UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

VÍRUS DA HEPATITE B E O HOSPEDEIRO HUMANO: INTEGRANDO GENÉTICA E EVOLUÇÃO

BIBIANE ARMILIATO DE GODOY

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

Orientador: Prof. Dr.Nelson Jurandi Rosa Fagundes

Porto Alegre

Outubro de 2018

Este trabalho teve como fonte financiadora o Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). Todas as análises foram desenvolvidas no Laboratório de Genética Humana e Evolução, Departamento de Genética do Instituto de Biociências da Universidade Federal do Rio Grande do Sul (UFRGS), contando com a colaboração do 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; e também da Dra. Tábita Hünemeier, Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, São Paulo, Brasil.

O aluno recebeu bolsa de estudos concedida pelo CNPq, vinculada ao Programa de Pós-Graduação em Genética e Biologia Molecular da UFRGS.

“O mais bonito da ciência é que ela nunca tem fim.”

Francisco Mauro Salzano

AGRADECIMENTOS

De tudo, os maiores aprendizados sempre virão das pessoas com quem tivemos a oportunidade de conviver. É isso que modifica uma vida inteira e é a isso que tenho a maior gratidão. Faço aqui alguns destaques:

Agradeço imensamente ao meu orientador, Nelson Fagundes, pela oportunidade de fazer parte de um projeto tão intrigante, por auxiliar e ser compreensivo diante de todas as dificuldades, por instigar o conhecimento nas mais diversas áreas, por ter participado de cada passo deste trabalho. Além de admirável enquanto professor e pesquisador, é também um exemplo de ser humano que guardarei comigo.

À professora Lavínia Faccini e à Dânae Longo que, ainda no início da minha graduação, abriram as portas da pesquisa científica para mim e carinhosamente me integraram ao Laboratório 113 onde, felizmente, me mantive ao longo de todos estes anos.

Ao querido professor Francisco Salzano, com o qual este projeto se iniciou ainda no meu mestrado, pela bela história de pesquisa que construiu e pelo exemplo de humildade que deixa para todos que tiveram o privilégio de sua convivência.

A todos os colegas do Laboratório 113, por tornarem os dias de trabalho mais alegres, pelo intercâmbio de conhecimentos e pela amizade, e ao querido Elmo Cardoso que de forma cordial sempre esteve disponível para ajudar no que fosse necessário.

Ao Vanderlei, por ter sido o alicerce fundamental para que eu conseguisse finalizar esse trabalho e por me mostrar, constantemente, que o amor reside no apoio, cuidado, carinho e paciência.

À minha irmã Liliane e à minha mãe Elisa, meus maiores exemplos de força e caráter, pelo incentivo incansável, pela cumplicidade e amizade diante dos momentos difíceis, por me

lembrarem sempre que o mais valioso da vida é poder compartilhar as experiências com quem amamos.

Por fim, deixo aqui também um agradecimento à Alexandra Elbakyan, desenvolvedora do repositório online *Sci-Hub*, pelo seu empenho na livre divulgação dos artigos científicos. Diante de uma realidade na qual produzir e acessar ciência são atividades tão dependentes de recursos financeiros, ao mesmo tempo que movimentos anticiência tomam maior espaço na sociedade, são louváveis seus esforços para tornar o conhecimento mais acessível e igualitário.

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

pgRNA: RNA pré-genômico

HCC: carcinoma hepatocelular

NTCP: polipepetídeo co-transportador de taurocolato de sódio

HDV: Vírus de Hepatite D

RESUMO

INTRODUÇÃO: Apesar da existência de uma vacina profilática, o vírus de Hepatite B (HBV) é um importante problema de saúde mundial, sendo responsável por cerca de 257 milhões de infecções crônicas incuráveis ao redor do mundo, que podem resultar em desfechos graves como cirrose e câncer hepático. Considerando a divergência filogenética entre genomas virais completos, o HBV humano tem sido classificado em dez linhagens (genótipos A-J). Em geral, esses genótipos podem ser relacionados a regiões geográficas e perfis étnicos específicos, conectando a dispersão viral à história populacional. Os genótipos de HBV mais divergentes (F e H) são considerados autóctones das Américas, e são usualmente relacionados a populações Nativas Americanas, que mostram as maiores taxas de prevalência para HBV nesse continente. Embora a distribuição global de genótipos do HBV seja sugestiva de uma origem antiga e um longo tempo de coevolução com o hospedeiro humano, não há consenso entre as teorias que abordam a história evolutiva do HBV.

