e D têm sido notificados em prevalências iguais, devido ao alto número de ondas migratórias para esses países (Chu *et al*., 2003; Kato *et al*., 2004). Isso é observado também com o subgenótipo A1, relacionado a populações africanas, mas que ocorre em diversas regiões do Brasil em decorrência do tráfico de escravos que resultou na imigração de milhões de africanos para o país (Motta-Castro *et al*., 2008; Santos *et al*. 2010). E ainda para o genótipo D, que ocorre em altas frequências no sul do Brasil devido à imigração italiana no final do século XIX e início do século XX (Campos *et al*., 2005).

Além da relação com aspectos históricos, a caracterização dos genótipos de HBV é importante para o melhor entendimento dos diferentes desfechos da doença (Tanwar e Dusheiki, 2012). Estudos prévios sugerem que os genótipos de HBV respondem de maneiras diferentes às terapias disponíveis (revisado por Tanwar e Dusheiko, 2012). Além disso, tem sido observada relação entre os genótipos e a severidade da doença, bem como com respostas diferenciadas a coinfecções (Sánchez-Tapias *et al*., 2002; Livingston *et al*., 2007; Kiesslich *et al*., 2009).

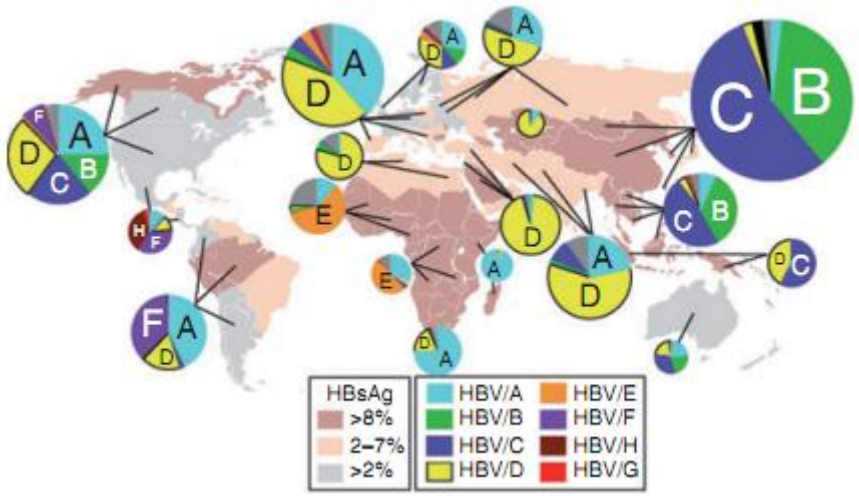


Figura 3: Prevalência de HBV e porcentagem de distribuição dos seus respectivos genótipos, com exceção dos mais recentemente descritos I e J (Kurbanov *et al*., 2010).

1.5. História evolutiva

Diferentes hipóteses já foram propostas para explicar a origem do HBV. A primeira delas sugere uma origem no Novo Mundo e posterior disseminação para os demais continentes há 400 anos atrás, após o período de colonização dos europeus, sendo sua principal evidência a posição filogenética dos genótipos F e H, considerados originários no Novo Mundo, e classificados como grupo irmão dos demais genótipos HBV de humanos e grandes macacos (Simmonds, 2001); a segunda propõe que HBV estava presente no ancestral comum entre os primatas do Velho Mundo e os macacos do Novo Mundo, ocorrendo uma co-especiação por volta de 35-10 milhões de anos atrás seguida de diversos eventos de transmissão entre espécies de primatas (Tatematsu, 2009); e a última sugere que HBV poderia estar presente entre os hominídeos que saíram da África por volta de 60-70 mil anos atrás (Norder *et al*., 1994; Magnius *et al*., 1995; Soares *et al*., 2012). Para os genótipos "americanos" F e H, seguindo a distribuição genotípica associada ao HBV, Arauz-Ruiz *et al.* (2002) propõem uma origem ainda na Ásia e posterior chegada na América junto aos primeiros colonizadores, entretanto análises visando datar a origem dos diferentes genótipos de HBV variam amplamente de acordo com as taxas evolutivas utilizadas (Godoy et al., 2013), dificultando uma melhor compreensão sobre a história evolutiva desse vírus em populações humanas.

A existência de uma forte estruturação geográfica na distribuição dos genótipos HBV, bem como o fato de determinados genótipos acompanharem perfis étnicos específicos - como se observa nos genótipos "Nativos Americanos" F e H, associados com populações Nativas Americanas; ou com o subgenótipo A1, associado com populações africanas e cuja distribuição corrobora parte dos registros históricos relacionados ao tráfico de escravos - fornecem fortes evidências de que HBV é um vírus de origem relativamente antiga que está segregando dentro das populações humanas há bastante tempo (Magiorkinis *et al*., 2005; Devesa *et al*., 2007; Schaefer, 2007; Andernach *et al*., 2009). Entretanto, análises visando datar a origem do HBV através de relógio molecular têm fornecido intervalos de tempo extremamente recentes que não condizem com os dados históricos associados à distribuição dos genótipos virais (Zhou *et al*., 2007; Alvarado-Mora *et al*., 2010; Torres *et al*., 2011). As taxas de substituição, pontos de calibração e consequentes taxas evolutivas determinadas para HBV variam bastante entre os autores (Osiowy *et al*., 2006; Torres *et al*., 2011; Paraskevis *et al*., 2013). Entre os fatores que influenciam na variação das taxas de substituição do HBV está a existência de uma polimerase viral sem atividade de correção de erro que poderia tornar as taxas de variação do HBV similares a de vírus de RNA e, consequentemente, maiores do que de outros vírus de DNA. Além disso, a presença de ORFs sobrepostas implica que a ocorrência de uma substituição sinônima em uma das ORFs pode potencialmente resultar em uma substituição não sinônima na ORF sobreposta, gerando taxas de substituição que podem variar entre regiões sobrepostas e não sobrepostas do genoma, dificultando as estimativas (Mizokami e Orito, 1999; Torres *et al.*, 2011).

Estudos recentes têm mostrado também que os pontos de calibração ocasionam grande variação nas taxas evolutivas, uma vez que a utilização de pontos de calibração muito recentes geram taxas rápidas demais e pontos de calibração muito antigos ocasionam o oposto (Ho *et al*., 2011; Crandall *et al*., 2012). A variação nas taxas evolutivas em diferentes escalas temporais poderia ser causado, entre outros fatores, pela existência de seleção purificadora sobre mutações não-sinônimas recentes que não permanecem na população por muito tempo, mas que inflam as estimativas baseadas em calibrações recentes (Kivisild *et al.*, 2006; Ho *et al.*, 2011). Esses trabalhos têm sugerido também que, idealmente, deveriam ser utilizados vários pontos de calibração, além de sequências datadas (isto é, amostradas no passado) de modo a abranger um intervalo de tempo

suficiente para a realização de adequada estimativa das taxas evolutivas (Firth *et al*., 2010). Nesse sentido, o estudo de amostras históricas ou antigas de HBV poderia contribuir na estimativa de taxas evolutivas mais razoáveis para esse vírus. Além disso, até o momento, nenhum estudo foi realizado com o HBV para avaliar se a seleção purificadora também pode afetar a estimativa das taxas evolutivas em diferentes escalas temporais.

CAPÍTULO II – JUSTIFICATIVA E OBJETIVOS

2. JUSTIFICATIVA E OBJETIVOS

2.1. Justificativa

A realização de um estudo descrevendo a prevalência do HBV e caracterizando geneticamente as linhagens virais circulantes entre populações Nativas Americanas teria grande importância pelo fato de contribuir para o aumento do conhecimento sobre o estado de endemismo dessa doença em populações nativas, além de auxiliar no entendimento da distribuição dessas linhagens e, principalmente, para guiar políticas de saúde voltadas para o controle dessa doença na região Amazônica e entre populações Nativas Americanas. Além disso, a oportunidade de estudar uma amostra altamente representativa das populações Nativas Americanas fornece a possibilidade de esclarecer questões históricas a respeito das populações amostradas, bem como contribuir para melhor elucidação da origem e história evolutiva do HBV na América do Sul.

2.2. Objetivos

2.2.1. Objetivo geral

Caracterizar o HBV circulante entre as populações Nativas Americanas através da descrição da prevalência e dos genótipos observados a fim de entender melhor a distribuição e origem do HBV na América do Sul e os processos históricos relacionados às populações amostradas.

2.2.2. Objetivos específicos

- Avaliar se há evidência da ação da seleção purificadora afetando a estimativa das taxas evolutivas do HBV.
- Relatar a prevalência de infecção por HBV nas populações Nativas Americanas amostradas, caracterizando os genótipos observados;
- Realizar a análise filogenética das sequências obtidas para auxiliar no melhor entendimento da história evolutiva do HBV na América do Sul.
- Associar a distribuição das linhagens virais encontradas com aspectos culturais e históricos.

CAPÍTULO III – ARTIGO 1

ORIGIN OF HBV AND ITS ARRIVAL IN THE AMERICAS – THE IMPORTANCE OF NATURAL SELECTION ON TIME ESTIMATES

ARTIGO PUBLICADO NA REVISTA ANTIVIRAL THERAPY.

Original article

Origin of HBV and its arrival in the Americas – the importance of natural selection on time estimates

*Bibiane A Godoy[1,2], Mónica V Alvarado-Mora[3], Michele S Gomes-Gouvêa[3], João R Rebello Pinho[3], Nelson JR Fagundes[1,2]**

[1]Genetics Department, Federal University of Rio Grande do Sul, Porto Alegre, Brazil
[2]National Institute of Medical Population Genetics, Porto Alegre, Brazil
[3]Laboratory of Tropical Gastroenterology and Hepatology 'João Alves de Queiroz and Castorina Bittencourt Alves', Institute of Tropical Medicine, Department of Gastroenterology, University of São Paulo School of Medicine, São Paulo, Brazil

*Corresponding author e-mail: nelson.fagundes@ufrgs.br

Background: The strong geographic structure shown by the global pattern of HBV lineages suggests an ancient origin for this virus; however, estimates based on the molecular clock suggest a very recent origin for the Native American genotypes F and H. In this study, we contribute to this debate by estimating the divergence times of genotypes F and H and by discussing how evolutionary rates estimated from recent samples may underestimate the divergence time of more ancient nodes in HBV phylogenies.
Methods: A total of 108 complete HBV genotype F and H genomes were compared to 44 reference genomes from other genotypes. Time estimates were based on a Bayesian method with evolutionary rates taken from the literature. To assess the pattern of substitutions in recent versus old branches we mapped the phylogenetic distribution of all mutations occurring in genotypes F and H using a maximum likelihood approach and compared the number of synonymous and non-synonymous mutations in young and old branches of HBV genotype F and H phylogeny using a χ^2 test.
Results: Estimated divergence times between genotypes F and H depend heavily on the evolutionary rate. While fast rates suggest a recent separation of these genotypes (approximately 800 years ago), slow rates suggest an earlier divergence (up to approximately 13,000 years ago). There is a clear excess of non-synonymous substitutions in the most recent branches of HBV phylogeny ($P=4.87\times10^{-15}$), most likely suggesting the action of purifying selection.
Conclusions: These results suggest that rates estimated based on recent samples will overestimate the evolutionary rate and underestimate the coalescence times for ancient nodes in HBV phylogeny.

Introduction

HBV is an important human pathogen affecting millions of individuals worldwide [1]. HBV is a DNA virus from the *Hepadnaviridae*, and has a strong phylogeographic structure, with 10 known evolutionary lineages (from A to J), typically referred to as 'genotypes', segregating in human populations [2]. In addition to such strong phylogenetic structure, HBV also presents a remarkable geographic structure, with specific genotypes being typical from specific geographic regions [3]. For example, genotypes F and H are considered as typically 'American' and are usually associated with Native American populations or Native American ethnic backgrounds [4–6]. Although recent migration events may blur a clear-cut geographical association between HBV lineages and geographic origin, several studies associate the frequency of certain HBV genotypes to more ancient migration events. For example, the finding of the (African) subgenotype A1 in Brazilian rural population isolates descending from African slaves brought to the country during colonial times is interpreted as being the result of the African slave trade during the 16th to 19th centuries [7,8]. Another example of a possible role of African slave trade for determining HBV distribution in the Americas is the finding of HBV genotype E in a poor rural Colombian community [9]. Likewise, the various prevalences of HBV genotypes in Argentina were interpreted as reflecting contributions of people from different ethnic backgrounds [10].

Despite such compelling historical accounts, which would suggest that HBV is a relatively old virus segregating within human populations for a long time, molecular

clock estimates for the origin of HBV and its genotypes provide a very different scenario for HBV evolution. Dating the most recent common ancestor for Colombian and African lineages of HBV genotype E resulted in a surprisingly recent date of 35 years ago [9]. Other studies estimated a very recent origin for the 'Native American' HBV genoypte F, of approximately 284 years ago [11], and an even more extreme date of 229 years ago has been found for the origin of HBV itself [12]. In common, all these studies use extremely sophisticated methods and software to estimate HBV evolutionary times based on dated samples collected recently [11,12] or data from intra-familiar transmissions [13]. Thus, while the geographic structure of HBV genotypes would be consistent with HBV being a relatively old virus with a moderate evolutionary rate, molecular clock studies suggest a very recent origin for HBV, with very fast evolutionary rates associated with it.

Recently, it has been suggested that time estimates based on molecular clocks may be highly dependent on the choice of the calibration of the evolutionary rate [14–18]. Evolutionary rates estimated from recent calibration points usually produce faster rates, while calibration with fossil data usually results in slower rates. Problems arise when evolutionary rates calibrated based on old nodes in the phylogeny are used to estimate times in recent nodes in the phylogeny and *vice versa*. One possible mechanism accounting for this phenomenon is purifying selection acting on deleterious substitutions that are still segregating in recent nodes of the phylogeny, but that would have been eliminated in more ancient nodes [15]. There is evidence that young human mitochondrial DNA haplogroups have an excess of non-synonymous substitutions as compared to old haplogroups [19], consistent with the purifying selection hypothesis if most non-synonymous mutations are deleterious or slightly deleterious. Therefore, if HBV evolutionary rates calibrated from recent samples are inflated, they would underestimate the deeper nodes in HBV phylogeny and may explain, at least partially, the discrepancy between 'molecular clock' and 'anthropological' accounts of HBV evolution.

In this study we compared the effect of different priors on the evolutionary rate to discuss how they impact divergence time estimates about the origin of genotypes F and H. We then compared the timing and distribution of HBV lineages in the Americas to discuss whether or not HBV is of pre-Columbian origin and if we could speculate its origins among the first American settlers who arrived in the Americas >15,000 years ago. Finally, we tested the hypothesis that purifying selection plays an important role in HBV evolution by comparing the number of synonymous and non-synonymous substitutions in lineages belonging to genotypes F and H and discuss its impact on the estimates of an evolutionary rate for HBV.

Methods

We retrieved from GenBank sequences of 126 complete HBV genomes belonging to genotypes F and H. Because this dataset was also used in the analyses of natural selection, we excluded all sequences showing deletions, as they represent frame-shift mutations or long deletions of likely deleterious effects. Identical sequences were also excluded. The final dataset for HBV genotypes F and H consisted of 108 sequences (Additional file 1). For the phylogenetic analysis we also used a dataset of 44 complete genomes from HBV isolates belonging to other genotypes (A–E, G, I; Additional file 1).

To estimate the time of key events in HBV evolution we used the phylogenetic method implemented in BEAST version 1.6.2 [20]. More specifically, we estimated the time to the most recent common ancestor (TMRCA) of all F subgenotypes, the TMRCA of genotypes F and H, the TMRCA of the clades F and H and the root of the tree, which represents the divergence of genotypes F and H from the other Old World genotypes. As the full dataset included only different sequences, we used a Yule model as the tree prior. The Yule model is a birth-only process and it is an efficient way of describing a tree which is not based on population sampling, in which case a coalescent-based prior would be preferable. To evaluate the effect of the choice of the evolutionary rate and type of molecular clock (whether strict or relaxed [21]) different analyses were performed: a strict molecular clock using the fast evolutionary rate estimated by Torres *et al.* [11] using dated samples (1.67×10^{-4}); a strict clock using a slow rate of 1.0×10^{-5}, as discussed by Torres *et al.* [11]; a strict clock using a range of 'likely' rates suggested by Torres *et al.* [11], between 2.0×10^{-5} and 6.0×10^{-6}; a lognormal relaxed clock using a fixed mean rate of 1.0×10^{-5}; and a lognormal relaxed clock using a range of values between 2.0×10^{-5} and 6.0×10^{-6} as above. The prior for range of most 'likely' values was modelled using a truncated normal prior with a mean of 1.0×10^{-5} and a standard deviation of 5.0×10^{-6}, with minimum and maximum bounds set at 2.0×10^{-5} and 6.0×10^{-6}, respectively. For all analyses we ran 25,000,000 steps of the Markov Chain Monte Carlo sampling every 1,000 steps, discarding the first 5,000 samples as burn-in.

To evaluate the effect of natural selection on different parts of HBV genotype F and H phylogeny and the possible implications on inferred evolutionary rates, we estimated a maximum likelihood (ML) tree in PAUP 4.0 [22] using the TVM+I+G model with values obtained from Modeltest 3.01 [23]. We then ran PAUP 4.0 to map all substitutions occurring in the HBV genotype F and H dataset across all branches of the estimated ML tree. No

bootstrap was performed for this analysis as we wanted the most likely tree estimate (under the likelihood criterion) to map mutations directly in each branch under the same criterion. Nonetheless, the general robustness of HBV phylogeny estimated from this dataset can be evaluated using the clade posterior probabilities provided by the Bayesian tree estimate. Mutations were then assigned as non-synonymous or synonymous depending on their effect on the affected codon using DnaSP version 5.0 [24]. Complex codons (that is, codons affected by multiple substitutions that could be either synonymous or non-synonymous depending on their ancestral state) were checked manually. Because the HBV genome presents overlapping genes, we initially considered a mutation affecting two genes simultaneously as doubly non-synonymous, doubly synonymous, or as simultaneously synonymous and non-synonymous, depending on the exact effect it produces on the HBV genome. However, to be certain that this approach did not introduce any bias to our results we also performed alternative analyses using only genome positions occurring on non-overlapping regions or considering each gene separately.

'Young' branches were defined as those occurring within genotypes (or subgenotypes in the case of F1 and F2), whereas 'old' branches were defined as those connecting genotypes among each other. The branches connecting subgenotypes F1a and F1b to their common ancestors, and those connecting F2a and F2b to their common ancestors were considered 'intermediate' and were therefore excluded from the initial analyses (Figure 1). To test if our results were robust to this choice, we also re-analysed our data considering these branches as either 'young' or 'old'. This categorization of branches according to their age, despite being simple, is a robust way of dealing with phylogenetic uncertainty, because all genotypes and subgenotypes are monophyletic and one can therefore be certain that branches occurring within these categories are younger than those defining viral genotypes. The number of non-synonymous and synonymous mutations between 'young' and 'old' branches were then compared using a χ^2 test in the programme PEPI 4.0 [25]. Under the null hypothesis that the action of natural selection is similar between these groups of branches this test should result in statistically non-significant values.