OBJETIVOS: O principal objetivo dessa tese é contribuir para um melhor conhecimento sobre o HBV circulante na América Latina, relacionando aspectos genéticos e evolutivos do vírus e de seu hospedeiro. Primeiramente, avaliamos o impacto de topologias alternativas para o HBV sobre inferências de aceleração de taxa e de seleção positiva para os genótipos F e H. A seguir, testamos a relação evolutiva entre as linhagens do genótipo D presentes na América Latina e na Europa. Finalmente, comparamos populações Nativas Americanas para diferenças genéticas no vírus e no hospedeiro que pudessem estar associadas a diferentes prevalências de HBV.

MATERIAL E MÉTODOS: Usamos abordagens filogenéticas baseadas em máxima verossimilhança e em análise Bayesiana para inferir hipóteses alternativas para o HBV. A evolução molecular dos genótipos F e H foi estimada baseada em cada hipótese filogenética. A seguir, usamos métodos filogenéticos e de genética de populações para comparar a variação genética em linhagens do genótipo D encontradas em populações Latino-Americanas e Europeias para inferir relações ancestral-descendente entre elas. Finalmente, usamos um racional caso-controle para comparar populações Nativas Americanas para a prevalência de HBV usando conjuntos de dados de SNPs genômicos genotipados nessas populações, bem como das linhagens de HBV isoladas das mesmas.

RESULTADOS E CONCLUSÕES: Os resultados mostraram um suporte maior para a topologia de HBV enraizada nos genótipos F-H, ressaltando que o acúmulo de diferenças observadas nessas linhagens é devido a uma divergência antiga, e não a uma aceleração causada por seleção positiva. Por outro lado, a inferência de seleção positiva foi robusta à incerteza filogenética. Os resultados também mostraram a influência de muitas fontes de genótipos D na América Latina, com um papel especial para a Itália, em nível de subgenótipo, na dispersão para o a região sul. Finalmente, nossos resultados são sugestivos de um papel importante do gene *RARB* na susceptibilidade do hospedeiro à infecção por HBV.

ABSTRACT

BACKGROUND: Despite the existence of a prophylactic vaccine, the Hepatitis B Virus (HBV) is a major global health problem, being responsible for about 257 million worldwide incurable chronic infections that can result in severe outcomes like cirrhosis and liver cancer. Considering the phylogenetic divergence among complete viral genomes, the human HBV has been classified into ten HBV lineages (genotypes A-J). In general, these genotypes can be related to specific geographical regions and ethnic profiles, connecting viral dispersal and host population history. The most divergent HBV genotypes (F and H) are considered autochthonous of the Americas and are usually related to Native Americans populations, who show the highest HBV prevalence in this continent. Despite the global HBV genotypes distribution is suggestive of an ancient origin and a long co-evolutionary time with its human host, there is no consensus among the theories addressing evolutionary history of HBV.

AIM: The major aim of this thesis is to contribute to the better knowledge of circulating HBV in Latin America addressing genetic and evolutionary aspects of the virus and its host. First, we evaluated the impact of alternative HBV topologies over inferences of rate acceleration and positive selection for genotypes F and H. Next, we tested the evolutionary relationship between genotype D lineages present in Latin America and Europe. Finally, we compared Native American populations for viral and host genetic differences that could be associated to different HBV prevalence.