Results

Time estimates for genotypes H and F, and subgenotypes of F

Not unexpectedly, the estimated date for the origin and divergence of HBV genotypes and subgenotypes is highly dependent on the evolutionary rate assumed. The root of the HBV tree can be dated as recently as approximately 800 years ago or as old as approximately 13,300 years ago, depending on the evolutionary rate assumed (Table 1), and represents the divergence between the clade formed by the Native American genotypes (F and H) and a clade formed by the remaining HBV genotypes. Importantly, depending on specific assumptions on the molecular clock model, the 95% CI included dates older than 12,000 years ago for all analyses except those based on the fast rate estimated by Torres *et al.* [11], and might be as old as approximately 19,000 years ago (Table 1). Concerning tree topology (Figure 2), the first divergence within the clade containing genotypes F and H separates genotype H from its sister genotype F. On its turn, the first divergence within genotype F separates subgenotype F1 from a sister clade containing all other F subgenotypes. These early divergences within genotypes F and H seem to be younger than 10,000 years ago. However, as previously mentioned, these estimates are highly sensitive to the choice of the evolutionary rate.

Effect of purifying selection on the estimated rates using recent samples

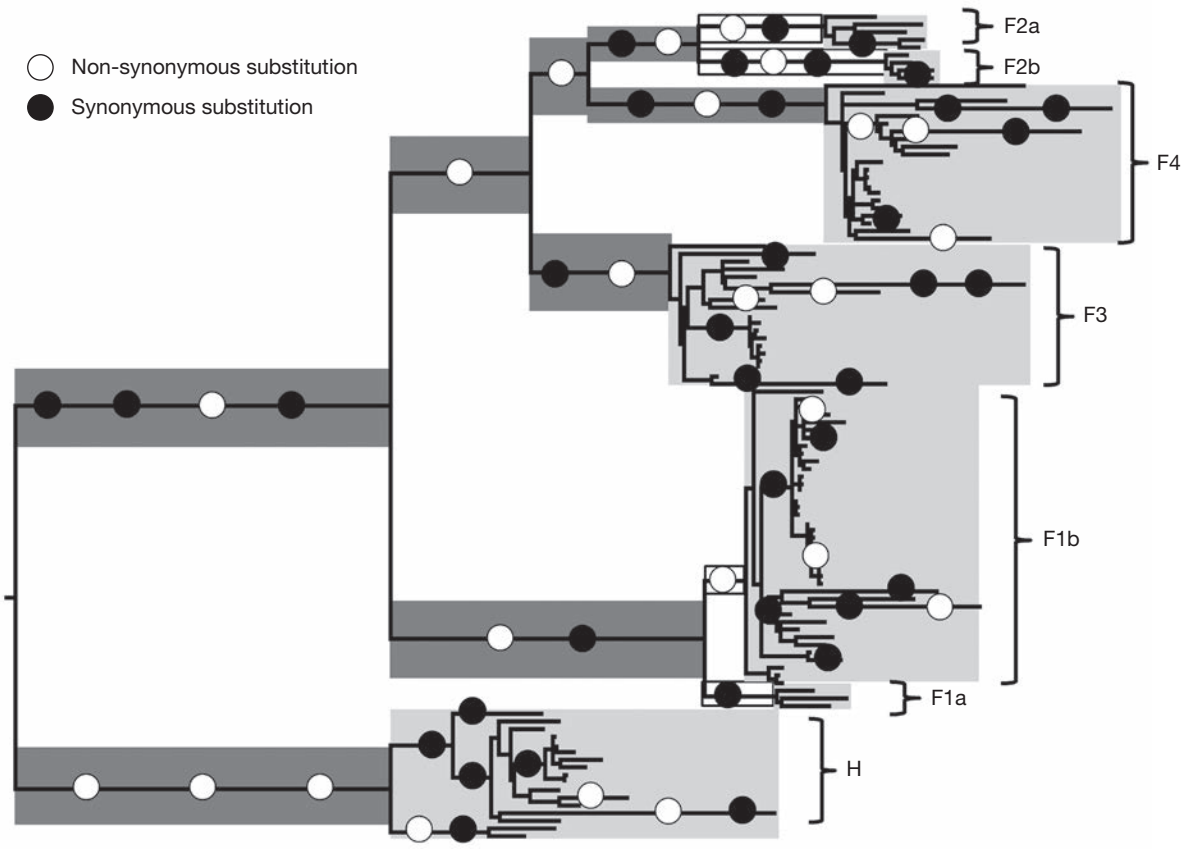
To test if the fast rates estimated from recent dated samples may be affected by the action of purifying selection on young branches, which would lead to an overestimation of the evolutionary rate (and a underestimation of divergence times), we compared the number of non-synonymous and synonymous substitutions in young versus old branches of HBV genotype F and H phylogeny (Figure 1). As shown in Table 2, there is a strong excess of non-synonymous substitutions in the young branches of HBV genotype F and H phylogeny ($P=4.87\times10^{-15}$ for the full genome comparison). This was also observed when the analysis considered only non-overlapping positions ($P=5.62\times10^{-15}$), or when individual genes were analysed separately, with the exception of gene X, which, despite an excess of non-synonymous to synonymous substitutions, did not show statistically significant values (Table 2). Importantly, considering 'intermediate' branches as either 'young' or 'old' did not change these results (Additional file 1).

Discussion

Estimates of HBV evolutionary rate: bias and solutions

The results presented here clearly indicate that there is an excess of non-synonymous substitutions in the young branches of HBV genotype F and H phylogeny. This provides a simple explanation for the discrepancy between 'historical' and 'molecular clock' estimates of HBV evolution, further suggesting that the latter is biased towards recent dates. These results contribute to the growing number of examples for which purifying selection acting on recent deleterious mutations will cause the evolutionary rate to be different

Figure 1. Schematic representation of the maximum likelihood HBV genotypes F and H tree based on complete genomes showing how mutations were counted along the phylogeny



Branches occurring within a given subgenotype were considered 'young' and are shown inside a light grey box. 'Intermediate' branches connect different lineages of the same subgenotype, as it is the case of F1a versus F1b, and F2a versus F2b. These are represented inside a white box. 'Old' branches are represented inside the dark grey box and connect different F subgenotypes and genotypes F and H. Non-synonymous and synonymous substitutions are represented scattered along the phylogeny as a white and black circles, respectively.

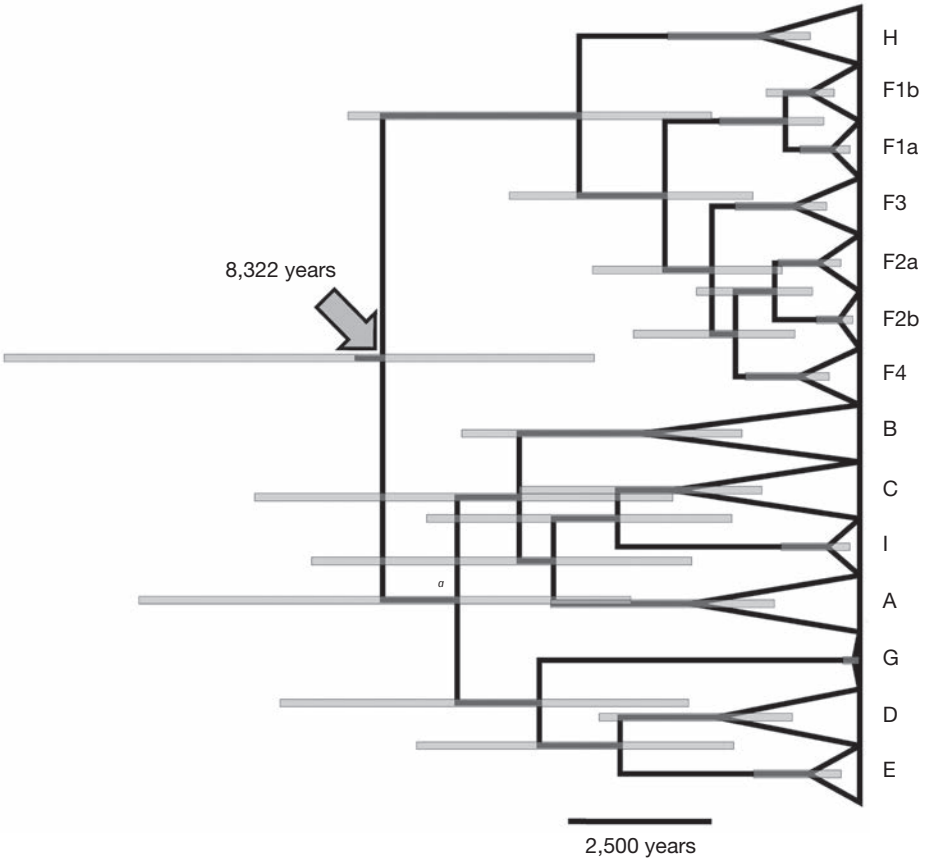
Table 1. Estimates of the time to the most recent common ancestor of specific nodes in HBV phylogeny according to the assumed evolutionary rate

	TMRCA in years (95% CI)				
Genotype	Strict clock, fixed fast rate of 1.67×10^{-4}	Strict clock, fixed slow rate of 1.0×10^{-5}	Strict clock, range of slow rates	Relaxed clock, fixed slow rate	Relaxed clock, range of slow rates
F1	85 (72, 100)	1,417 (1,185, 1,655)	1,140 (656, 2,053)	1,597 (1,107, 2,289)	1,300 (630, 2,451)
F2	113 (96, 132)	1,896 (1,602, 2,188)	1,522 (904, 2,770)	1,853 (1,255, 2,591)	1,488 (739, 2,765)
F3	116 (96, 137)	1,936 (1,610, 2,308)	1,568 (885, 2,816)	1,552 (999, 2,389)	1,164 (577, 2,173)
F4	86 (73, 100)	1,438 (1,226, 1,671)	1,158 (679, 2,095)	1,597 (952, 1,830)	1,072 (537, 1,994)
F	255 (229, 282)	4,254 (3,825, 4,712)	3,410 (2,033, 6,109)	4,134 (3,014, 5,457)	3,398 (630, 2,451)
H	103 (86, 120)	1,711 (1,443, 1,996)	1,374 (781, 2,474)	2,126 (1,418, 3,087)	1,773 (867, 3,351)
F and H	389 (350, 430)	6,498 (5,854, 7,170)	5,214 (3,093, 9,284)	5,769 (4,222, 7,917)	4,894 (2,587, 8,927)
Root	796 (732, 863)	13,295 (12,268, 14,413)	10,689 (6,438, 18,895)	8,657 (6,907, 10,510)	8,322 (4,631, 14,919)

Data are the time to the most recent common ancestor (TMRCA). 95% CI is the 95% credible interval for these data.

in young or old portions of the phylogeny [16,17]. Importantly, such excess would lead to an overestimation of the long-term evolutionary rate when calibrating HBV evolution with recent dated samples or data from intra-familiar transmissions. For example, Zhou and Holmes [12] estimate evolutionary rates in the order of 10^{-4}, with corresponding TMRCAs younger than 300 years ago. Our results suggest that this is an overestimation of the evolutionary rate, obtained using recent calibrations, with corresponding underestimated TMRCAs. However, even when only mutations occurring in 3rd codon positions and non-overlapping regions were considered, these authors found a global TMRCA for human HBV of approximately 477 years ago (with the 95% highest posterior density [HPD] estimate from 57 to 1,509 years ago) and an evolutionary rate of 9.63×10^{-4}, with a wide 95% HPD between 4.1×10^{-5} and 1.92×10^{-3}. If most mutations

Figure 2. Schematic representation of the Bayesian HBV phylogeny based on complete genomes



The scale bar represents a time span of 2,500 years. The arrow indicates the median estimate for the tree root assuming a relaxed clock with a range of 'slow rates', as explained in the text. The grey bars indicate the 95% CI for all time estimates. [a]All groups represented in this phylogeny had posterior probability values of at least 0.99, except the one indicated, which had a posterior probability of 0.74.

Table 2. Distribution of non-synonymous or synonymous nucleotide substitutions occurring in 'young' or 'old' branches according to the different partitions analysed

	Complete HBV genome		Non-overlapping positions		Gene Pol		Gene PreS/S		Gene PreC/C		Gene X	
Partition	NS	S	NS	S	NS	S	NS	S	NS	S	NS	S
Young branches	992	866	375	437	447	437	189	187	149	148	160	62
Old branches	170	334	44	198	104	167	26	69	2	60	24	15
P-value[a]	4.87×10^{-15}		5.62×10^{-15}		4.40×10^{-4}		6.25×10^{-5}		9.74×10^{-12}		0.183	

Data are number of non-synonymous (NS) or synonymous (S) substitutions in young and old branches. [a]*P*-value by χ^2 test.

in 3rd codon positions are synonymous, would this estimate be accurate? In highly linked systems, even synonymous (neutral) substitutions may be eliminated from the population if they are linked to deleterious non-synonymous mutations. Thus, if purifying natural selection is an important factor in HBV evolution, as our results suggest, in the long-term both deleterious non-synonymous and neutral synonymous substitutions linked to them will be eliminated from the population. The conclusion, therefore, is that using only synonymous substitutions is unlikely to alleviate the problems posed by purifying selection on evolutionary rate estimates. Moreover, a slower evolutionary rate in HBV is also consistent with the recent report of an HBV-C2 subgenotype found in a mummy from Korea dated to 330 years ago [26]. The excess of non-synonymous substitutions in young branches could also reflect positive selection in response to vaccine or immune escape. However, we think this is an unlikely explanation since implementation of vaccination programmes is very recent, and programmes still do not cover most parts of Latin American communities [27].

Another important implication of our findings is that this effect may be important for time estimates obtained from other viruses, as in these cases calibration points are usually given by intra-familiar transmissions or recent dated samples, which are recent events that may overestimate the long-term evolutionary rate. We suggest that better estimates of HBV evolutionary rate will be obtained by including known historical events, such as African slave trade to Latin America, as calibration points, or by obtaining HBV isolates from old samples, as the case of the Korean mummy [26]. However, it should be noted that the difference between the short- and long-term evolutionary rates due to purifying selection is not the only problem for accurate estimates of HBV evolutionary time scale. It has been shown that strains expressing hepatitis B e antigen (HBeAg) have evolutionary rate orders of magnitude faster than those estimated for HBeAg-negative strains [28,29]. In the future it will be important to understand the interplay between these two factors in determining the bias on HBV evolutionary rate estimates.

Implications for the origin of HBV in the Americas

Results obtained with the fast HBV evolutionary rate would be consistent with a post-Columbian origin of HBV in the Americas. However, we think 'fast rate' estimates are flawed, as they are incompatible with convincing historical evidence about the distribution of HBV lineages and also do not take into account the effects of purifying selection, as discussed thus far. Thus, if HBV in the Americas is pre-Columbian, could we know when it arrived there? For genotypes other than F and H the answer seems to be simple: they were carried here by African/European or Asian immigrants; but what about the 'indigenous' genotypes F and H?

Recent molecular studies about the peopling of the Americas suggest that the first New World settlers came from Asia through the Bering Strait around 18,000 years ago, following a fast expansion route through the Pacific [30,31]. This is corroborated by the fact that the oldest known archaeological site in the Americas (Monte Verde, dated at approximately 14,500 years ago) lies in Southern Chile [32]. After this initial entry, recurrent gene flow between Native Americans and Asians would have started around 9,750 years ago [33]. We can consider that these represent two alternative opportunities for HBV to spread from Asia to the Americas. The distribution of human HBV genotypes led Arauz-Ruiz *et al.* [34] to suggest that HBV New World genotypes F and H arrived from Asia with the initial settlers. As shown in Table 2, mutation rates in the order of 1.0×10^{-5} are easily compatible with divergence between genotypes F and H from the remaining genotypes older than 10,000 years ago, although they are hardly older than 18,000 years ago, for example. However, the uncertainty about the 'real' evolutionary rate of HBV makes it impossible to distinguish between different hypotheses or suggest that an alternative explanation should be invoked. Similarly, it is not possible to determine if these genotypes differentiated in America or if they arose elsewhere and were carried to the Americas.

The tree topology for HBV genotype F and H (Figure 2) is in agreement to those previously reported for HBV and HBV genotype F phylogeny [11,35]. Nonetheless, the fact that these two genotypes are sister clades in the phylogeny suggests that both differentiated in the Americas and that a 'pre-FH genotype' existed in the continent some time ago. Interestingly, the two most ancient divergences among the New World genotypes separate genotypes H and F1, respectively, from the remaining lineages. Genotype H is mostly distributed in Mexico and Central America [5,34], whereas subgenotype F1 is more common in Central America (F1a) [6,36] and in the Pacific coast of South America [26]. Therefore, it would be tempting to assume such divergences occurred as the first Native American human populations were moving southwards through the South American Pacific coast [30–32], but this would be highly speculative at this point, as these early divergences within genotypes F and H seem to be therefore more recent than the estimated times for the arrival of the first Native American populations in this region [32]. Better estimates of HBV evolutionary rate, taking into account purifying selection, will be crucial for an improved knowledge about the arrival of genotypes F and H in the Americas.

Acknowledgements

This study was funded by the National Council of Scientific and Technological Development (CNPq; grant number 573993/2008-4 to the National Institute of Medical Population Genetics), and supported by Fundação Faculdade de Medicina and Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo. We would like to thank two anonymous reviewers for their helpful comments on an earlier version of the manuscript.

Disclosure statement

The authors declare no competing interests.

Additional file

Additional file 1: Supplementary material including the GenBank accession numbers for the complete HBV genomes for genotypes A to I included in this study, and distributions of synonymous or non-synonymous nucleotide substitutions occurring in 'young' or 'old' branches according to the different partitions that were analysed, can be found at http://www.intmedpress.com/uploads/documents/AVT-12-SP-2655_Godoy_Add_file1.pdf

References

1. Lai CL, Ratziu V, Yuen MF, Poynard T. Viral hepatitis B. *Lancet* 2003; **362:**2089–2094.
2. Yu H, Yuan Q, Ge SX, *et al*. Molecular and phylogenetic analyses suggest an additional hepatitis B virus genotype 'I'. *PLoS ONE* 2010; **5:**e9297.
3. Schaefer S. Hepatitis B virus taxonomy and hepatitis B virus genotypes. *World J Gastroenterol* 2007; **13:**14–21.
4. Devesa M, Pujol FH. Hepatitis B virus genetic diversity in Latin America. *Virus Res* 2007; **127:**177–184.
5. Roman S, Tanaka Y, Khan A, *et al*. Occult hepatitis B in the genotype H-infected Nahuas and Huichol native Mexican population. *J Med Virol* 2010; **82:**1527–1536.
6. Alvarado-Mora MV, Romano CM, Gomes-Gouvea MS, *et al*. Molecular characterization of the Hepatitis B virus genotypes in Colombia: a Bayesian inference on the genotype F. *Infect Genet Evol* 2011; **11:**103–108.
7. Motta-Castro AR, Martins RM, Araujo NM, *et al*. Molecular epidemiology of hepatitis B virus in an isolated Afro-Brazilian community. *Arch Virol* 2008; **153:**2197–2205.
8. Santos AO, Alvarado-Mora MV, Botelho L, *et al*. Characterization of hepatitis B virus (HBV) genotypes in patients from Rondonia, Brazil. *Virol J* 2010; **7:**315.
9. Alvarado-Mora MV, Romano CM, Gomes-Gouvea MS, Gutierrez MF, Carrilho FJ, Pinho JRR. Molecular epidemiology and genetic diversity of hepatitis B virus genotype E in an isolated Afro-Colombian community. *J Gen Virol* 2010; **91:**501–508.
10. Piñeiro y Leone FG, Pezzano SC, Torres C, *et al*. Hepatitis B virus genetic diversity in Argentina: dissimilar genotype distribution in two different geographical regions; description of hepatitis B surface antigen variants. *J Clin Virol* 2008; **42:**381–388.
11. Torres C, Piñeiro y Leone FG, Pezzano SC, Mbayed VA, Campos RH. New perspectives on the evolutionary history of hepatitis B virus genotype F. *Mol Phylogenet Evol* 2011; **59:**114–122.
12. Zhou Y, Holmes EC. Bayesian estimates of the evolutionary rate and age of hepatitis B virus. *J Mol Evol* 2007; **65:**197–205.
13. Hannoun C, Horal P, Lindh M. Long-term mutation rates in the hepatitis B virus genome. *J Gen Virol* 2000; **81:**75–83.
14. Ho SYW, Larson G. Molecular clocks: when times are a-changin'. *Trends Genet* 2006; **22:**79–83.
15. Ho SYW, Phillips MJ, Cooper A, Drummond AJ. Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Mol Biol Evol* 2005; **22:**1561–1568.
16. Ho SYW, Saarma U, Barnett R, Haile J, Shapiro B. The effect of inappropriate calibration: three case studies in molecular ecology. *PLoS ONE* 2008; **3:**e1615.
17. Ho SYW, Lanfear R, Bromham L, *et al*. Time-dependent rates of molecular evolution. *Mol Ecol* 2011; **20:**3087–3101.
18. Crandall ED, Sbrocco EJ, DeBoer TS, *et al*. Expansion dating: calibrating molecular clocks in marine species from expansions onto the Sunda Shelf following the last glacial maximum. *Mol Biol Evol* 2012; **29:**707–719.
19. Kivisild T, Shen P, Wall D, *et al*. The role of selection in the evolution of human mitochondrial genomes. *Genetics* 2006; **172:**373–387.
20. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 2007; **7:**214.
21. Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence. *PLoS Biol* 2006; **4:**e88.
22. Swofford DL. *PAUP*: Phylogenetic Analysis using Parsimony (*and other methods)*. Version 4.0b10. Sunderland, MA: Sinauer Associates 2003.
23. Posada D, Crandall KA. Modeltest: testing the model of DNA substitution. *Bioinformatics* 1998; **14:**817–818.
24. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 2009; **25:**1451–1452.
25. Abramson JH, Gahlinger PM. *Computer programs for epidemiologists: PEPI v 4.0*. Salt Lake City, UT: Sagebrush Press 2001.
26. Bar-Gal GK, Kim MJ, Klein A, *et al*. Tracing hepatitis B virus to the 16th century in a Korean mummy. *Hepatology* 2012; **56:**1671–1680.
27. Ott JJ, Stevens GA, Groeger J, Wiersma ST. Global epidemiology of hepatitis B virus infection: new estimates of age-specific HBsAg seroprevalence and endemicity. *Vaccine* 2012; **30:**2212–2219.
28. Lim SG, Cheng Y, Guindon S, *et al*. Viral quasi-species during hepatitis B e antigen seroconversion. *Gastroenterology* 2007; **133:**951–958.
29. Harrison A, Lemey P, Hurles M, *et al*. Genomic analysis of hepatitis B virus reveals antigen state and genotype as sources of evolutionary rate variation. *Viruses* 2011; **3:**83–101.
30. Fagundes NJR, Kanitz R, Eckert R, *et al*. Mitochondrial population genomics supports a single pre-Clovis origin with a coastal route for the peopling of the Americas. *Am J Hum Genet* 2008; **82:**583–592.
31. Tamm E, Kivisild T, Reidla M, *et al*. Beringian standstill and spread of Native American founders. *PLoS ONE* 2007; **2:**e829.
32. Dillehay TD. *Monte Verde: a late Pleistocene settlement in Chile. The archaeological context and interpretation*. Vol. 2. Washington, DC: Smithsonian Institution Press 1997.
33. Ray N, Wegmann D, Fagundes NJR, Wang S, Ruiz-Linares A, Excoffier L. A statistical evaluation of models for the initial settlement of the American continent emphasizes the importance of gene flow with Asia. *Mol Biol Evol* 2010; **27:**337–345.
34. Arauz-Ruiz P, Norder H, Robertson BH, Magnius LO. Genotype H: a new Amerindian genotype of hepatitis B virus revealed in Central America. *J Gen Virol* 2002; **83:**2059–2073.

35. Venegas M, Alvarado-Mora MV, Villanueva RA, *et al*. Phylogenetic analysis of hepatitis B virus genotype F complete genome sequences from Chilean patients with chronic infection. *J Med Virol* 2011; **83:**1530–1536.

36. Arauz-Ruiz P, Norder H, Visona KA, Magnius LO. Genotype F prevails in HBV infected patients of hispanic origin in Central America and may carry the precore stop mutant. *J Med Virol* 1997; **51:**305–312.

Accepted 13 January 2013; published online 21 June 2013

Supplementary Information

Table S1 – Complete HBV genomes from Genotypes F and H included in this study

Table S2 – Complete HBV genomes from Genotypes A-E, G, and I includes in this study

Table S3. Distribution of non-synonymous (NS) or synonymous (S) nucleotide substitutions occurring in “Young” or “Old” branches according to the different partitions analyzed. “Intermediate” branches were considered as “Young”.

Table S4. Distribution of non-synonymous (NS) or synonymous (S) nucleotide substitutions occurring in “Young” or “Old” branches according to the different partitions analyzed. “Intermediate” branches were considered as “Old”.

Table S1 – Complete HBV genomes from Genotypes F and H included in this study

Subgenotype	**Country**	**GenBank**
F1a	Costa Rica	AY090458
F1a	Costa Rica	AY090459
F1a	El Salvador	AY090461
F1b	Argentina	AF223963
F1b	Argentina	AF223964
F1b	Argentina	AY179735
F1b	Argentina	DQ823091
F1b	Argentina	DQ823093
F1b	Argentina	DQ823094

F1b	Argentina	DQ823095
F1b	Argentina	EU366118
F1b	Argentina	EU366133
F1b	Argentina	FJ657525
F1b	Chile	HM585186
F1b	Chile	HM585187
F1b	Chile	HM585188
F1b	Chile	HM585189
F1b	Chile	HM585190
F1b	Chile	HM585191
F1b	Chile	HM585192
F1b	Chile	HM585193
F1b	Chile	HM585194
F1b	Chile	HM585195
F1b	Chile	HM585196
F1b	Chile	HM585197
F1b	Chile	HM585198
F1b	Chile	HM585199

F1b	Chile	HM585200
F1b	Chile	HM590474
F1b	Chile	HM627320
F1b	Chile	HM622135
F1b	Ireland	HQ378247
F1b	Japan	AB086397
F1b	Japan	AB116654
F1b	Peru	EU670262
F1b	USA	AB064316
F1b	USA	JN792917
F1b	USA	JN792918
F1b	USA	JN792919
F1b	USA	JN792920
F1b	USA	JN792922
F1b	Venezuela	AB116552
F2a	Brazil	X69798
F2a	Nicaragua	AY090455
F2a	Venezuela	AY311369

F2a	Venezuela	DQ899142
F2a	Venezuela	DQ899143
F2b	Venezuela	DQ899144
F2b	Venezuela	DQ899145
F2b	Venezuela	DQ899146
F2b	Venezuela	DQ899147
F3	Colombia	FJ589067
F3	Colombia	X75663
F3	Panama	AB116549
F3	Panama	AB116550
F3	Venezuela	AB036905
F3	Venezuela	AB036910
F3	Venezuela	AB036911
F3	Venezuela	AB036913
F3	Venezuela	AB036914
F3	Venezuela	AB036915
F3	Venezuela	AB036916
F3	Venezuela	AB036919

F3	Venezuela	AB036920
F3	Venezuela	AB116551
F3	Venezuela	AY311370
F3	Venezuela	DQ899148
F3	Venezuela	DQ899149
F3	Venezuela	DQ899150
F3	Venezuela	FJ589066
F4	Argentina	AF223962
F4	Argentina	AF223965
F4	Argentina	AY179734
F4	Argentina	DQ776247
F4	Argentina	DQ823086
F4	Argentina	DQ823087
F4	Argentina	DQ823088
F4	Argentina	DQ823089
F4	Argentina	DQ823090
F4	Argentina	EU366116
F4	Argentina	EU366132

F4	Argentina	FJ657519
F4	Argentina	FJ657522
F4	Argentina	FJ657528
F4	Bolivia	AB166850
F4	Bolivia	AB214516
F4	Bolivia	AB365446
F4	Bolivia	AB365447
F4	Bolivia	AB365449
F4	Bolivia	AB365450
F4	France	X75658
H	Argentina	FJ356715
H	Argentina	FJ356716
H	Indonesia	EU498228
H	Japan	AB179747
H	Japan	AB205010
H	Japan	AB266536
H	Japan	AB275308
H	Japan	AB298362

H	Japan	EF157291
H	Mexico	AB516393
H	Mexico	AB516395
H	Mexico	HM066946
H	Mexico	HM117850
H	Mexico	HM117851
H	Nicaragua	AY090454
H	Nicaragua	AY090457
H	USA	AY090460

Table S2 – Complete HBV genomes from other genotypes included in this study

Subgenotype	**Country**	**GenBank**
A1	Malawi	AB076679
A1	South Africa	AY233279
A2	Belgium	AF090841
A2	France	AJ309369
A3	-	AB19495
A3	Gabon	AM18412
A5	Haiti	FJ692595

A5	Haiti	FJ692598
B1	Japan	AB010292
B1	Japan	AB073838
B2	Japan	AB073837
B2	Taiwan	AB073841
B4	Vietnam	AB100695
B4	Vietnam	AY033072
B3	Indonesia	AB033554
B3	Indonesia	D00331
B5	Philippines	AB219426
B6	Greenland	AB287325
B6	Canada	DQ463787
C1	Thailand	AB074755
C1	India	DQ315781
C2	Japan	AB113875
C2	China	AB205123
C3	Polynesia	X75656
C3	New Caledonia	X75665

C5	Philippines	AB241111
C5	Vietnam	AF241410
C4	Australia	AB048704
D1	Russia	AB126581
D1	Egypt	AB104709
D2	Japan	AB109475
D2	India	AB090268
D3	France	AJ344117
D3	South Africa	AY233296
D4	Australia	AB048703
D4	Papua New Guinea	AB033559
D5	India	DQ315779
E	Ivory Coast	AB091255
E	Madagascar	DQ060830
E	Senegal	AY739675
G	Germany	AF405706
G	USA	AB056513
I	Vietnam	AF241407

I	Laos	FJ023661

Table S3. Distribution of non-synonymous (NS) or synonymous (S) nucleotide substitutions occurring in “Young” or “Old” branches according to the different partitions analyzed. “Intermediate” branches were considered as “Young”.

Partition	Complete HBV Genome		Non-Overlapping positions		Gene Pol		Gene PreS/S		Gene PreC/C		Gene X	
	NS	S	NS	S	NS	S	NS	S	NS	S	NS	S
Young branches	1024	963	384	501	464	493	193	206	150	165	164	65
Old branches	170	334	44	198	104	167	26	69	2	60	24	15
P-value*	8.93×10^{-13}		8.06×10^{-13}		0.003		2.14×10^{-4}		7.33×10^{-11}		0.204	

*Chi-square test

Table S4. Distribution of non-synonymous (NS) or synonymous (S) nucleotide substitutions occurring in “Young” or “Old” branches according to the different partitions analyzed. “Intermediate” branches were considered as “Old”.

Partition	Complete HBV Genome		Non-Overlapping positions		Gene Pol		Gene PreS/S		Gene PreC/C		Gene X	
	NS	S	NS	S	NS	S	NS	S	NS	S	NS	S
Young branches	992	866	375	437	447	437	189	187	149	148	160	62
Old branches	202	431	53	262	121	223	30	88	3	77	28	18
P-value*	9.42×10^{-21}		8.04×10^{-20}		1.19×10^{-6}		2.15×10^{-6}		5.81×10^{-14}		0.131	

*Chi-square test

CAPÍTULO IV – ARTIGO 2

HIGH PREVALENCE OF HBV/A1 SUBGENOTYPE REVEALS DIFFERENTIAL POPULATION DYNAMICS IN NATIVE SOUTH AMERICANS

Manuscrito em preparação a ser submetido à revista Plos One.

High prevalence of HBV/A1 subgenotype reveals differential population dynamics in Native South Americans

Bibiane A. Godoy[1], Michele S. Gomes-Gouvêa[2], Mónica V. Alvarado-Mora[2], João R. R. Pinho[2], Francisco M. Salzano[1], Nelson J. R. Fagundes[1*]

[1]Genetics Department, Federal University of Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brazil

[2]Laboratory of Tropical Gastroenterology and Hepatology "João Alves de Queiroz and Castorina Bittencourt Alves", Institute of Tropical Medicine, Department of Gastroenterology, University of São Paulo School of Medicine, São Paulo, SP, Brazil

*Corresponding author e-mail: nelson.fagundes@ufrgs.br

Abstract

Hepatitis B virus (HBV) is a hepatotropic DNA virus that presents a partially double-stranded circular genome. Based on sequence divergence of the complete genome, ten HBV genotypes (A-J) have been described, being F and H considered as indigenous from the Americas. HBV genotypes present a remarkable geographic structure which may reflect historic patterns of human migrations. In South America, the Amazon basin is considered an area of high HBV endemism, and high prevalence rates have been observed in Native American populations. The aim of this study was to characterize the HBV circulating in a historical sample of Native South Americans. The sample consisted of 1070 individuals belonging 38 populations, collected between the years 1965 and 1997. Presence of HBV-DNA was checked by quantitative real-time PCR, and determination of HBV genotypes was performed through amplification and sequencing followed by phylogenetic analysis of a fragment partially comprising HBsAg and Pol coding regions (S/Pol) of viral genome. Bayesian Skyline Plot was realized in order to compare the viral population dynamics of HBV/A1 of Native Americans and other Brazilian strains. A total of 109 individuals were positive for HBV-DNA (~10%), and 70 samples were successfully sequenced and genotyped. The most frequent subgenotype observed was A1 (66 - 94%), related with African populations and their descendants. The Skyline Plot analysis showed a marked population expansion in HBV/A1 from Native Americans occurring more recently (1945-1970) than in other Brazilian strains. Our results suggest that historic processes that contributed to formation of HBV/A1 circulating in Native American are related with more recent migratory waves towards the Amazon basin, which generated a different viral dynamics in this region.

Introduction

Hepatitis B Virus (HBV) belongs to the *Hepadnaviridae* family, which has marked tropism for hepatocytes [1]. The encapsulated viral genome presents a partially double stranded circular DNA with approximately 3,200 bp that displays four overlapping open reading frames [2]. The transmission can occur vertically – from mother to child - and horizontally – by parenteral or sexual way, due to the high concentrations that HBV reaches in body fluids [3]. The consequences of transmission range from asymptomatic infection to chronic hepatitis, which can progresses to complications such as liver cirrhosis and hepatocellular carcinoma [4].

The comparison between sequences and phylogenetic groups led to the classification of 10 different lineages of HBV (A to J), named "genotypes", which show more than 8% of nucleotide divergence at the whole genome level [5]. These genotypes are further divided into "subgenotypes", according to the inter-genotypic divergences between 4% and 8% [6]. HBV genotypes and subgenotypes present a strong geographic structure, where specific genotypes are associated with specific regions or ethnic profiles, reflecting historical aspects and migration patterns of human populations [7-9]. Some examples include the most divergent genotype F (HBV/F), considered as autochthonous to South America and related to Native American populations [10-12], and the A1 subgenotype (HBV/A1), which is related to African populations and common in many regions of Brazil due to the African slave trade during the colonial period (16^{th}-19^{th} centuries) [11, 13-19].

Although the availability of a vaccine has produced a significant reduction in the overall rates of infection [20], currently about 2 billion people worldwide are infected and 350 million people are chronic carriers, what maintains HBV as a serious problem of public health. Considering South America, the most endemic area comprises the Amazon basin, and the highest prevalence rates have been observed in Native American populations [10, 11, 21-23]. The isolated location, hygiene and sanitation conditions, and some cultural practices are part of the factors that can contribute to this figure. In addition, these populations have been associated with high rates of fulminant hepatitis in cases of superinfection with Hepatitis D Virus [24-26], suggesting a possible relationship between the genetic background present in these populations, the combination of viral genotypes and the severity of disease [10, 26, 27].

Despite HBV/F is considered the indigenous HBV genotype occurring in Native South American populations, the complex pattern of Latin America colonization has created a mosaic of HBV genotypes, with European, African, and Asian HBV genotypes occurring due to population migration [11, 28, 29]. In this context, the association of HBV genotypes with different disease outcomes and response to antiviral therapy [30, 31] makes important the characterization of the HBV lineages in a given region or population. The aim of this study is to performer a molecular characterization of HBV circulating in a large historical sample of Native South Americans in order to provide more information about the epidemiologic situation of HBV and better understand the historic processes that contributed to formation of these populations.

Materials and Methods

Sample and Ethics Statements

The sample is part of a historic bank collected in different periods between the years 1965 and 1997 with logistic support of Brazilian governmental agencies (National Indian Foundation - FUNAI). All participants collaborated voluntarily to the realization of the study. Ethical approval was provided by Brazilian National Ethics Commission (CONEP nº123/1998), according to all the ethic practices required at the time. This bank comprises genomic DNA previously extracted from total blood or plasma for use in earlier studies [review in 32]. All samples were assessed anonymously. The subjects consisted of 1070 Native Americans individuals from different 38 populations located in Amazonia (30 populations; 78.9% of the samples populations), Chaco (3; 7.9%), Brazilian Central Plateau (2; 5.3%), Southern South America (1; 2.6%), and Southern Brazil (2; 5.3%). These populations belong to 10 linguistic groups, and sample size ranged from 4 to 55 individuals (Parintintin and Mekranoti, respectively). Table 1 describes all the information collected about each studied population, whose location is shown in Figure 1.

HBV detection and amplification

We considered detection of HBV DNA as an indication of individuals carrying HBV. We verified the presence of HBV DNA by quantitative real-time PCR with primers previously described that allows the amplification of any HBV genotype [33]. To characterize the HBV genotypes of positive samples, we realized a nested PCR with

primers previously described that result in the amplification of a fragment of 1306 bp partially comprising the HBsAg and Polymerase ORFs (S/Pol) [34]. For positive samples that failed to amplify in this first stage, we realized a second nested PCR to amplify of a fragment of 734 bp that overlaps the same genomic region [26]. After purification of the PCR products through ExoSAP kit (GE Healthcare), Sanger sequencing was performed in Macrogen (Macrogen Inc., Seoul, South Korea). We checked the quality of the sequences and assembled a consensus sequence for each individual using the software Geneious [35]. Sequences used in this work were deposited in GenBank with accession numbers: xxx-xxx.

HBV genotyping analyses

We based the assignment of the HBV sequences obtained in this study on phylogenetic analyses using reference sequences from each HBV genotype obtained in GenBank. In addition, several sequences isolated in Brazil were included in the analysis for comparison (n=352; Table S1). The alignment was done MEGA 5 software [38] and edited manually to correct small insertion-deletion misalignments. All the sequences included in the dataset covered totally or partially a 1306pb of HBsAg and DNA polymerase coding regions (S/Pol) of viral genome. Bayesian phylogenetic analysis was performed in BEAST v.1.7.5 [36], with 20,000,000 steps of the Markov Chain Monte Carlo (MCMC) sampling every 1,000 steps, and discarding the first 2,000,000 steps as burn-in. A strict clock was performed with a rate of $1.0x10^{-5}$, which was shown to be roughly consistent with HBV/F history and the Peopling of the Americas [37,38]. We choose a Yule model as the tree prior, and the nucleotide substitution model (GTR+G+I) was determined according with MEGA 5 software [39]. The maximum clade credibility (MCC) tree was obtained using TreeAnnotator v1.7.5 by summarizing the 18,000 trees sampled after burn-in.