METHODS: We used phylogenetic approaches based on maximum likelihood and Bayesian analysis to infer alternative hypothesis for HBV. Molecular evolution of F and H genotypes was estimated based on each phylogenetic hypothesis. Next, we used phylogenetic and population genetic methods to compare the genetic variation in genotype D lineages found in Latina American and European populations to infer ancestral-descendent relationships among them. Finally, we used a case-control rationale to compare Native American populations for HBV prevalence using datasets of genomewide SNPs genotyped in these populations, as well as HBV lineages isolated from them.

RESULTS AND CONCLUSIONS: Our results showed a much higher support for a HBV topology rooted in F-H genotypes, highlighting that the accumulation of differences observed in these lineages is due to an old divergence and not to an acceleration caused by

positive selection. On the other hand, the occurrence of positive selection was robust to phylogenetic uncertainty. In addition, our results showed the influence of many D genotype sources in Latin America, with a special role for Italy in the dispersion at subgenotype level in the Southern region. Finally, our results are suggestive of an important role of *RARB* gene on the host susceptibility to HBV infection.

CAPÍTULO I – INTRODUÇÃO

1.1 A família Hepadnaviridae

O Vírus da Hepatite B (HBV) pertence à família Hepadnaviridae, a qual reúne pequenos vírus de DNA envelopado (~3 a 3.3kpb) que apresentam tropismo por hepatócitos (Ganem and Prince 2004). Esse grupo abrange os únicos vírus de animais que replicam seu genoma via um intermediário de RNA, através da atividade de transcriptase reversa presente na polimerase viral (Summers and Mason 1982; Locarnini et al. 2013). Dentro da família existem dois gêneros espécie-específicos: *Orthohepadnavirus* e *Avihepadnavirus*, que infectam mamíferos e aves, respectivamente (Suh et al. 2013). Com a identificação de elementos virais endógenos – uma consequência da inserção e fixação de fragmentos do DNA viral no genoma hospedeiro -, foram encontradas evidências de hepadnavírus no genoma de crocodilianos, tartarugas, cobras (Gilbert et al. 2014; Suh et al. 2014) e, mais recentemente, também em peixes e anfíbios (Dill et al. 2016), indicando que estes vírus tiveram capacidade de infectar os cinco principais grupos de vertebrados: mamíferos, aves, répteis, anfíbios e peixes.

Em primatas não humanos, HBV tem sido isolado de chimpanzés, gibões, gorilas e orangotangos, os quais mostram linhagens específicas, mas que estão dentro do espectro de variação das linhagens humanas (Simmonds 2001). Em termos filogenéticos, as linhagens humanas são parafiléticas em relação às linhagens dos grandes primatas (e.g. Souza et al. 2018). Embora não exista registro de infecção humana por HBV oriundo de outros primatas, ocorrência de infecção por linhagens humanas e também de outras espécies de primatas são observadas nesses animais (Rasche et al. 2016). Considerando os macacos do Novo Mundo, até o momento HBV só foi identificado no macaco-barrigudo (*Lagothrix lagotricha*) (Lanford et al. 1998) e no macaco-prego-do-peito-amarelo (*Sapajus xanthosternos*) (Souza et al. 2018). As linhagens virais obtidas dessas espécies constituem, do ponto de vista filogenético, um grupo-irmão em relação às demais, com diferenças genéticas mais pronunciadas dessas últimas (Souza et al. 2018).

Uma análise recente comparando diferentes famílias virais e seus hospedeiros mostrou que Hepadnaviridae apresenta eventos frequentes de co-divergência, o que é refletido em padrões de história evolutiva semelhantes. Em conjunto, essas características sugerem uma origem viral antiga, envolvendo eventos de alternância de hospedeiros e um longo período de co-evolução (Dill et al. 2016; Geoghegan et al. 2017; Lauber et al. 2017).

Apesar da ocorrência em hospedeiros tão diferentes, o HBV compartilha uma organização genômica e estrutural que é semelhante para todos os membros da família Hepadnaviridae.