Population dynamics of HBV/A1

Population size dynamics for HBV/A1 was estimated through the Bayesian Skyline Plot generated in Tracer 1.5 [40]. We used the HKY+G substitution model as suggested by MEGA 5, and the same clock model used in the phylogenetic analysis. We ran the MCMC analysis in BEAST 1.7.5 using 50,000,000 steps, discarding the first 5,000,000 as burn-in. Initially, we used all HBV/A1 sequences identified in the sample with sampled dates as an

additional prior for time estimates. Next, to evaluate if the population dynamics in the Native American sample was different from that found for HBV/A1 in Brazil as a whole, we repeated the analysis using HBV/A1 sequences from Brazil retrieved from GenBank (Table S3). In this case, sampled data was not included in the analysis. To guarantee that the pattern of the Bayesian Skyline Plot was indifferent to the sampled date information we reanalyzed the sequences found in our sample without the sampled date information. As expected, the general pattern of population size chance was the same either with or without sample date information, which some differences affecting only the time scale of the analysis (Data not shown).

Spatial analysis

In order to evaluate the spatial distribution of HBV prevalence we used a Kriging technique [41] in the software ArcGIS 10® to compare the geographic location of the samples, HBV prevalence, and reported time of contact with non-Native populations.

Results

Of the 1070 individuals included in this work, 109 were positive for HBV (10.2%). The prevalence of HBV varied between the populations (Table 1), reaching the highest values in Txucahamãe (76.7%) and Parakanã (66.7%). Seventeen populations did not show any individual positive for HBV. Among the HBV positive samples, 70 were successfully sequenced and genotyped. We obtained the largest (1306 bp) fragment of the S/Pol region for 53 samples, while for the remaining 17 samples we obtained the 734 bp fragment. There is no obvious relationship between prevalence and time of the first contact with non-Native populations (Figure 2). Indeed, non-parametric Spearman correlation between prevalence and year of first contact with non-Natives was not statistically significant (rho=-0.217, P=0.191, for the whole data; rho=-0.167, P=0.470 excluding populations for which prevalence was zero). On the other hand, prevalence have a clear geographic component, which higher values in Eastern Amazon, and lower values in Western Amazon (Figure 2).

The Bayesian phylogenetic tree (Figure 3) show good support values for the keys nodes defining each genotype or subgenotype. Surprisingly, the most frequent subgenotype detected was A1, present in 94.3% of the positive individuals. Native Americans strains of

HBV/A1 formed a closely related cluster in the tree, but due to low support values for several internal nodes, it was not possible to evaluate the evolutionary relationship of these lineages in further detail. The remaining four individuals who carried other genotypes had linages belonging to genotype F. Two individuals (2.8%) from the Amazonian population Apalaí had subgenotype F2a, while one individual from each of the populations Aché and Ayoreo, both in the Chaco region, had subgenotype F4 (two individuals in total – 2.8%). All populations showed only one circulating subgenotype. We were unable to amplify 39 samples that were positives in the first screening, and in consequence of this, five populations having individuals positive for HBV DNA could not be characterized.

The Bayesian Skyline Plot analysis shows an evident viral population size expansion in the sample of Native American A1 strains that seems to have started around 1945 up to 1970 (Figure 4a). The population dynamics observed in Native American A1 strains differs markedly of pattern presented in other A1 strains from Brazil, where there is evidence of a much older expansion ~1,100 years ago (Figure 4b). Importantly, while the history of the A1 subgenotype in Native Americans is clearly more recent than the African Slave trade to South America, the dynamics of the Brazilian population predates this event, corroborating the idea that A1 was transmitted to Native American populations brought to South America relatively recently, and suggesting that Brazilian A1 lineages are paraphyletic. In other words, Brazil actually contains a diverse phylogenetic set of A1 lineages whose most recent common ancestor is in Africa.

Discussion

The overall HBV prevalence in our sample, of 10.2% is in agreement to the general figure of high prevalence of HBV (>8%) reported in Native American populations [23, 42] and in the Amazon region [22, 43], which is where most of our sample came from. Indeed, considering only Amazonian populations (see Table 1) the overall prevalence in our sample is of 10.9%. We found three different HBV subgenotypes circulating in the studied Native American populations: A1 (94.3%), F2a (2.85%) and F4 (2.85%). HBV/F is thought to be, together with HBV/H, one of the indigenous genotypes from the Americas. Because of this, it is generally assumed that it should be the most prevalent in isolated indigenous communities [43], but our results do not support this intuition. Concerning the two HBV/F subgenotypes found in our sample, the F2a subgenotype is more common in

Venezuela and in some parts of Brazil, mainly in the North region [6, 44], while the F4 subgenotype is more frequent in the Chaco region, mainly in Bolivia and Argentina [46]. The locations of the tribes Apalaí - that presented 2 individuals HBV/F2a – as well as Aché and Ayoreo – each one presented an individual HBV/F4 – corroborate this previous description.

HBV/A has many subgenotypes and is found mainly in Africa, Europe, and North America. Subgenotype A1 has been associated with African populations and their descendants [18, 46, 47]. Although the F genotype had been described as "Amerindian", HBV/A1 was, by far, the most prevalent in our sample, occurring in 13 populations. Importantly, this expansion does not merely reflects a general "Brazilian" population dynamic, as it is clear from the different patterns exhibited by the Bayesian Skyline Plot estimated for other samples from Brazil. Currently, HBV/A1 is the most frequent subgenotype in Brazil [11] and its specific increase in frequency in Native populations have been attributed to migratory movements toward western Amazonia during the late of 19th and early 20th centuries associated with the Amazon "rubber boom" which impacted regional economy [44, 48]. However, the Bayesian Skyline Plot analysis of the HBV/A1 lineages found in our sample revealed a more recent population expansion, around 1945-1970. One possibility is that this expansion is associated with the "second rubber boom" that occurred during the second world war between 1942 and 1945 in a Brazil-USA agreement to supply allies armies with rubber due to the Japanese control of Southeast Asia plantations [49].

Nonetheless, the geographical location of the tribes with highest prevalence of HBV in Eastern Amazon partially overlaps to the referred "arc of deforestation" that comprises the southeast area of the Amazon basin where most of the original forest have been altered as peopled migrated from other Brazilian regions towards the Amazon [50]. The "arc of deforestation" region attracted people of many regions of Brazil since 1958 when the construction of the Belém-Brasília Highway began. This was followed by the construction of other highways that continued to attract people living in eastern Brazil as work force, culminating in the construction of Transamazon Highway in the early 1970s [51].

The migratory movement initiated by the construction of such roads or caused by migration into the Amazon during the second "rubber boom" could have served as vehicle

for the spreading of HBV/A1 from other regions of Brazil toward the Native American populations. In both cases, it would generate a viral epidemic dynamics different from that observed for Brazilian HBV/A1 lineages found in non-Native American populations. However, the migration into the Amazon during the second "rubber boom" directed many workers into Western Amazon [49], and these regions do not show a high HBV prevalence. On the other hand, the effects of road construction and the "arc of deforestation" system are geographically more consistent with our results. Of course, our time estimates are partially dependent on the mutation rate assumed for HBV. A mutation rate of $1.0x10^{-5}$ is consistent with the divergence of genotypes F and H associated with the Peopling of the Americas [37, 38]. However, this could be considered as a conservative value for more recent events due to the effect of purifying selection over HBV [38]. A faster mutation rate would make our estimates more recent, and therefore, even the second "rubber boom" would be too old to be compatible with our estimates, lending further support for an effect of the "arc of deforestation" system.

In this study we found a single subgenotype circulating in each Native American population and the frequent prevalence of HBV/A1. Interestingly, large urban populations have many different lineages of HBV circulating (e.g. 11, 28, 29), the fact that all these small populations have a single lineage may suggest that stronger genetic drift at the pathogen level or that inter-genotype competition may accelerate the extinction of other pathogenic lineages. Other studies on small rural communities in Brazil have also shown a single HBV genotype [16, 18]. The occurrence of HBV/A1 in tribes of different regions and linguistic classification seems indicate that specific population habits do not seem to have a strong influence in genotype distribution. For these populations, horizontal transmission, mainly the intra-familiar way, is considered as the most frequent mode of contagion [52, 53]. Furthermore, cultural practices present in some native populations such as scarifications, bloodletting, piercing/tattooing or pre-mastication of food can increase the risk of infection [23]. Our study reports the mostly presence of HBV/A1 in a larger sample of Native Americans and provide additional insights about the epidemiologic situation of these populations. Recent studies have shown an interaction between pathogen and host, where the genetic profiles in addition with the ancestries of both display an important role under the outcome of disease [54]. Therefore, the genetic profile present in the Native American populations could produce different clinical responses towards

“African” viral genotypes. Further studies with Native American populations need be carried considering information about genetic of pathogen and individual and progression of disease in order to provide better guidelines of politics of public healthy for these populations.

Acknowledgements

This study was funded by National Council of Scientific and Technological Development (CNPq), and supported by Fundação Faculdade de Medicina da Universidade de São Paulo. We also thank Rafael Bisso Machado for help with sample information and Marcelo Zagonel Oliveira for help with ArcGIS and the Kriging approach.

References

1. Ganem D, Prince AM. (2004) Hepatitis B virus infection natural history and clinical consequences. N Engl J Med 350: 1118–1129.

2. Wei Y, Neuveut C, Tiollais P, et al. (2010) Molecular biology of the hepatitis B virus and role of the X gene. Pathol Biol 58(4): 267-72.

3. Alvarado-Mora MV, Pinho JRR. (2013) Epidemiological update of hepatites B, C and delta in Latin America. Antiviral Therapy 18: 429-433.

4. Dandri M, Locarnini S. (2012) New insight in the pathobiology of hepatitis B virus infection. Gut 61 Suppl 1: i6–17.

5. Yu H, Yuan Q, Ge SX, et al. (2010) Molecular and phylogenetic analyses suggest an additional hepatitis B virus genotype "I". PLoS One 5: e9297.

6. Alvarado-Mora MV, Pinho JRR (2013) Distribution of HBV genotypes in Latin America. Antiviral Therapy 2013 18: 459-465.

7. Campos RH, Mbayed VA, Pineiro Y, et al. (2005) Molecular epidemiology of hepatitis B virus in Latin America. J ClinVirol 34 Suppl 2: S8-S13.

8. Andernach IE, Nolte C, Pape JW, et al. (2009) Slave trade and hepatitis B virus genotypes and subgenotypes in Haiti and Africa. Emerg Infect Dis 15(8): 1222–8.

9. Paraskevis D, Magiorkinis G, Magiorkinis E, et al. (2013) Dating the origin and dispersal of hepatitis B virus infection in humans and primates. Hepatology 57(3): 908-16.

10. Devesa M, Pujol FH. (2007) Hepatitis B virus genetic diversity in Latin America. Virus Res 127(2): 177-84.

11. Mello FCA, Souto FJD, Nabuco LC, et al. (2007) Hepatitis B virus genotypes circulating in Brazil: molecular characterization of genotype F isolates. BMC Microbiol 7: 103.

12. Dias AL, Oliveira CM, Castilho MC, et al. (2012) Molecular characterization of the hepatitis B virus in autochthonous and endogenous populations in the Western Brazilian Amazon. Rev Soc Bras Med Trop. 45(1): 9-12.

13. Araujo NM, Mello FCA, Yoshida CFT, et al. (2004) High proportion of subgroup A' (genotype A) among Brazilian isolates of Hepatitis B virus. Arch Virol 149: 1383–1395.

14. Motta-Castro AR, Martins RM, Araujo NM, et al. (2008) Molecular epidemiology of hepatites B vírus in na isolated Afro-Brazilian community. Arch Virol 153: 2197-2205.

15. Oliveira CM, Farias IP, Fonseca JCF et al. (2008) Phylogeny and molecular genetic parameters of different stages of hepatitis B virus infection in patients from the Brazilian Amazon. Arch Virol 153: 823–830.

16. Matos MAD, Reis NRS, Kozlowski AG et al. (2009) Epidemiological study of hepatitis A, B and C in the largest Afro-Brazilian isolated community. Trans R Soc Trop Med Hyg. 103(9): 899-905.

17. Santos AO, Alvarado-Mora MV, Botelho L, et al. (2010) Characterization of hepatitis B vírus (HBV) genotypes in patients from Rondônia, Brazil. Virol J 7: 315.

18. Alvarado-Mora MV, Botelho L, Gomes-Gouvêa MS, et al. (2011) Detection of Hepatitis B virus subgenotype A1 in a Quilombo community from Maranhão, Brazil. Virol J 8: 415.

19. Moura IF, Lopes EP, Alvarado-Mora MV, et al. (2013) Phylogenetic analysis and subgenotypic distribution of the hepatitis B virus in Recife, Brazil. Infect Genet Evol 14: 195-9.

20. Zanetti AR, Damme PV, Shouval D. (2008) The global impact of vaccination against hepatitis B: A historical overview. Vaccine 26(49): 6266-73.

21. Tanaka, J. (2000) Hepatitis B epidemiology in Latin America. Vaccine 18 Suppl 1: S17-9.

22. Parana, R., Almeida, D. (2005) HBV epidemiology in Latin America. J. Clin. Virol. 34, S130–S133.

23. Duarte MC, Cardona N, Poblete F, et al. (2010) A comparative epidemiological study of hepatitis B and hepatitis D virus infections in Yanomami and Piaroa Amerindians of Amazonas State, Venezuela. Trop Med Int Health 15(8): 924-33.

24. Casey JL, Niro GA, Engle RE, 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.

25. Quintero A, Uzcategui N, Loureiro CL, et al. (2001) Hepatitis delta virus genotypes I and III circulate associated with hepatitis B virus genotype F in Venezuela. J Med Virol 64: 356–359.

26. Gomes-Gouvêa MS, Soares MC, Bensabath G, et al. (2009) Hepatitis B virus and hepatitis delta virus genotypes in outbreaks of fulminant hepatitis (Labrea black fever) in the western Brazilian Amazon region. J Gen Virol 90(Pt 11): 2638-43.

27. Nakano T, Shapiro CN, Hadler SC, et al. (2001).Characterization of hepatitis D virus genotype III among Yucpa Indians in Venezuela. J Gen Virol 82: 2183–2189.

28. Pereira LMMB, Martelli CMT, Merchán-Hamann E, et al. (2009) Population-Based Multicentric Survey of Hepatitis B Infection and Risk Factor Differences among Three Regions in Brazil. Am J Trop Med Hyg 81(2): pp. 240–247.

29. Bertolini DA, Gomes-Gouvea MS, Carvalho-Mello IM, et al. (2012) Hepatitis B virus genotypes from European origin explains the high endemicity found in some areas from southern Brazil. Infect Genet Evol 12: 1295–1304.

30. Kim BK, Revill PA, Ahn SH. (2011) HBV genotypes: relevance to natural history, pathogenesis and treatment of chronic hepatitis B. Antivir Ther 16(8): 1169-86.

31. Tanwar S, Dusheiko G. (2012) Is there any value to hepatitis B virus genotype analysis? Curr Gastroenterol Rep 14(1): 37-46.

32. Salzano FM (2002). Molecular variability in Amerindians: widespread but uneven information. An Acad Bras Cienc 74(2): 223-63.

33. Sitnik R, Paes A, Mangueira CP, et al. (2010) A real-time quantitative assay for hepatitis B DNA virus (HBV) developed to detect all HBV genotypes. Rev Inst Med Trop Sao Paulo 52(3): 119-24.

34. Alvarado-Mora MV, Romano CM, Gomes-Gouvea MS, et al. (2011) Molecular characterization of the Hepatitis B virus genotypes in Colombia: a Bayesian inference on the genotype F. Infect Genet Evol 11(1): 103–108.

35. Drummond AJ, Ashton B, Buxton S, et al. (2010) Geneious v5.3. Available at: http://www.geneious.com

36. Drummond AJ, Rambaut A. (2007) BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol 7: 214.

37. Torres C, Piñeiro y Leone FG, Pezzano SC, et al. (2011) New perspectives on the evolutionary history of hepatitis B virus genotype F. Mol Phylogenet Evol 59: 114-122.

38. Godoy BA, Alvarado-Mora MV, Gomes-Gouvêa MS, et al. (2013) Origin of HBV and its arrival in the Americas--the importance of natural selection on time estimates. Antivir Ther 18(3 Pt B): 505-12.

39. Tamura K, Peterson D, Peterson N, et al. (2011) MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Molecular Biology and Evolution 28: 2731-2739.

40. Drummond AJ, Rambaut A, Shapiro B & Pybus OG (2005) Mol Biol Evol 22, 1185-1192.

41. Camargo, E. C. G. (1997) Desenvolvimento, implementação e teste de procedimentos geoestatísticos (krigeagem) no sistema de processamento de informações georreferenciadas (SPRING). São José dos Campos: INPE, 1997. 124p.

42. Cardona NE, Loureiro CL, Garzaro DJ, et al. (2011) Unusual presentation of hepatitis B serological markers in an Amerindian community of Venezuela with a majority of occult cases. Virol J 8: 527.

43. Castilho MC, Oliveira CMC, Gimaque JBL, et al. (2012) Epidemiology and Molecular Characterization of Hepatitis B Virus Infection in Isolated Villages in the Western Brazilian Amazon. Am J Trop Med Hyg 87(4) pp. 768–774.

44. Mello FC, Araujo OC, Lago BV, et al. (2013) Phylogeography and evolutionary history of hepatitis B virus genotype F in Brazil. Virol J 10: 236.

45. Venegas M, Alvarado-Mora MV, Villanueva RA, et al. (2011) Phylogenetic analysis of hepatitis B virus genotype F complete genome sequences from Chilean patients with chronic infection. J Med Virol 83(9): 1530-6.

46. Hannoun C, Söderström A, Norkrans G, et al. (2005) Phylogeny of African complete genomes reveals a West African genotype A subtype of hepatitis B virus and relatedness between Somali and Asian A1 sequences. J Gen Virol 86(Pt 8): 2163-7.

47. Kramvis A, Kew MC. (2007) Molecular characterization of subgenotype A1 (subgroup Aa) of hepatitis B virus. Hepatol Res 37(s1): S27-32.

48. Ribeiro D (1995) O Povo Brasileiro: a formação e o sentido do Brasil. São Paulo, Brazil: Companhia das Letras.

49. Martinello P (2004) A Batalha da Borracha na Segunda Guerra Mundial. Rio Branco, Brazil: EDUFAC.

50. Durieux L, Machado LAT, Laurent H. (2003) The impact of deforestation on cloud cover over the Amazon arc of deforestation. Remote Sensing of Environment 86: 132–140.

51. Fearnside PM. (2005) Deforestation in Brazilian Amazonia: History, Rates, and Consequences. Conservation Biology 19(3): 680-688.

52. Brasil LM, da Fonseca JC, de Souza RB, et al. (2003) Prevalence of hepatitis B virus markers within household contacts in the state of Amazonas. Rev Soc Bras Med Trop 36: 565–570.

53. Lobato C, Tavares-Neto J, Rios-Leite M, t al. (2006) Intrafamilial prevalence of hepatitis B virus in Western Brazilian Amazon region: epidemiologic and biomolecular study. J Gastroenterol Hepatol 21: 863–868.

54. Kodaman N, Pazos A, Schneider BG, et al. Human and Helicobacter pylori coevolution shapes the risk of gastric disease. PNAS 2014; 111(4):1455-1460.

55. Lewis MP. (2009) Ethnologue: languages of the World,16th ed. Dallas: SIL International. Online version:http://www.ethnologue.com/.

56. Callegari-Jacques SM, Salzano FM. (1999) Brazilian Indian/non-Indian interactions and their effects. Ciência e Cultura, São Paulo, v. 51, n. 3-4, p. 166-174. 1999.

Figure Legends:

Figure 1: Geographical location of 38 Native American populations included in this study.

Figure 2: Prevalence and HBV genotypes observed in a sample of Native Americans according to the year of first contact with non-Native populations.
Circle sizes are proportional to the prevalence of HBV in each population. Different symbols indicate different genotypes.

Figura 3: The maximum clade credibility (MCC) tree estimated by Bayesian analysis of 352 S/POL sequences with 1306 bp of HBV strains. The posterior probabilities (PP) of key nodes (internal nodes) are depicted above the respective nodes. PP<0.90 are not shown. Samples obtained from Native American individuals are shown in red. Other Brazilian strains are shown in blue. Some clades were collapsed for better visualization and are detailed in Table S2.

Figure 4: Bayesian Skyline Plot analyzes showing the pattern of population dynamics to
a) HBV/A1 strains isolated from Native American individuals. b) HBV/A1 strains isolated from urban Brazilian individuals. The thick red solid line is the median estimate, and the thick grey lines show the 95% highest posterior density limits.

Table 1: Characterization of a sample of 38 Native South American populations included in this study.

Population	Linguistic classification[1]	Region	Geographical coordinates	Sampling	First contact with non-Native populations[2]	n	+HBV/DNA	Prevalence (%)
Jamamadi	Arauan	Amazon	7°15’S; 66°41’W	1986	1963	23	00	0
Mapuche	Araucanian	Southern South America	39°10’ -41°20’S; 68°37’ -70°22’W	1990	1962	28	00	0
Baniwa	Arawakan	Amazon	1°N, 67°50'W	1976	1933	43	01	2.3
Tiryió	Carib	Amazon	2°N; 56°W	1970	1940	30	00	0
Wai-Wai	Carib	Amazon	0°40’S; 58°W	1988	1936	52	00	0
Arara	Carib	Amazon	3°30’ -4°20’S; 53°0’ -54°10’W	1985	1962	23	01	4.3
Apalaí	Carib	Amazon	1°20’N; 54°40’W	1983	1967	16	02	12.5
Kali'na	Carib	Amazon	5°31’N; 53°47’W	1991	1971	20	01	5
Pacaás Novos	Chapacura-Wanham	Amazon	11°N; 63°W	1981	1951	30	04	13.3
Kaingang	Macro-Ge	Southern Brazil	28°S; 51°20’W	1988	1963	25	00	0
Mekranoti	Macro-Ge	Amazon	8°40’S; 54°W	1969	1914	55	05	9.1
Xavánte	Macro-Ge	Brazilian Central Plateau	14°S; 52°30’W	1990	1962	28	00	0
Xikrin	Macro-Ge	Amazon	5°55’S; 51°W	1970	1942	28	00	0
Gorotire	Macro-Ge	Amazon	7°44’S; 51°10’W	1974	1940	34	15	44.1

Krahó	Macro-Ge	Brazilian Central Plateau	8°S; 47°15’W	1974	1952	22	12	54.5
Kuben-Kran-Kegn	Macro-Ge	Amazon	8°10’S; 52°8’W	1965	1935	30	05	16.7
Txukahamãe	Macro-Ge	Amazon	10°20’S; 53°5’W	1970	1940	30	23	76.7
Lengua	Mascoian	Chaco	22°45’S; 58°5’W	1994	1965	29	01	3.4
Mura	Mura	Amazon	3°34'S, 59°12'W	1985	1956	29	00	0
Cinta-Larga	Tupi	Amazon	9°50’ –12°30’S; 59°10’ –60°50’W	1987	1957	30	02	6.7
Wayampi	Tupi	Amazon	1°N; 53°W	1980	1932	48	02	4.2
Sateré Mawé	Tupi	Amazon	3°S; 57°W	1982	1952	30	00	0
Tenharim	Tupi	Amazon	8°20’S; 62°W	1986	1962	24	00	0
Asurini	Tupi	Amazon	3°35’ -4°12’S; 49°40’ -52°26’W	1984	1954	30	03	10.0
Karitiana	Tupi	Amazon	9°30’S; 64°15’W	1986	1956	30	03	10.0
Parakanã	Tupi	Amazon	5°22’S; 51°17’W	1984	1954	30	20	66.7
Suruí	Tupi	Amazon	5°58’ -10°50’S; 48°39’ -61°10’W	1990	1962	28	00	0
Urubu-Kaapor	Tupi	Amazon	2° -3°S; 46° -47°W	1983	1954	29	00	0
Uru-Eu-Wau-Wau	Tupi	Amazon	62ºW - 12ºS	1986	1980	06	00	0
Parintintin	Tupi	Amazon	64ºW - 7ºS	1986	1982	04	00	0
Munduruku	Tupi	Amazon	6°23’S; 59°9’W	1985	1961	24	00	0
Guarani	Tupi	Southern Brazil	23°6’S; 55°12’W	1988	1961	27	01	3.7
Gavião	Tupi	Amazon	10°10’S; 61°8’W	1990	1961	29	00	0

Aché	Tupi	Chaco	23°30’ –24°10’S; 55°50’ – 56°30’W	1997	1967	30	02	6.7
Araweté	Tupi	Amazon	5°9’S; 52°22’W	1986	1956	30	01	3.3
Zoró	Tupi	Amazon	10°20’S; 60°20’W	1990	1959	31	00	0
Karipuna	Tupi	Amazon	2°50’N; 54°W	1983	1975	08	04	50.0
Ayoreo	Zamucoan	Chaco	19°S; 60°30’W	1994	1967	27	01	3.7
						1070	**109**	**10,2**

1 – According with Lewis [54]

2 – According with Callegari-Jacques and Salzano [55]

Figure 1:

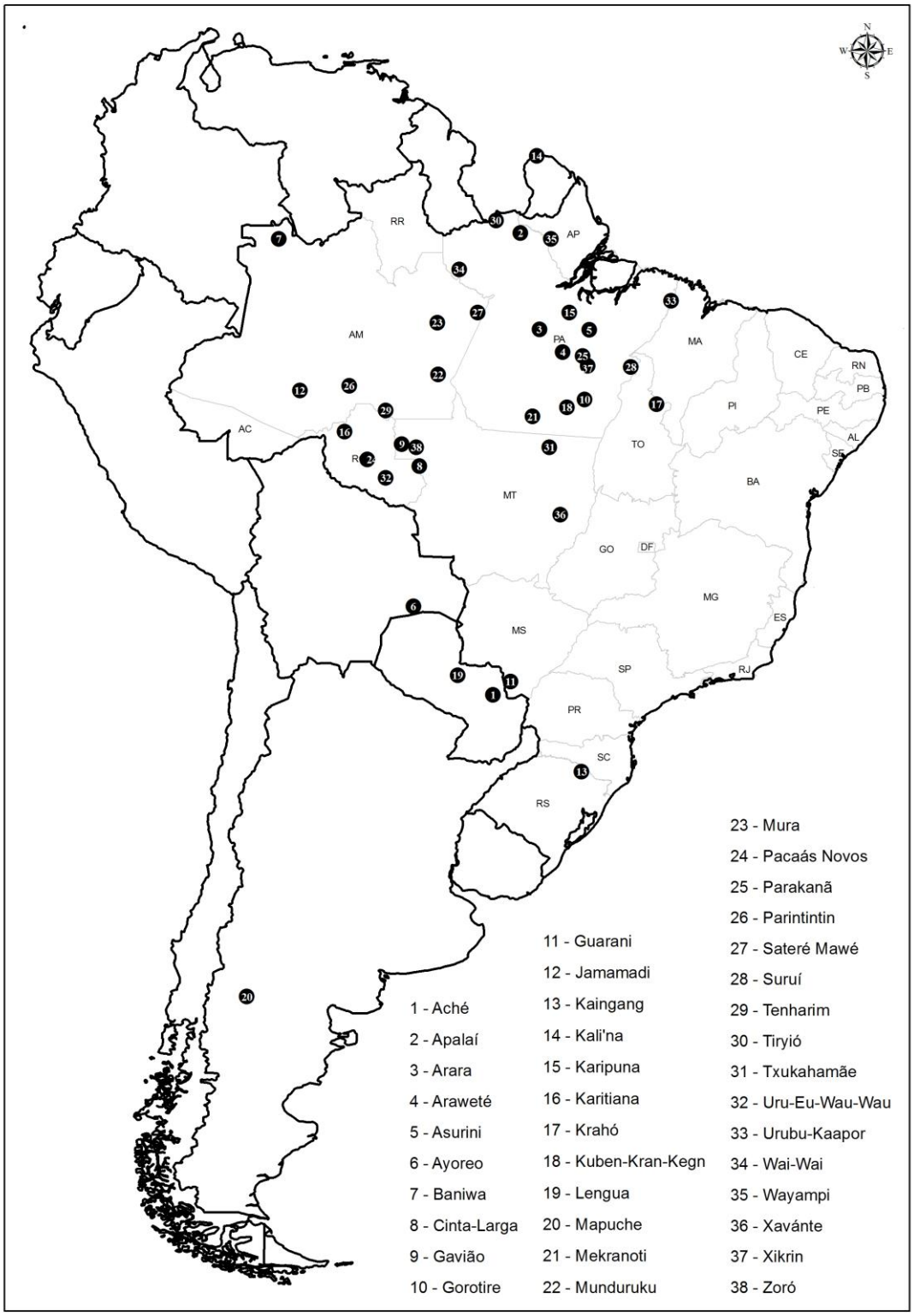

N
W
E
S
RR
AP
AM
PA
MA
CE
RN
PB
PI
PE
AL
SE
AC
TO
BA
MT
GO
DF
MG
ES
MS
SP
RJ
PR
SC
RS
1 - Aché
2 - Apalaí
3 - Arara
4 - Araweté
5 - Asurini
6 - Ayoreo
7 - Baniwa
8 - Cinta-Larga
9 - Gavião
10 - Gorotire
11 - Guarani
12 - Jamamadi
13 - Kaingang
14 - Kali'na
15 - Karipuna
16 - Karitiana
17 - Krahó
18 - Kuben-Kran-Kegn
19 - Lengua
20 - Mapuche
21 - Mekranoti
22 - Munduruku
23 - Mura
24 - Pacaás Novos
25 - Parakanã
26 - Parintintin
27 - Sateré Mawé
28 - Suruí
29 - Tenharim
30 - Tiryió
31 - Txukahamãe
32 - Uru-Eu-Wau-Wau
33 - Urubu-Kaapor
34 - Wai-Wai
35 - Wayampi
36 - Xaváante
37 - Xikrin
38 - Zoró

Figure 2:

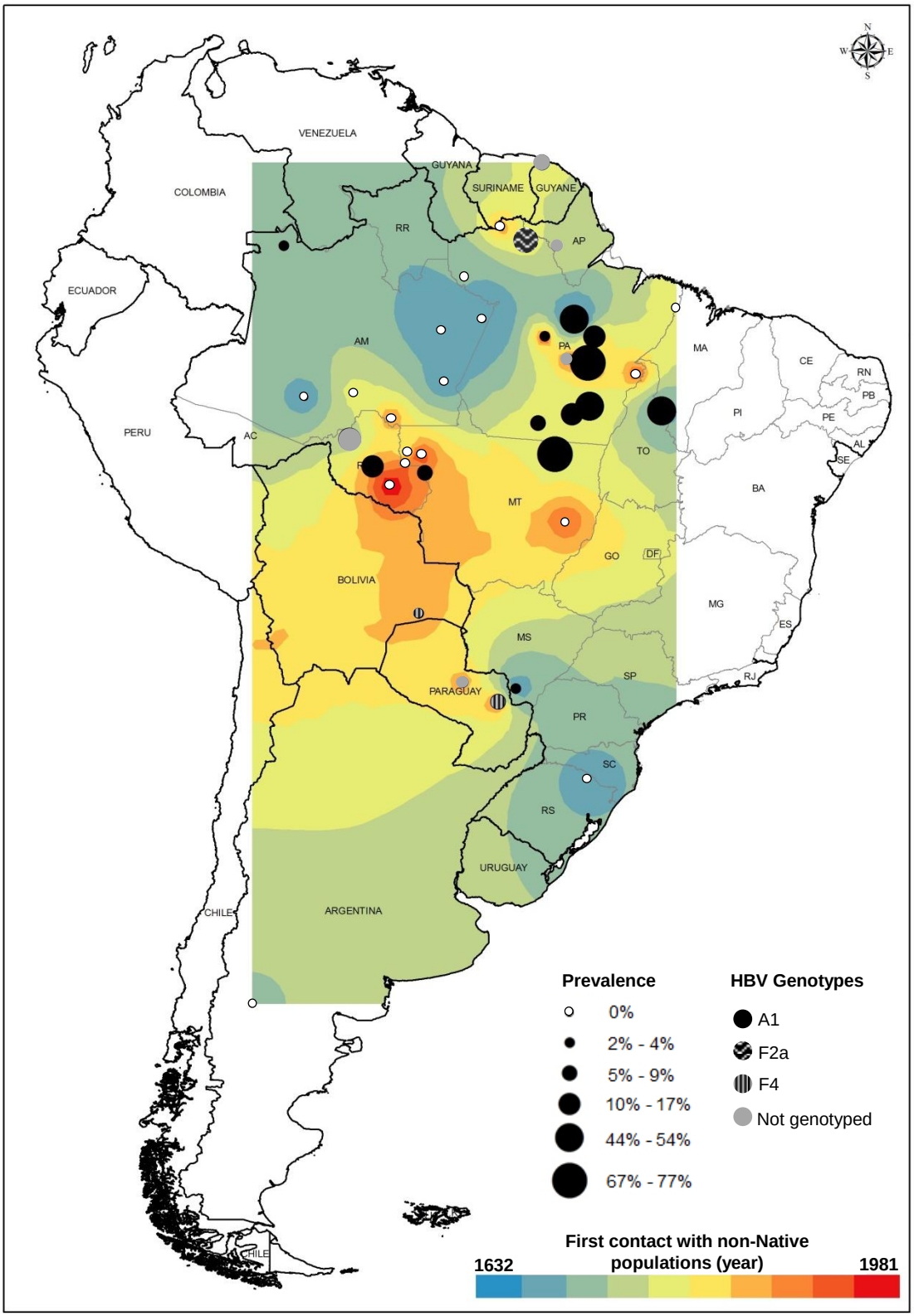

VENEZUELA
GUYANA
SURINAME
GUYANE
COLOMBIA
RR
AP
ECUADOR
AM
PA
MA
CE
RN
PB
PI
PE
AL
SE
PERU
AC
TO
BA
MT
GO
DF
BOLIVIA
MG
ES
MS
SP
RJ
PARAGUAY
PR
SC
RS
URUGUAY
CHILE
ARGENTINA
CHILE
Prevalence
0%
2% - 4%
5% - 9%
10% - 17%
44% - 54%
67% - 77%
HBV Genotypes
A1
F2a
F4
Not genotyped
First contact with non-Native populations (year)
1632
1981

Figure 3:

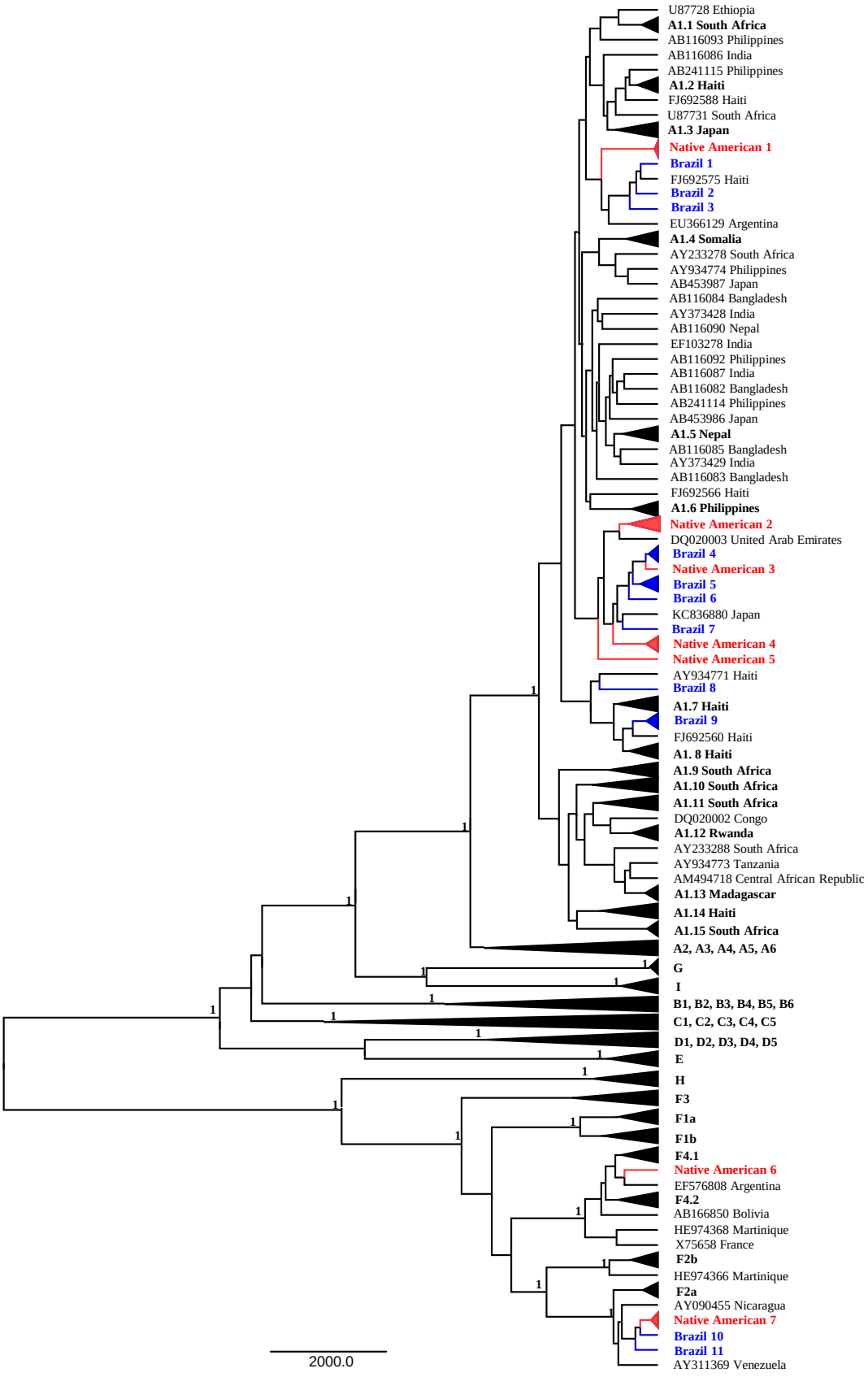
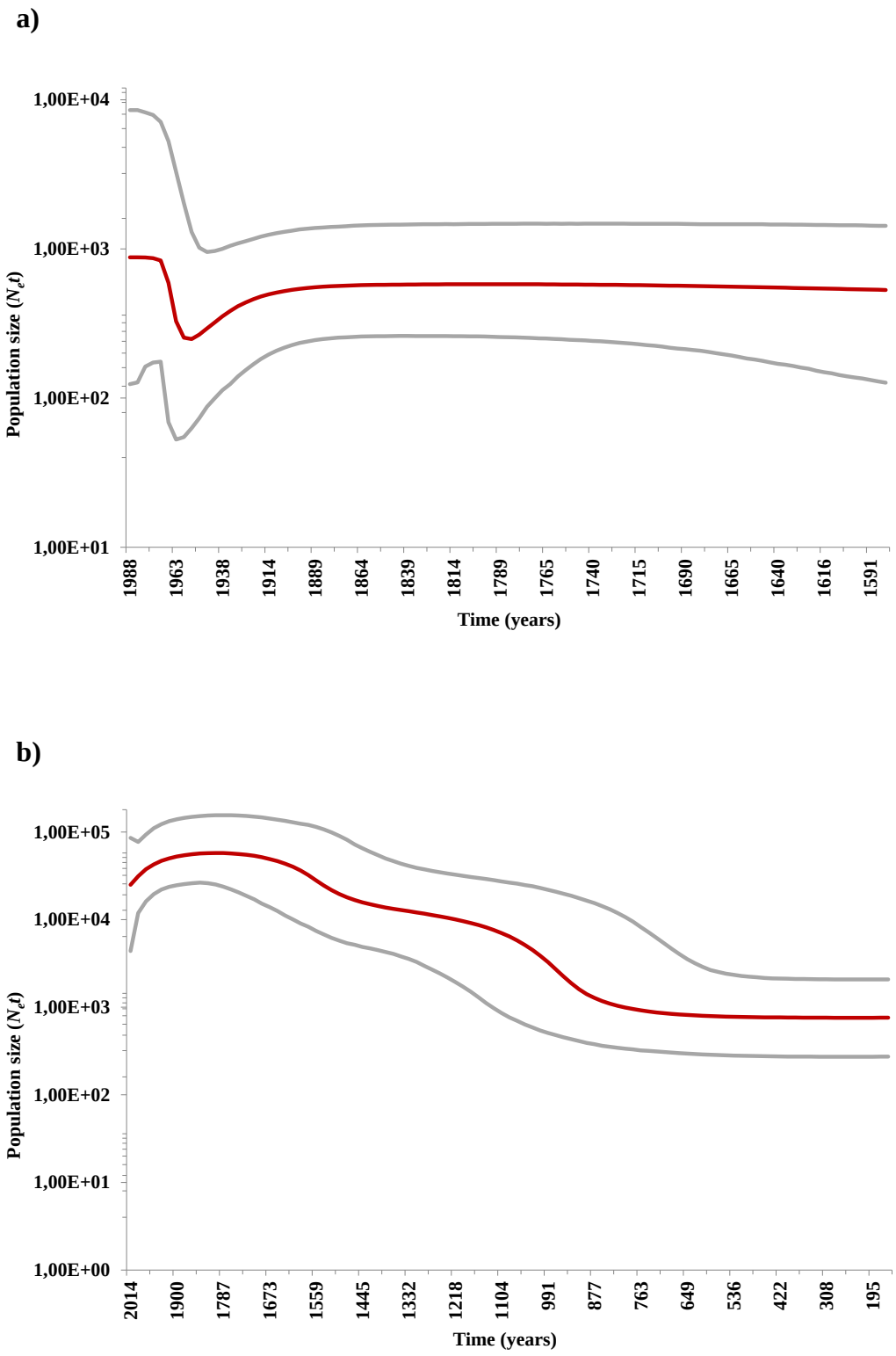