1.2 O Vírus da Hepatite B em humanos

1.2.1 Estrutura

A partícula infecciosa do HBV apresenta uma estrutura esférica e envelopada que abrange o nucleocapsídeo viral, dentro do qual se encontra o DNA (Ganem and Prince 2004). O genoma do HBV possui estrutura relaxada, circular (rcDNA) e parcialmente dupla-fita, com diferentes comprimentos: a fita maior (negativa) possui cerca de 3200 pares de base enquanto a sua complementar (positiva) apresenta um tamanho que varia de 50 a 100% da primeira (Wei et al. 2010). Sua organização genômica é compactada e inteiramente codificante, apresentando 4 fases abertas de leitura (ORFs) parcialmente sobrepostas (Ganem and Prince 2004):

- P - codifica a polimerase viral, uma proteína multifuncional que apresenta adicional atividade de transcriptase reversa (Dandri and Locarnini 2012);

- S (ou Pré-S1/ Pré-S/ S) - inteiramente sobreposta à P e dividida em três domínios, os quais levam à produção das proteínas de superfície que formam o envelope do nucleocapsídeo. Todas elas possuem a mesma terminação 3', mas apresentam diferentes códons de iniciação, o que configura a produção de três proteínas de tamanhos diferentes: a proteína grande (LHBs) é essencial para a entrada do vírus na célula, apresentando função importante relacionada à capacidade infectante; a média (MHBs) não possui funções estabelecidas; e a pequena (SHBs) é a mais abundante delas, constitui o antígeno de superfície (HBsAg) e é sintetizada no retículo endoplasmático (Valaydon and Locarnini 2017). Essas mesmas proteínas são também utilizadas pelo Vírus de Hepatite D (HDV), uma vez que este não produz seu próprio antígeno de superfície, tornando necessária a existência de infecção prévia por HBV para que a infecção ocorra (Zhang et al. 2018).

- C (ou Pré-C/ C) - quando traduzido completamente, codifica o antígeno secretado "e" (HBeAg), uma proteína acessória que funciona como marcador sorológico de infecção ativa e tem sido apontada com necessária no processo de persistência da infecção (Milich and Liang 2003). O domínio pré-C produz o antígeno core (HBcAg) que compõe o nucleocapsídeo que abriga o DNA viral (Dandri and Locarnini 2012).

- X - codifica a proteína regulatória HBx que ocorre exclusivamente nos mamíferos (van Hemert et al. 2011), e cuja função está relacionada com a expressão gênica vírus-hospedeiro, sendo requerida para o estabelecimento da infecção e manutenção da replicação ativa do vírus. Sua atividade tem sido associada como o fator principal no potencial carcinogênico do HBV (Wei et al. 2010; Lucifora et al. 2011).

1.2.2 Transmissão e Epidemiologia do HBV

A situação epidemiológica do HBV varia bastante ao redor do mundo. As maiores prevalências são observadas na região do Oeste do Pacífico e também na África, onde mais de 6% da população se encontra infectada pelo vírus, seguido pela região do Mediterrâneo que apresenta prevalências gerais maiores que 3% (WHO 2017). HBV é transmitido pelo contato com sangue ou outros fluidos corporais de pessoas infectadas (Trépo et al. 2014). O vírus pode ser passado de forma vertical, da mãe para o filho, e também de forma horizontal pelas vias parenteral e sexual (Lin et al. 1990; Li et al. 2015). As principais vias de transmissão variam de acordo com a situação epidemiológica da região, bem como com os hábitos e condições das populações (Zanetti et al. 2008; Alvarado-Mora et al. 2011; Godoy et al. 2016). De uma forma geral, regiões de altas prevalências apresentam taxas maiores de infecção perinatal ou transmissão vertical, enquanto locais com prevalência de baixa a intermediária apresentam rotas de transmissão predominantemente parenteral e sexual (Stevens et al. 1975). Ainda assim, situações diferenciadas podem ser observadas mesmo em áreas de baixa prevalência, como é o caso de grupos mais vulneráveis, imigrantes, e usuários de drogas, por exemplo (Montenegro and Stephens 2006; Alvarado-Mora et al. 2011; McCarthy et al. 2013; Richter et al. 2014; Godoy et al. 2016; Coppola et al. 2017).