U87728 Ethiopia
A1.1 South Africa
AB116093 Philippines
AB116086 India
AB241115 Philippines
A1.2 Haiti
FJ692588 Haiti
U87731 South Africa
A1.3 Japan
Native American 1
Brazil 1
FJ692575 Haiti
Brazil 2
Brazil 3
EU366129 Argentina
A1.4 Somalia
AY233278 South Africa
AY934774 Philippines
AB453987 Japan
AB116084 Bangladesh
AY373428 India
AB116090 Nepal
EF103278 India
AB116092 Philippines
AB116087 India
AB116082 Bangladesh
AB241114 Philippines
AB453986 Japan
A1.5 Nepal
AB116085 Bangladesh
AY373429 India
AB116083 Bangladesh
FJ692566 Haiti
A1.6 Philippines
Native American 2
DQ020003 United Arab Emirates
Brazil 4
Native American 3
Brazil 5
Brazil 6
KC836880 Japan
Brazil 7
Native American 4
Native American 5
AY934771 Haiti
Brazil 8
A1.7 Haiti
Brazil 9
FJ692560 Haiti
A1. 8 Haiti
A1.9 South Africa
A1.10 South Africa
A1.11 South Africa
DQ020002 Congo
A1.12 Rwanda
AY233288 South Africa
AY934773 Tanzania
AM494718 Central African Republic
A1.13 Madagascar
A1.14 Haiti
A1.15 South Africa
A2, A3, A4, A5, A6
G
I
B1, B2, B3, B4, B5, B6
C1, C2, C3, C4, C5
D1, D2, D3, D4, D5
E
H
F3
F1a
F1b
F4.1
Native American 6
EF576808 Argentina
F4.2
AB166850 Bolivia
HE974368 Martinique
X75658 France
F2b
HE974366 Martinique
F2a
AY090455 Nicaragua
Native American 7
Brazil 10
Brazil 11
AY311369 Venezuela
2000.0

Figure 4:


a)
1,00E+04
1,00E+03
1,00E+02
1,00E+01
Population size ($N_e t$)
1988
1963
1938
1914
1889
1864
1839
1814
1789
1765
1740
1715
1690
1665
1640
1616
1591
Time (years)
b)
1,00E+05
1,00E+04
1,00E+03
1,00E+02
1,00E+01
1,00E+00
Population size ($N_e t$)
2014
1900
1787
1673
1559
1445
1332
1218
1104
991
877
763
649
536
422
308
195
Time (years)

Supplementary Information

Table S1: Sequences comprising totally or partially a 1306pb of HBsAg and DNA polymerase coding regions (S/Pol) of HBV genome included in this study.

Table S2: Sequences included in the collapsed branches or not specified in the phylogeny.

Table S3: Sequences of HBV/A1 from Brazilian non-Native Americans included in Bayesian Skyline Plot analysis.

Table S1: Sequences comprising totally or partially a 1306pb of HBsAg and DNA polymerase coding regions (S/Pol) of HBV genome included in this study.

Subgenotype	Accession number	Country
A1	AB116082	Philippines
A1	AB116083	Bangladesh
A1	AB116084	Bangladesh
A1	AB116085	Bangladesh
A1	AB116086	India
A1	AB116087	India
A1	AB116088	Nepal
A1	AB116089	Nepal
A1	AB116090	Nepal
A1	AB116091	Phillippines
A1	AB116092	Philippines
A1	AB116093	Philippines
A1	AB116094	Phillippines
A1	AB241115	Philippines
A1	AB453986	Japan
A1	AB453987	Japan
A1	AB453988	Japan
A1	AB453989	Japan
A1	AF297623	South Africa
A1	AF297625	South Africa
A1	AM494718	Central African Rebublic
A1	AY233274	South Africa
A1	AY233275	South Africa
A1	AY233276	South Africa
A1	AY233277	South Africa
A1	AY233278	South Africa
A1	AY233279	South Africa
A1	AY233281	South Africa

A1	AY233282	South Africa
A1	AY233283	South Africa
A1	AY233284	South Africa
A1	AY233285	South Africa
A1	AY233287	South Africa
A1	AY233288	South Africa
A1	AY233289	South Africa
A1	AY373428	India
A1	AY373429	India
A1	AY934765	South Africa
A1	AY934766	South Africa
A1	AY934768	Somalia
A1	AY934769	Somalia
A1	AY934770	Somalia
A1	AY934771	Haiti
A1	AY934773	Tanzania
A1	AY934774	Philippines
A1	DQ020002	Congo
A1	DQ020003	United Arab Emirates
A1	EF103278	India
A1	EU366129	Argentina
A1	FJ692557	Haiti
A1	FJ692558	Haiti
A1	FJ692559	Haiti
A1	FJ692560	Haiti
A1	FJ692561	Haiti
A1	FJ692562	Haiti
A1	FJ692563	Haiti
A1	FJ692564	Haiti
A1	FJ692565	Haiti
A1	FJ692566	Haiti
A1	FJ692567	Haiti

A1	FJ692568	Haiti
A1	FJ692569	Haiti
A1	FJ692571	Haiti
A1	FJ692572	Haiti
A1	FJ692574	Haiti
A1	FJ692575	Haiti
A1	FJ692576	Haiti
A1	FJ692577	Haiti
A1	FJ692578	Haiti
A1	FJ692579	Haiti
A1	FJ692580	Haiti
A1	FJ692581	Haiti
A1	FJ692582	Haiti
A1	FJ692584	Haiti
A1	FJ692585	Haiti
A1	FJ692587	Haiti
A1	FJ692588	Haiti
A1	FJ692590	Haiti
A1	FJ692591	Haiti
A1	FJ692592	Haiti
A1	FM199974	Rwanda
A1	FM199977	Rwanda
A1	FM199978	Rwanda
A1	FM199979	Rwanda
A1	FM199980	Rwanda
A1	FM199981	Rwanda
A1	FN821460	Madagascar
A1	FN821461	Madagascar
A1	FN821462	Madagascar
A1	HM101114	Brazil
A1	HM101115	Brazil
A1	HM101116	Brazil

A1	HM101117	Brazil
A1	HM101118	Brazil
A1	HM101119	Brazil
A1	HM101120	Brazil
A1	HM101121	Brazil
A1	HM101122	Brazil
A1	HM101123	Brazil
A1	HM101124	Brazil
A1	HM101129	Brazil
A1	HM772994	Brazil
A1	HM772995	Brazil
A1	HM772996	Brazil
A1	HM772997	Brazil
A1	KC836880	Japan
A1	U87727	South Africa
A1	U87728	Ethiopia
A1	U87731	Haiti
A1	U87733	South Africa
A2	AY934763	Gambia
A2	HE974364	Martinique
A2	HE974367	Martinique
A2	HE674374	Martinique
A2	HE974376	Martinique
A2	HE974383	Martinique
A3	AB194950	Cameroon
A3	AB194951	Cameroon
A3	AB194952	Cameroon
A3	AM184125	Gabon
A3	AM184126	Gabon
A4	AY934764	Gambia
A5	FJ692554	Nigeria
A5	FJ692555	Nigeria

A5	FJ692556	Nigeria
A5	FJ692593	Haiti
A5	FJ692594	Haiti
A5	FJ692595	Haiti
A5	FJ692596	Haiti
A5	FJ692597	Haiti
A5	FJ692598	Haiti
A5	FJ692599	Haiti
A5	FJ692600	Haiti
A5	FJ692601	Haiti
A5	FJ692602	Haiti
A5	FJ692603	Haiti
A5	FJ692604	Haiti
A5	FJ692605	Haiti
A5	FJ692606	Haiti
A5	FJ692607	Haiti
A5	FJ692608	Haiti
A5	FJ692609	Haiti
A5	FJ692610	Haiti
A5	FJ692611	Haiti
A5	FJ692612	Haiti
A5	FJ692613	Haiti
A6	GQ331046	Belgium
A6	GQ331047	Belgium
A6	GQ331048	Belgium
B1	AB010292	Japan
B1	AB073838	Japan
B2	AB073837	Japan
B2	AB073841	Taiwan
B3	AB033554	Indonesia
B3	D00331	Indonesia
B4	AB100695	Vietnam

B4	AY033072	Vietnam
B5	AB219426	Philippines
B6	AB287325	Greeland
B6	DQ463787	Canada
C1	AB074755	Thailand
C1	DQ315781	India
C2	AB113875	Japan
C2	AB205123	China
C3	X75656	Polynesia
C3	X75665	New Caledonia
C4	AB048704	Australia
C5	AB241111	Philippines
C5	AF241410	Vietnam
D1	AB104709	Egypt
D1	AB126581	Russia
D2	AB090268	India
D2	AB109475	Japan
D3	AJ344117	France
D3	AY233296	South Africa
D4	AB033559	Papua New Guinea
D4	AB048703	Australia
D5	DQ315779	India
E	AB091255	Ivory
E	AY739675	Germany
E	DQ060830	Madagascar
F1a	AY090456	Nicaragua
F1a	AY090458	Costa Rica
F1a	AY090459	Costa Rica
F1a	AY090461	Nicaragua
F1b	AB064316	USA
F1b	AB086397	Japan
F1b	AB116552	Venezuela

F1b	AB116654	Japan
F1b	AF223964	Argentina
F1b	AY179735	Argentina
F1b	DQ823094	Argentina
F1b	EU670262	Peru
F1b	FJ657525	Argentina
F1b	FJ657529	Argentina
F1b	FJ709458	Chile
F1b	FJ709459	Chile
F1b	FJ709460	Chile
F1b	HE981182	Argentina
F1b	HE981183	Argentina
F1b	HM467761	Colombia
F1b	HM585186	Chile
F1b	HM585187	Chile
F1b	HQ378247	Ireland
F1b	JN792913	Alaska
F1b	JN792915	Alaska
F1b	JN792917	USA
F1b	JN792918	USA
F1b	JN792921	Alaska
F1b	JN792922	USA
F2a	AY090455	Nicaragua
F2a	AY311369	Venezuela
F2a	DQ899142	Venezuela
F2a	DQ899143	Venezuela
F2a	HM101130	Brazil
F2a	X69798	Brazil
F2b	DQ899144	Venezuela
F2b	DQ899145	Venezuela
F2b	DQ899146	Venezuela
F2b	DQ899147	Venezuela

F2b	HE974366	Martinique
F3	AB036905	Venezuela
F3	AB036906	Venezuela
F3	AB036907	Venezuela
F3	AB036908	Venezuela
F3	AB036909	Venezuela
F3	AB036910	Venezuela
F3	AB036911	Venezuela
F3	AB036912	Venezuela
F3	AB036913	Venezuela
F3	AB036914	Venezuela
F3	AB036915	Venezuela
F3	AB036916	Venezuela
F3	AB036917	Venezuela
F3	AB036918	Venezuela
F3	AB036919	Venezuela
F3	AB036920	Venezuela
F3	AB116549	Panama
F3	AB116550	Panama
F3	AB116551	Venezuela
F3	AY311370	Venezuela
F3	DQ899149	Venezuela
F3	DQ899150	Colombia
F3	EF397987	Colombia
F3	EF397990	Colombia
F3	FJ589066	Venezuela
F3	FJ589067	Colombia
F3	HM467760	Colombia
F3	HM467762	Colombia
F3	HM467771	Colombia
F3	HM467772	Colombia
F3	HM467784	Colombia

F3	HM467785	Colombia
F3	HM467786	Colombia
F3	X75663	Colombia
F4	AB166850	Bolivia
F4	AB365446	Bolivia
F4	AB365447	Bolivia
F4	AB365448	Bolivia
F4	AB365449	Bolivia
F4	AB365450	Bolivia
F4	AB365453	Bolivia
F4	AF223965	Argentina
F4	DQ823086	Argentina
F4	EF576812	Argentina
F4	AY179734	Argentina
F4	DQ776247	Argentina
F4	DQ823087	Argentina
F4	DQ823088	Argentina
F4	DQ823089	Argentina
F4	DQ823090	Argentina
F4	EU366116	Argentina
F4	EU366132	Argentina
F4	FJ657519	Argentina
F4	FJ657522	Argentina
F4	FJ657528	Argentina
F4	JX079937	Argentina
H	AB298362	Japan
H	AY090457	Nicaragua
H	AY090460	USA
H	FJ356716	Argentina
H	HM117851	Mexico

Table S2: Sequences included in the collapsed branches or not specified in the phylogeny.

Genotype	Collapsed branch	Accession number	Population/ Country
A1	Native American 1	-	Pacaás Novos / Brazil
A1	Native American 1	-	Guarani / Brazil
A1	Native American 2	-	Baniwa / Brazil
A1	Native American 2	-	Cinta-Larga / Brazil
A1	Native American 2	-	Karipuna / Brazil
A1	Native American 2	-	Kuben Kran Kegn / Brazil
A1	Native American 2	-	Mekranoti / Brazil
A1	Native American 2	-	Parakanã / Brazil
A1	Native American 2	-	Txucahamãe / Brazil
A1	Native American 3	-	Asuriní / Brazil
A1	Native American 4	-	Gorotire / Brazil
A1	Native American 5	-	Arara / Brazil
F4	Native American 6	-	Aché / Paraguay
F2a	Native American 7	-	Apalaí / Brazil
A1	Brazil 1	HM101122	Brazil
A1	Brazil 2	HM101121	Brazil
A1	Brazil 3	HM101123	Brazil
A1	Brazil 4	HM772994	Brazil
A1	Brazil 4	HM772995	Brazil
A1	Brazil 4	HM772996	Brazil
A1	Brazil 4	HM772997	Brazil
A1	Brazil 5	HM101118	Brazil
A1	Brazil 5	HM101119	Brazil
A1	Brazil 5	HM101120	Brazil
A1	Brazil 6	HM101124	Brazil
A1	Brazil 7	HM101129	Brazil
A1	Brazil 8	HM101114	Brazil
A1	Brazil 9	HM101115	Brazil

A1	Brazil 9	HM101116	Brazil
A1	Brazil 9	HM101117	Brazil
A1	Brazil 10	HM101130	Brazil
A1	Brazil 11	X69798	Brazil
A1	A1.1	U87727	South Africa
A1	A1.1	U87733	South Africa
A1	A1.2	FJ692564	Haiti
A1	A1.2	FJ692565	Haiti
A1	A1.2	FJ692567	Haiti
A1	A1.2	FJ692590	Haiti
A1	A1.2	FJ692591	Haiti
A1	A1.3	AB453988	Japan
A1	A1.3	AB453989	Japan
A1	A1.4	AY934768	Somalia
A1	A1.4	AY934769	Somalia
A1	A1.4	AY934770	Somalia
A1	A1.5	AB116088	Nepal
A1	A1.5	AB116089	Nepal
A1	A1.6	AB116091	Phillippines
A1	A1.6	AB116094	Phillippines
A1	A1.7	FJ692558	Haiti
A1	A1.7	FJ692561	Haiti
A1	A1.7	FJ692562	Haiti
A1	A1.8	FJ692557	Haiti
A1	A1.8	FJ692559	Haiti
A1	A1.8	FJ692563	Haiti
A1	A1.8	FJ692568	Haiti
A1	A1.8	FJ692569	Haiti
A1	A1.8	FJ692571	Haiti
A1	A1.8	FJ692572	Haiti
A1	A1.8	FJ692574	Haiti
A1	A1.8	FJ692576	Haiti

A1	A1.8	FJ692577	Haiti
A1	A1.8	FJ692578	Haiti
A1	A1.8	FJ692579	Haiti
A1	A1.8	FJ692580	Haiti
A1	A1.8	FJ692581	Haiti
A1	A1.8	FJ692582	Haiti
A1	A1.8	FJ692584	Haiti
A1	A1.8	FJ692585	Haiti
A1	A1.9	AY233276	South Africa
A1	A1.9	AY233279	South Africa
A1	A1.9	AY233287	South Africa
A1	A1.10	AF297625	South Africa
A1	A1.10	AY233274	South Africa
A1	A1.10	AY233277	South Africa
A1	A1.10	AY233282	South Africa
A1	A1.10	AY233283	South Africa
A1	A1.10	AY233284	South Africa
A1	A1.10	AY233285	South Africa
A1	A1.10	AY233289	South Africa
A1	A1.11	AF297623	South Africa
A1	A1.11	AY934765	South Africa
A1	A1.11	AY934766	South Africa
A1	A1.12	FM199974	Rwanda
A1	A1.12	FM199977	Rwanda
A1	A1.12	FM199979	Rwanda
A1	A1.12	FM199980	Rwanda
A1	A1.12	FM199981	Rwanda
A1	A1.13	FN821460	Madagascar
A1	A1.13	FN821461	Madagascar
A1	A1.13	FN821462	Madagascar
A1	A1.14	FJ692587	Haiti
A1	A1.14	FJ692592	Haiti