Na América as prevalências gerais de HBV são baixas, entretanto a América Latina conta com a maioria dos casos da infecção e apresenta situações epidemiológicas preocupantes em algumas regiões (PAHO 2016). Um exemplo disso é a região da Amazônia e as populações Nativas Americanas. A região é endêmica para HBV e HDV, e as condições de vida das populações, que podem envolver localização isolada e condições precárias de saneamento, além de práticas culturais específicas como tatuagem, escarificação, compartilhamento de objetos ou perfuração para colocação de adornos, por exemplo, acabam contribuindo para um risco aumentado de transmissão (Gomes-Gouvea et al. 2009; WHO

2017). Com isso, estudos mostram que as prevalências em determinadas populações podem alcançar os 80%, revelando uma situação bastante grave (Torres 1996; Braga 2004; Godoy et al. 2016).

Como consequência mais frequente da infecção em adultos, há ocorrência de hepatite aguda, onde a imunidade do indivíduo realiza a depuração viral de forma espontânea e assintomática. Entretanto, em alguns casos esse vírus pode persistir na forma de uma infecção crônica que não possui cura e pode resultar em desfechos graves como a cirrose, carcinoma hepatocelular (HCC) ou falência hepática (Ganem and Prince 2004). Essa persistência está bastante relacionada com a idade no período da infecção. Estima-se que mais de 90% das crianças infectadas antes do primeiro ano de vida desenvolvam hepatite crônica, enquanto em adultos esse número varia entre 0 e 2% (Ozasa et al. 2006; Yuen et al. 2018).

A disponibilidade de uma vacina efetiva como estratégia profilática produziu uma redução significativa nas taxas globais de infecção, mortalidade de crianças em decorrência da doença, taxas de portadores crônicos e incidência de HCC na infância (Zanetti et al. 2008; Schweitzer et al. 2015). Entretanto, existem cerca de 257 milhões de pessoas cronicamente infectadas por HBV atualmente, o que tem produzido mais de um milhão de mortes anuais em decorrência de patologias hepáticas associadas à infecção (Block et al. 2007; WHO 2017). Esses números motivaram a Organização Mundial da Saúde a adicionar as hepatites virais junto às principais doenças infecciosas que representam uma ameaça à saúde global e precisam de esforços conjuntos para serem combatidas (Chen et al. 2015; Locarnini et al. 2016; Revill et al. 2016).

1.2.3 Replicação do HBV

O processo celular completo pelo qual a infecção por HBV ocorre está representado na Figura 1 (Revill and Locarnini 2016). A entrada do HBV no hepatócito é mediada por uma ligação reversível e de baixa afinidade (Leistner et al. 2008) seguida por uma ligação de alta especificidade entre a região Pré-S1 da proteína maior do envelope viral ao polipepetídeo co-transportador de taurocolato de sódio (NTCP) – um receptor específico com fundamental atividade para a entrada tanto do HBV como do HDV no hepatócito. Estudos mostram que a expressão deste receptor em outros animais produz suscetibilidade à

infecção, sendo indicativo de que este é um fator limitante para ocorrência da infecção em humanos (Ni et al. 2014).

Após a ligação a este receptor, o vírus é internalizado pelo hepatócito através de endocitose (Yan and Li 2015). Na sequência, ocorre a remoção do envelope viral e do nucleocapsídeo e o rcDNA é transportado para o núcleo onde é convertido através da interação com proteínas do hospedeiro a uma molécula circular covalentemente fechada (cccDNA) - uma estrutura funcional de minicromossomo que servirá de molde para a transcrição de todos os RNAs virais (Rabe et al. 2003; Dandri and Locarnini 2012). Nesse processo, os RNAs menores irão originar todas as proteínas virais, enquanto o RNA maior denominado de pré-genômico (pgRNA) será transportado para o citoplasma, onde receberá a estrutura do nucleocapsídeo e servirá de molde para a replicação do DNA viral através da atividade de transcriptase reversa da polimerase. Essa conversão do RNA longo em DNA ocorre de maneira sequencial: a primeira fita (-) é sintetizada a partir do RNA encapsidado, enquanto a segunda (+) utilizará a fita recém-sintetizada como molde (Ganem and Prince 2004).