A1	A1.15	AY233275	South Africa
A1	A1.15	AY233281	South Africa
A2	A2, A3, A4, A5, A6	AY934763	Gambia
A2	A2, A3, A4, A5, A6	HE974364	Martinique
A2	A2, A3, A4, A5, A6	HE974367	Martinique
A2	A2, A3, A4, A5, A6	HE674374	Martinique
A2	A2, A3, A4, A5, A6	HE974376	Martinique
A2	A2, A3, A4, A5, A6	HE974383	Martinique
A3	A2, A3, A4, A5, A6	AB194950	Cameroon
A3	A2, A3, A4, A5, A6	AB194951	Cameroon
A3	A2, A3, A4, A5, A6	AB194952	Cameroon
A3	A2, A3, A4, A5, A6	AM184125	Gabon
A3	A2, A3, A4, A5, A6	AM184126	Gabon
A4	A2, A3, A4, A5, A6	AY934764	Gambia
A5	A2, A3, A4, A5, A6	FJ692554	Nigeria
A5	A2, A3, A4, A5, A6	FJ692555	Nigeria
A5	A2, A3, A4, A5, A6	FJ692556	Nigeria
A5	A2, A3, A4, A5, A6	FJ692593	Haiti
A5	A2, A3, A4, A5, A6	FJ692594	Haiti
A5	A2, A3, A4, A5, A6	FJ692595	Haiti
A5	A2, A3, A4, A5, A6	FJ692596	Haiti
A5	A2, A3, A4, A5, A6	FJ692597	Haiti
A5	A2, A3, A4, A5, A6	FJ692598	Haiti
A5	A2, A3, A4, A5, A6	FJ692599	Haiti
A5	A2, A3, A4, A5, A6	FJ692600	Haiti
A5	A2, A3, A4, A5, A6	FJ692601	Haiti
A5	A2, A3, A4, A5, A6	FJ692602	Haiti
A5	A2, A3, A4, A5, A6	FJ692603	Haiti
A5	A2, A3, A4, A5, A6	FJ692604	Haiti
A5	A2, A3, A4, A5, A6	FJ692605	Haiti
A5	A2, A3, A4, A5, A6	FJ692606	Haiti
A5	A2, A3, A4, A5, A6	FJ692607	Haiti

A5	A2, A3, A4, A5, A6	FJ692608	Haiti
A5	A2, A3, A4, A5, A6	FJ692609	Haiti
A5	A2, A3, A4, A5, A6	FJ692610	Haiti
A5	A2, A3, A4, A5, A6	FJ692611	Haiti
A5	A2, A3, A4, A5, A6	FJ692612	Haiti
A5	A2, A3, A4, A5, A6	FJ692613	Haiti
A6	A2, A3, A4, A5, A6	GQ331046	Belgium
A6	A2, A3, A4, A5, A6	GQ331047	Belgium
A6	A2, A3, A4, A5, A6	GQ331048	Belgium
G	G	AB056513	USA
G	G	AF405706	HE
I	I	AF241407	Vietnam
I	I	FJ023661	Laos
B1	B1, B2, B3, B4, B5, B6	AB010292	Japan
B1	B1, B2, B3, B4, B5, B6	AB073838	Japan
B2	B1, B2, B3, B4, B5, B6	AB073837	Japan
B2	B1, B2, B3, B4, B5, B6	AB073841	Taiwan
B3	B1, B2, B3, B4, B5, B6	AB033554	Indonesia
B3	B1, B2, B3, B4, B5, B6	D00331	Indonesia
B4	B1, B2, B3, B4, B5, B6	AB100695	Vietnam
B4	B1, B2, B3, B4, B5, B6	AY033072	Vietnam
B5	B1, B2, B3, B4, B5, B6	AB219426	Philippines
B6	B1, B2, B3, B4, B5, B6	AB287325	Greeland
B6	B1, B2, B3, B4, B5, B6	DQ463787	Canada
C1	C1, C2, C3, C4, C5	AB074755	Thailand
C1	C1, C2, C3, C4, C5	DQ315781	India
C2	C1, C2, C3, C4, C5	AB113875	Japan
C2	C1, C2, C3, C4, C5	AB205123	China
C3	C1, C2, C3, C4, C5	X75656	Polynesia
C3	C1, C2, C3, C4, C5	X75665	New Caledonia
C4	C1, C2, C3, C4, C5	AB048704	Australia
C5	C1, C2, C3, C4, C5	AB241111	Philippines

C5	C1, C2, C3, C4, C5	AF241410	Vietnam
D1	D1, D2, D3, D4, D5	AB104709	Egypt
D1	D1, D2, D3, D4, D5	AB126581	Russia
D2	D1, D2, D3, D4, D5	AB090268	India
D2	D1, D2, D3, D4, D5	AB109475	Japan
D3	D1, D2, D3, D4, D5	AJ344117	France
D3	D1, D2, D3, D4, D5	AY233296	South Africa
D4	D1, D2, D3, D4, D5	AB033559	Papua New Guinea
D4	D1, D2, D3, D4, D5	AB048703	Australia
D5	D1, D2, D3, D4, D5	DQ315779	India
E	E	AB091255	Ivory
E	E	AY739675	Germany
E	E	DQ060830	Madagascar
H	H	AB298362	Japan
H	H	AY090457	Nicaragua
H	H	AY090460	USA
H	H	FJ356716	Argentina
H	H	HM117851	Mexico
F1a	F1a	AY090456	Nicaragua
F1a	F1a	AY090458	Costa Rica
F1a	F1a	AY090459	Costa Rica
F1a	F1a	AY090461	Nicaragua
F1b	F1b	AB064316	USA
F1b	F1b	AB086397	Japan
F1b	F1b	AB116552	Venezuela
F1b	F1b	AB116654	Japan
F1b	F1b	AF223964	Argentina
F1b	F1b	AY179735	Argentina
F1b	F1b	DQ823094	Argentina
F1b	F1b	EU670262	Peru
F1b	F1b	FJ657525	Argentina
F1b	F1b	FJ657529	Argentina

F1b	F1b	FJ709458	Chile
F1b	F1b	FJ709459	Chile
F1b	F1b	FJ709460	Chile
F1b	F1b	HE981182	Argentina
F1b	F1b	HE981183	Argentina
F1b	F1b	HM467761	Colombia
F1b	F1b	HM585186	Chile
F1b	F1b	HM585187	Chile
F1b	F1b	HQ378247	Ireland
F1b	F1b	JN792913	Alaska
F1b	F1b	JN792915	Alaska
F1b	F1b	JN792917	USA
F1b	F1b	JN792918	USA
F1b	F1b	JN792921	Alaska
F1b	F1b	JN792922	USA
F2a	F2a	DQ899142	Venezuela
F2a	F2a	DQ899143	Venezuela
F2b	F2b	DQ899144	Venezuela
F2b	F2b	DQ899145	Venezuela
F2b	F2b	DQ899146	Venezuela
F2b	F2b	DQ899147	Venezuela
F2b	F2b	HE974366	Martinique
F3	F3	AB036905	Venezuela
F3	F3	AB036906	Venezuela
F3	F3	AB036907	Venezuela
F3	F3	AB036908	Venezuela
F3	F3	AB036909	Venezuela
F3	F3	AB036910	Venezuela
F3	F3	AB036911	Venezuela
F3	F3	AB036912	Venezuela
F3	F3	AB036913	Venezuela
F3	F3	AB036914	Venezuela

F3	F3	AB036915	Venezuela
F3	F3	AB036916	Venezuela
F3	F3	AB036917	Venezuela
F3	F3	AB036918	Venezuela
F3	F3	AB036919	Venezuela
F3	F3	AB036920	Venezuela
F3	F3	AB116549	Panama
F3	F3	AB116550	Panama
F3	F3	AB116551	Venezuela
F3	F3	AY311370	Venezuela
F3	F3	DQ899149	Venezuela
F3	F3	DQ899150	Colombia
F3	F3	EF397987	Colombia
F3	F3	EF397990	Colombia
F3	F3	FJ589066	Venezuela
F3	F3	FJ589067	Colombia
F3	F3	HM467760	Colombia
F3	F3	HM467762	Colombia
F3	F3	HM467771	Colombia
F3	F3	HM467772	Colombia
F3	F3	HM467784	Colombia
F3	F3	HM467785	Colombia
F3	F3	HM467786	Colombia
F3	F3	X75663	Colombia
F4	F4.1	AB166850	Bolivia
F4	F4.1	AB365446	Bolivia
F4	F4.1	AB365447	Bolivia
F4	F4.1	AB365448	Bolivia
F4	F4.1	AB365449	Bolivia
F4	F4.1	AB365450	Bolivia
F4	F4.1	AB365453	Bolivia
F4	F4.1	AF223965	Argentina

F4	F4.1	DQ823086	Argentina
F4	F4.1	EF576812	Argentina
F4	F4.2	AY179734	Argentina
F4	F4.2	DQ776247	Argentina
F4	F4.2	DQ823087	Argentina
F4	F4.2	DQ823088	Argentina
F4	F4.2	DQ823089	Argentina
F4	F4.2	DQ823090	Argentina
F4	F4.2	EU366116	Argentina
F4	F4.2	EU366132	Argentina
F4	F4.2	FJ657519	Argentina
F4	F4.2	FJ657522	Argentina
F4	F4.2	FJ657528	Argentina
F4	F4.2	JX079937	Argentina

Table S3: Sequences of HBV/A1 from Brazilian non-Native Americans included in Bayesian Skyline Plot analysis.

Subgenotype	Accession number	Country
A1	FJ174794	Brazil
A1	FJ174795	Brazil
A1	FJ174796	Brazil
A1	FJ174797	Brazil
A1	FJ174799	Brazil
A1	FJ174800	Brazil
A1	HM101113	Brazil
A1	HM101114	Brazil
A1	HM101115	Brazil
A1	HM101116	Brazil
A1	HM101117	Brazil
A1	HM101118	Brazil
A1	HM101119	Brazil
A1	HM101120	Brazil
A1	HM101121	Brazil
A1	HM101122	Brazil
A1	HM101123	Brazil
A1	HM101124	Brazil
A1	HM101129	Brazil
A1	HBU55220	Brazil
A1	HBU55221	Brazil
A1	HBU55222	Brazil
A1	EF547829	Brazil
A1	EF547830	Brazil
A1	EF547831	Brazil
A1	EF547832	Brazil
A1	EF547833	Brazil
A1	EF547834	Brazil
A1	EF547835	Brazil

A1	EF547836	Brazil
A1	EF547837	Brazil
A1	EF547838	Brazil
A1	EF547839	Brazil
A1	EF547840	Brazil
A1	EF547841	Brazil
A1	EF547842	Brazil
A1	EF547843	Brazil
A1	EF547844	Brazil
A1	EF547845	Brazil
A1	EF547846	Brazil
A1	EF547847	Brazil
A1	EF547848	Brazil
A1	EF547849	Brazil
A1	EF547850	Brazil
A1	EF547851	Brazil
A1	EF547852	Brazil
A1	EF547853	Brazil
A1	EF547854	Brazil
A1	EF547855	Brazil
A1	EF547856	Brazil
A1	AY344098	Brazil
A1	AY344099	Brazil
A1	AY344100	Brazil
A1	AY344101	Brazil
A1	AY344102	Brazil
A1	AY344103	Brazil
A1	AY344104	Brazil
A1	AY344105	Brazil
A1	AY344106	Brazil
A1	AY344107	Brazil
A1	AY344108	Brazil

A1	AY344109	Brazil
A1	AY344110	Brazil
A1	AY344111	Brazil
A1	AY344112	Brazil
A1	HM772994	Brazil
A1	HM772995	Brazil
A1	HM772996	Brazil
A1	HM772997	Brazil
A1	JF784214	Brazil
A1	JF784215	Brazil
A1	JF784216	Brazil
A1	JF784217	Brazil
A1	JF784218	Brazil
A1	JF784219	Brazil
A1	JF784223	Brazil
A1	JF784224	Brazil
A1	JF784225	Brazil
A1	JF784226	Brazil
A1	JF784227	Brazil
A1	JF784228	Brazil
A1	JF784229	Brazil
A1	JF784230	Brazil
A1	JF784231	Brazil
A1	JF784232	Brazil
A1	JF784234	Brazil
A1	JF784236	Brazil
A1	JF784239	Brazil
A1	JF784240	Brazil
A1	JF784241	Brazil
A1	JF784242	Brazil
A1	JF784244	Brazil
A1	JF784246	Brazil

A1	JF784247	Brazil
A1	JF784248	Brazil
A1	JF784249	Brazil
A1	JF784250	Brazil
A1	HQ646059	Brazil
A1	HQ646060	Brazil
A1	HQ646061	Brazil
A1	HQ646062	Brazil
A1	HQ646063	Brazil
A1	HQ646064	Brazil
A1	HQ646065	Brazil
A1	HQ646066	Brazil
A1	HQ646067	Brazil
A1	HQ646068	Brazil
A1	HQ646069	Brazil
A1	HQ646070	Brazil
A1	HQ646071	Brazil
A1	HQ646072	Brazil
A1	HQ646073	Brazil
A1	HQ646074	Brazil
A1	HQ646075	Brazil
A1	HQ646077	Brazil
A1	HQ646078	Brazil
A1	HQ646079	Brazil
A1	HQ646080	Brazil
A1	HQ646081	Brazil
A1	HQ646082	Brazil
A1	HQ646083	Brazil
A1	HQ646084	Brazil
A1	HQ646086	Brazil
A1	HQ646087	Brazil
A1	HQ646088	Brazil

A1	HQ646089	Brazil
A1	HQ646090	Brazil
A1	HQ646091	Brazil
A1	HQ646092	Brazil
A1	HQ646093	Brazil
A1	HQ646094	Brazil
A1	HQ646095	Brazil
A1	HQ646096	Brazil
A1	HQ646097	Brazil
A1	HQ646098	Brazil
A1	JF784220	Brazil
A1	JF784221	Brazil
A1	JF784235	Brazil
A1	JF784243	Brazil
A1	EF690524	Brazil
A1	EU 264113	Brazil
A1	EU 264114	Brazil
A1	EU 264115	Brazil
A1	EU 264116	Brazil
A1	EU 264117	Brazil
A1	EU 264118	Brazil
A1	EU 264119	Brazil
A1	EU 264120	Brazil
A1	EU 264121	Brazil
A1	EU 264122	Brazil
A1	EU 264123	Brazil
A1	EU 264124	Brazil
A1	EU 264125	Brazil
A1	EU 264126	Brazil
A1	EU 264127	Brazil
A1	EU 264128	Brazil
A1	EU 264129	Brazil

A1	EU 264130	Brazil
A1	EU 264131	Brazil
A1	EU 264132	Brazil
A1	EU 264133	Brazil
A1	EU 264134	Brazil
A1	EU 264135	Brazil
A1	EU 264136	Brazil
A1	EU 264138	Brazil
A1	EU 264139	Brazil
A1	EU 264140	Brazil
A1	EU 264141	Brazil
A1	EU 264143	Brazil
A1	EU 264144	Brazil
A1	EU 264145	Brazil
A1	EU 264146	Brazil
A1	EU 264147	Brazil
A1	EU 264149	Brazil
A1	EU 264150	Brazil
A1	EU 264151	Brazil
A1	EU 264152	Brazil
A1	EU 264154	Brazil
A1	EU 264155	Brazil
A1	EU 264156	Brazil
A1	FJ010658	Brazil
A1	FJ010659	Brazil
A1	FJ010660	Brazil
A1	FJ010661	Brazil
A1	JF815647	Brazil

CAPÍTULO V – DISCUSSÃO GERAL

5. DISCUSSÃO GERAL

5.1. HBV e Populações Nativas Americanas

Os genótipos F e H são considerados autóctones do continente americano e possuem uma distribuição bastante variável entre os países. O genótipo H é altamente prevalente no México, enquanto o genótipo F, que é posteriormente classificado em 4 subgenótipos (Devesa *et al.*, 2008), apresenta uma ampla distribuição entre os países americanos sendo relacionado, principalmente, com Nativos Americanos. O subgenótipo F1a é predominante na América Central, enquanto seu grupo irmão F1b ocorre no Peru, Chile e Argentina. Subgenótipo F2a ocorre no Norte do Brasil e na Venezuela, enquanto o subgenótipo F2b é observado apenas na Venezuela. F3 é descrito na Colômbia, Panamá e Venezuela, e F4 é restrito à região do Chaco, principalmente na Bolívia e Argentina (revisado por Alvarado-Mora *et al.*, 2013b). No presente estudo nós mostramos a alta prevalência de HBV em uma amostra histórica de Nativos Americanos (10,2%) e descrevemos o marcante predomínio do subgenótipo A1 nessas populações (94,3%). A baixa frequência de genótipos F (F2a -2,85%; F4 - 2,85%) foi inicialmente surpreendente pelo fato deste genótipo estar relacionado, principalmente, com populações Nativas Americanas isoladas (Castilho *et al.*, 2012), o que corresponderia à amostra utilizada nesse estudo. Entretanto, embora exista maior prevalência do genótipo F nos demais países da América Latina, no Brasil, devido a maior miscigenação populacional, uma distinta distribuição genotípica é observada, com o genótipo A sendo descrito como o mais frequente (Sitnik *et al.*, 2004; Mello *et al.*, 2007; Alvarado-Mora *et al.*, 2013b). Essa maior frequência, especialmente de subgenótipo A1, é atribuída ao tráfico de escravos africanos para o Brasil durante o período colonial (século XVI a XIX) que teria servido como um dispersor desta cepa viral (Mello *et al.*, 2007). Um estudo prévio conduzido por Mello *et al* (2013), também relata baixas frequências de genótipo F e a predominante ocorrência do genótipo A na região da bacia amazônica, sugerindo um mudança no perfil genotípico das populações nativas em decorrência do fluxo migratório em direção à Amazônia ocorrido entre o fim do século XIX e o início do século XX. Nossas análises de dinâmica populacional do subgenótipo A1 mostram padrões diferentes entre as cepas virais isoladas de populações Nativas Americanas e de outras amostras brasileiras, sugerindo que o padrão observado na nossa amostra é associado a um período de tempo e processos

históricos diferentes daqueles relacionados ao Brasil como um todo. Nossos resultados sugerem que após a chegada do subgenótipo A1 em diferentes regiões do Brasil (provavelmente associado ao tráfico de escravos africanos), a ocupação da Amazônia para extração de recursos da floresta e construção de rodovias por volta de 1950-1970, (Martinello, 2004; Fearnside, 2005), resultou na transmissão deste subgenótipo para as populações Nativas Americanas.

Nossos resultados mostram a ocorrência de um único genótipo de HBV circulante por população. A existência de baixas condições de higiene e saneamento, bem como práticas culturais específicas presentes nessas populações (tatuagem e colocação de adornos, por exemplo) poderiam facilitar a transmissão de HBV (Duarte *et al.*, 2010; Te *et al.*, 2010) e, com isso, a ocorrência de coinfecção com diferentes cepas. A falta de associação entre genótipo de HBV e classificação linguística ou localização das tribos é indicativa de que hábitos específicos não parecem influenciar muito sobre a disseminação da doença na nossa amostra. Não é possível avaliar se essas populações não possuíam HBV antes da chegada do genótipo A1 ou se poderia ter havido uma substituição do genótipo F pelo genótipo A1 após a chegada deste. O grande predomínio do subgenótipo A1 poderia ser visto como um indício de competição inter-genotípica ou ocorrência de deriva genética intra-populacional fixando um único subgenótipo do HBV. Entretanto, a baixa prevalência de HBV (e, portanto, de HBV-F) em diversas populações Nativas na porção oeste da Amazônia pode ser vista como uma indicação mais forte de uma prevalência baixa do HBV nessa região antes da chegada do subgenótipo A1.

A falta de informações clínicas sobre os indivíduos positivos para HBV representa uma limitação do nosso trabalho. Populações Nativas Americanas apresentam alta incidência de infecção oculta de HBV (Roman *et al*., 2010; Cardona *et al*., 2011) e são associadas com desfechos mais agressivos da doença (Casey *et al*., 1996; Quintero *et al*., 2001; Parana *et al.*, 2005). Além disso, estudos recentes mostram que perfis genéticos do indivíduo e do patógeno, juntamente com a ancestralidade de ambos influenciam nas manifestações da doença (Devesa *et al*., 2007; Kodaman et al., 2014), mostrando que a análise de informações genéticas e clínicas tem grande importância. Nesse sentido, a presença de um genótipo "africano" de HBV em populações de ancestralidade Nativa Americana representa um achado importante para o monitoramento de hepatites virais

nessas populações e para guiar políticas de saúde pública de prevenção da doença e controle de seus efeitos.

5.2. HBV e História Evolutiva

A origem e história evolutiva do HBV ainda não estão completamente entendidas. Se por um lado evidências como a forte estruturação geográfica dos genótipos de HBV ou a detecção de um genótipo atual do vírus (C2) em uma múmia coreana de 330 anos (Bar-Gal *et al.*, 2013) são indicativos de uma origem antiga, por outro, algumas estimativas de datação utilizando relógio molecular resultam em datas extremamente recentes que não condizem com processos históricos relacionados com a sua distribuição (Zhou *et al*., 2007; Alvarado-Mora *et al*., 2010; Torres *et al*., 2011). A datação molecular para a origem e divergência dos genótipos de HBV é extremamente dependente da taxa evolutiva assumida. Como a taxa evolutiva está diretamente relacionada com a taxa de substituição utilizada, fatores como a existência de uma polimerase viral sem atividade de correção de erro e presença de ORFs sobrepostas que acabam por gerar variação nos padrões evolutivos de diferentes regiões ao longo do genoma, têm gerado dificuldade para a determinação de uma taxa evolutiva coerente (Alvarado-Mora *et al.*, 2011; Torres *et al.*, 2011).

Mantendo o foco sobre a história do HBV no continente americano, através da utilização de amostras disponíveis dos dois genótipos originários da América (F e H), nossas análises mostraram um excesso de substituições não sinônimas ocorrendo em ramos recentes da filogenia, quando comparados a ramos antigos ou intermediários, sugerindo a ocorrência de seleção purificadora na eliminação de mutações deletérias presentes nos ramos recentes. Esse excesso observado em amostras recentes pode estar influenciando na geração de taxas evolutivas rápidas demais, o que pode tornar as datações para HBV mais recentes do que o esperado (Ho *et al*., 2011; Crandall *et al*., 2012). Estudos prévios tem sugerido que amostras datadas sejam utilizadas nas análises a fim de obter estimativas mais coerentes (Firth *et al*., 2010). Entretanto, a disponibilidade de sequências datadas nos bancos de sequências ainda é limitada e, mesmo quando esse dado está disponível, pode conter algum tipo de viés que afete a análise dos dados (Bouckaert *et al.*, 2013).

Para origem dos genótipos ‘americanos’ F e H, nossos resultados mostraram que, utilizando taxas evolutivas previamente estimadas (Torres *et al.*, 2011) e diferentes

modelos de relógio molecular, este evento pode ser tão antigo quanto ~19,000 anos, condizendo com a chegada dos primeiros hominídeos na América, ou tão recentes quanto ~6,000 anos, o que discordaria da história e distribuição vinculada ao HBV. Essa enorme discrepância entre as análises, unida à incerteza da taxa evolutiva vinculada ao HBV impede a definição de datações mais coerentes e o melhor entendimento da origem e evolução deste vírus na América. Desta forma, ocorrência de seleção purificadora, e a utilização de amostras datadas de maneira mais específica bem como de diversos pontos de calibração devem ser consideradas na realização das análises de datação molecular para HBV a fim de esclarecer melhor a história evolutiva vinculada a este vírus.

CAPÍTULO VI – CONCLUSÕES E PERSPECTIVAS

6. CONCLUSÕES E PERSPECTIVAS

Nossos resultados mostram a alta prevalência do subgenótipo A1 de HBV em uma amostra histórica de Nativos Americanos e evidencia a ocorrência de diferentes dinâmicas populacionais para este subgenótipo no Brasil. Além disso, mostramos que análises de datação para HBV são altamente dependentes dos parâmetros utilizados, sendo necessário considerar a ocorrência de seleção purificadora, além de amostras datadas e diferentes pontos de calibração a fim de produzir estimativas mais razoáveis.

As principais perspectivas deste trabalho envolvem o *screening* para mutações de resistência a medicamentos nas amostras Nativas Americanas, seguido por uma melhor caracterização do subgenótipo A1 de HBV através da obtenção de mais amostras brasileiras que possam fornecer também dados clínicos e imunológicos para inferência mais completas sobre a patogenicidade do HBV, além de esclarecer o processo histórico de expansão do sugenótipo A1 ao longo da bacia Amazônica. Finalmente, para um melhor entendimento da história evolutiva do HBV, a avaliação do papel da seleção purificadora deverá ser realizada também para os demais genótipos.

CAPÍTULO VII – REFERÊNCIAS BIBLIOGRÁFICAS

7. REFERÊNCIAS BIBLIOGRÁFICAS

Alvarado-Mora MV, Romano CM, Gomes-Gouvêa MS, Gutierrez MF, *et al*. Molecular epidemiology and genetic diversity of hepatitis B virus genotype E in an isolated Afro-Colombian community. *J Gen Virol* 2010; **91:**501-508.

Alvarado-Mora MV, Pinho JRR. Epidemiological update of hepatites B, C and delta in Latin America. *Antiviral Therapy* 2013a; **18:**429-433.

Alvarado-Mora MV, Pinho JRR (2013b) Distribution of HBV genotypes in Latin America. *Antiviral Therapy* 2013b; **18:**459-465.Alvarado-Mora MV, Locarnini S, Rizzetto M *et al.* An update on HDV: virology, pathogenesis and treatment. *Antiviral Therapy* 2013c; **18:**541-548.

Andernach IE, Nolte C, Pape JW, et *al*. Slave trade and hepatitis B virus genotypes and subgenotypes in Haiti and Africa. *Emerg Infect Dis* 2009; **15(8):**1222–8.

Arauz-Ruiz P, Norder H, Robertson BH, *et al*. Genotype H: a new Amerindian genotype of hepatitis B virus revealed in Central America. *J Gen Virol* 2002; **83:**2059–2073.

Blitz L, Pujol F, Swenson P, *et al*. Antigenic diversity of hepatitis B virus strain of genotype F in Amerindians and other population groups from Venezuela. *J Clin Microbiol* 1998; **36:**648-51.

Block TM, Guo H, Guo JT. Molecular virology of hepatitis B virus for clinicians. *Clin Liver Dis* 2007; **11:**685 e 706, vii.

Bouckaert R, Alvarado-Mora MV, Pinho JRR. Evolutionary rates and HBV: issues of rate estimation with Bayesian molecular methods. *Antivir Ther* 2013; **18:**497-503.

Braga WS. Hepatitis B and D virus infection within Amerindians ethnic groups in the Brazilian Amazon: epidemiological aspects. *Rev. Soc. Bras. Med. Trop* 2004; **37:**S9–S13.

Bruss V, Ganem D. The role of envelope proteins in hepatitis B virus assembly. *Proc Natl Acad Sci USA* 1991; **88:**1059-63.

Campos RH, Mbayed VA, Pineiro Y, *et al*. Molecular epidemiology of hepatitis B virus in Latin America. *J ClinVirol* 2005; **34 Suppl 2:**S8-S13.

Cardona NE, Loureiro CL, Garzaro DJ, *et al*. Unusual presentation of hepatitis B serological markers in an Amerindian community of Venezuela with a majority of occult cases. *Virology Journal* 2011, **8:**527.

Casey JL, Niro GA, Engle RE, et al. 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* 1996; **174:**920–926.

Castilho MC, Oliveira CMC, Gimaque JBL, *et al*. Epidemiology and Molecular Characterization of Hepatitis B Virus Infection in Isolated Villages in the Western Brazilian Amazon. *Am J Trop Med Hyg* 2012; **87(4) pp.** 768–774.

Chang MH, Chen CJ, Lai MS, *et al*. Universal hepatitis B vaccination in Taiwan and the incidence of hepatocellular carcinoma in children. Taiwan Childhood Hepatoma Study Group. *N Engl J Med* 1997; **336:**1855–9.

Chang MH, You SL, Chen CJ, *et al*. Decreased incidence of hepatocellular carcinoma in hepatitis B vaccines: a 20-year follow-up study. *J Natl Cancer Inst* 2009; **101:**1348–55.

Chu CJ, Keefe EB, Han SH, *et al*. Hepatitis B virus genotypes in the United States: results of a nationwide estudy. *Gastroenterology* 2003; **125:**444–451.

Crandall ED, Sbrocco EJ, DeBoer TS, *et al*. Expansion dating: calibrating molecular clocks in marine species from expansions onto the Sunda Shelf following the last glacial maximum. *Mol Biol Evol* 2012; **29:**707–719.

Dandri M, Locarnini S. New insight in the pathobiology of hepatitis B virus infection. *Gut* 2012; **61 Suppl 1:**i6–17.

Dane DS, Cameron CH, Briggs M. Virus-like particles in the serum of patients with Australia-antigen-associated hepatitis. *Lancet* 1970; **1:**695-8.

Devesa M, Pujol FH. Hepatitis B virus genetic diversity in Latin America. *Virus Res* 2007; **127(2):**177-84.

Devesa M, Loureiro CL, Rivas Y, *et al.* Subgenotype Diversity of Hepatitis B Virus American Genotype F in Amerindians From Venezuela and the General Population of Colombia. *Journal of Medical Virology* 2008; **80:**20–26.

Duarte MC, Cardona N, Poblete F, *et al.*A comparative epidemiological study of hepatitis B and hepatitis D virus infections in Yanomami and Piaroa Amerindians of Amazonas State, Venezuela. *Trop Med Int Health* 2010; **15(8):**924-33.

Echevarria JM, Leon P. Epidemiology of viruses causing chronic hepatitis among populations from the Amazon Basin and related ecosystems. *Cad Saude Publica* 2003; **19:**1583–1591.

Fearnside PM. Deforestation in Brazilian Amazonia: History, Rates, and Consequences. *Conservation Biology* 2005; **19(3):**680-688.

Firth C, Kitchen A, Shapiro B, *et al*. Using time-structured data to estimate evolutionary rates of double-stranded DNA viruses. *Mol Biol Evol* 2010; **27:**2038–2051.

Franco E, Bagnato B, Marino MG, *et al*. Hepatitis B: Epidemiology and prevention in developing countries. *World J Hepatol* 2012; **4(3):**74-80.

Ganem D. Assembly of hepadnaviral virions and subviral particles. *Curr Top Micro-biol Immunol* 1991; **168:**61-83.

Ganem D, Prince AM. Hepatitis B virus infection natural history and clinical consequences. *N Engl J Med* 2004; **350:**1118–1129.

Gaspar AM, Yoshida CF. Geographic distribution of HBsAg subtypes in Brazil. *Mem Inst Oswaldo Cruz* 1987; **82:**253-8.

Godoy BA, Alvarado-Mora MV, Gomes-Gouvêa MS, *et al.* Origin of HBV and its arrival in the Americas--the importance of natural selection on time estimates. *Antivir Ther* 2013; **18(3 Pt B):** 505-12.Ho SYW, Lanfear R, Bromham L, *et al*. Time-dependent rates of molecular evolution. *Mol Ecol* 2011; **20:**3087–3101.

Kato H, Gish RG, Bzowej N, *et al*. Eight genotypes (A–H) of hepatitis B virus infecting patients from San Francisco and their demographic, clinical, and virological characteristics. *J Med Virol* 2004; **73:**516–521.

Kao JH, Hsu HM, Shau WY, *et al*. Universal hepatitis B vaccination and the decreased mortality from fulminant hepatitis in infants in Taiwan. *J Pediatr* 2001; **139:**349–52.

Kiesslich D, Crispim MA, Santos C, *et al*. Influence of hepatitis B virus (HBV) genotype on the clinical course of disease inpatients coinfected with HBV and hepatitis delta virus. *J Infect Dis* 2009; **199(11):**1608–11.

Klingmüller U, Schaller H. Hepadnavirus infection requires interaction between the viral pre-S domain and a specific hepatocellular receptor. *J Virol* 1993; **67:**7414-22.

Kodaman N, Pazos A, Schneider BG, *et al.* Human and Helicobacter pylori coevolution shapes the risk of gastric disease. *PNAS* 2014; **111(4):**1455-1460.

Kurbanov F, Tanaka Y, Mizokami M. Geographical and genetic diversity of the human hepatitis B virus. *Hepatology Researc* 2010; **40:**14–30.

Livingston SE, Simonetti J, Bulkow L, *et al*. Clearance of Hepatitis B e antigen in patients with chronic hepatitis B and genotypes A, B, C, D and F. *Gastroenterology*. 2007; **133:**1452–7

Lucifora J, Arzberger S, Durantel D, *et al*. Hepatitis B virus X protein is essential to initiate and maintain virus replication after infection. *J Hepatol* 2011; **55:**996-1003.

Magiorkinis EN, Magiorkinis GN, Paraskevis DN, *et al*. 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* 2005; **349:**165-171.

Magnius LO, Norder H. Subtypes, genotypes and molecular epidemiology of the hepatitis B virus as reflected by sequence variability of the S-gene. *Intervirology* 1995; **38:**24-34.

Martinello P: A Batalha da Borracha na Segunda Guerra Mundial. Rio Branco, Brazil: EDUFAC; 2004.

Mello FCA, Souto FJD, Nabuco LC, *et al.* Hepatitis B virus genotypes ci rculating in Brazil: molecular characterization of genotype F isolates. *BMC Microbiol* 2007; **7:**103.

Mizokami M, Orito E. Molecular evolution of hepatitis viruses. *Intervirology* 1999; **42:**159-165.

Motta-Castro AR, Martins RM, Araujo NM, et al. Molecular epidemiology of hepatites B vírus in na isolated Afro-Brazilian community. *Arch Virol* 2008; **153:**2197-2205.

Norder H, Hammas B, Lee SD, *et al*. Genetic relatedness of hepatitis B virus stratus with diverse geographic origin and natural variations in the primary structure of hepatitis B surface antigen. *J Gen Virol* 1993; **74:**1341-8.

Norder H, Courouce AM, Magnius LO. Complete genomes, phylogenetic relatedness, and structural proteins of six strains of the hepatitis B virus, four of which represent two new genotypes. *Virology* 1994; **198:**489-503.

Newbold JE, Xin H, Tencza M, *et al*. The covalently closed duplex form of the hepadnavirus genome exists in situ as a heterogeneous population of viral minichromosomes. *J Virol* 1995; **69:**3350–7.

Osiowy C, Giles E, Tanaka Y, *et al*. Molecular evolution of hepatitis B virus over 25 years. *J Virol* 2006; **80:**10307–10314.

Ott JJ, Stevens GA, Groeger J, Wiersma ST. Global epidemiology of hepatitis B virus infection: New estimates of age-specific HBsAg seroprevalence and endemicity. *Vaccine* 2012; **30:**2212–2219

Parana R, Almeida D. HBV epidemiology in Latin America. *J. Clin.Virol* 2005; **34:**S130–S133.

Paraskevis D, Magiorkinis G, Magiorkinis E, *et al*. Dating the origin and dispersal of hepatitis B virus infection in humans and primates. *Hepatology* 2013; **57(3):**908-16.

Pollack JR, Ganem D. Site-specific RNA binding by a hepatitis B virus reverse transcriptase initiates two distinct reactions: RNA packaging and DNA synthesis. *J Virol* 1994; **68:**5579-87.

Pujol FH, Devesa M. Genotypic variability of hepatitis viruses associated with chronic infection and the development of hepatocellular carcinoma. *J Clin Gastroenterol* 2005; **39:**611–618.

Quintero A, Uzcategui N, Loureiro CL, et al. Hepatitis delta virus genotypes I and III circulate associated with hepatitis B virus genotype F in Venezuela. *J Med Virol* 2001; **64:**356–359.

Ribeiro NRC, Campos GS, Angelo ALD, *et al*. Distribution of hepatitis B virus genotypes among patients with chronic infection. *Liver Int* 2006; **26:**636–642.

Roman S, Tanaka Y, Khan A, Kurbanov F, *et al.* Occult hepatitis B in the genotype H-infected Nahuas and Huichol native Mexican population. *J Med Virol* 2010, **82:**1527-1536.

Santos AO, Alvarado-Mora MV, Botelho L, et al. Characterization of hepatites B vírus (HBV) genotypes in patients from Rondônia, Brazil. *Virol J* 2010; **7:**315.

Sánchez-Tapias JM, Costa J, Mas A, *et al*. Influence of hepatitis B virus genotype on the long-term outcome of chronic hepatitis B in western patients. *Gastroenterology* 2002; **123:**1848-56.

Schaefer S. Hepatitis B virus taxonomy and hepatitis B virus genotypes. World J *Gastroenterol* 2007; **13:**14-21

Simmonds P. Reconstructing the origins of human hepatitis viruses. *Philos Trans R Soc Lond B Biol Sci* 2001; **356:**1013-1026.

Sitnik R, Pinho JR, Bertolini DA, *et al.* Hepatitis B vírus genotypes and precore and core mutants in Brazilian patients. *J Clin Microbiol* 2004; **42:**2455-2460.

Soares P, Alshamali F, Pereira JB, *et al*. The Expansion of mtDNA Haplogroup L3 within and out of Africa. *Molecular biology and evolution* 2012; **29:**915-927.

Stuyver L, De Gendt S, Van Geyt C, *et al*. A new genotype of hepatitis B virus complete genome and phylogenetic relatedness. *J Gen Virol* 2000; **81:**67-74.

Tanaka J. Hepatitis B epidemiology in Latin America. *Vaccine* 2000; **18:**S17–S19.

Tanwar S, Dusheiko G. Is there any value to hepatitis B virus genotype analysis? *Curr Gastroenterol Rep* 2012; **14(1):**37-46.

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

Te HS, Jensen DM. Epidemiology of Hepatitis B and C Viruses: A Global Overview. *Clin Liver Dis* 2010; **14:**1–21.

Torres JR. Hepatitis B and hepatitis delta virus infection in South America. *Gut* 1996; **38:**S48–S55.

Torres C, Piñeiro y Leone FG, Pezzano SC, et al. New perspectives on the evolutionary history of hepatitis B virus genotype F. *Mol Phylogenet Evol* 2011; **59:**114-122.

Viana S, Paranfi R, Moreira RC, *et al*. High prevalence of hepatitis B virus in the Western Brazilian Amazon. *Am J Trop Med Hyg* 2005; **73:**808-14.

Yu H, Yuan Q, Ge SX, *et al*. Molecular and phylogenetic analyses suggest an additional hepatitis B virus genotype "I". *PLoS One* 2010; **5:**e9297.

Wei Y, Neuveut C, Tiollais P, *et al*. Molecular biology of the hepatitis B virus and role of the X gene. *Pathol Biol* 2010; **58(4):**267-72.

Zhou Y, Holmes EC. Bayesian estimates of the evolutionary rate and age of hepatitis B virus. *J Mol Evol* 2007; **65:**197-205.

Zoulim F, Saputelli J, Seeger C. Woodchuck hepatitis virus X protein is required for viral infection in vivo. *J Virol* 1994; **68:**2026-30.