Nesse processo pode ocorrer pareamento inespecífico e a consequente produção de estruturas lineares de DNA dupla-fita, as quais podem se integrar ao DNA do hospedeiro através de recombinação (Beck and Nassal 2007; Valaydon and Locarnini 2017). De uma forma geral, eventos de integração ao genoma do hospedeiro são pouco frequentes em infecção por HBV, mas costumam aumentar com a persistência da infecção e são associados à ocorrência de carcinoma hepatocelular (HCC) (Zhao et al. 2016). Devido à organização estrutural complexa do vírus, neste processo normalmente ocorre apenas a preservação do domínio menor da ORF S (Yuen et al. 2018). Na maior parte das vezes, no entanto, a estrutura circular é formada normalmente e o vírion maduro recebe o envelope proteico, sendo liberado do hepatócito com capacidade para infectar novas células (Ganem and Prince 2004; Wei et al. 2010; Dandri and Locarnini 2012; Valaydon and Locarnini 2017). Por outro lado, parte dessas estruturas em formação (com DNA linear ou circular) será reencaminhado para o núcleo, mantendo um reservatório de moldes transcricionais que servirão para reativar a replicação viral, o que representa um importante desafio clínico para a cura da doença (Petersen et al. 2016; Valaydon and Locarnini 2017).

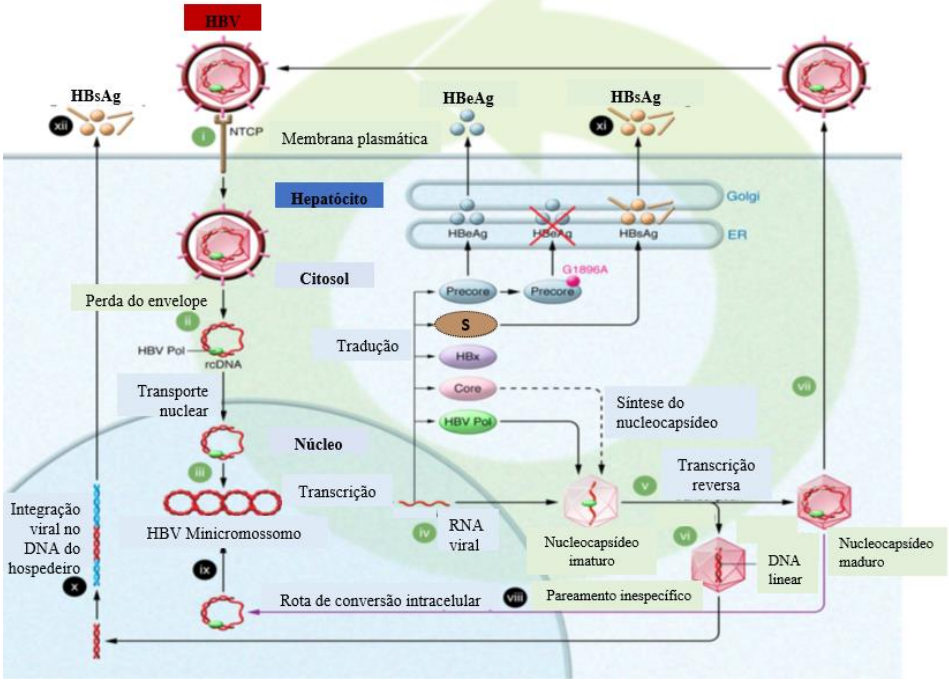


Figura 1: Representação da replicação do HBV (modificado de Revill e Locarnini, 2016).

Muitas das proteínas produzidas durante este processo são utilizadas como marcadores sorológicos da doença e são representativas do tipo de resposta e progressão que está em curso no indivíduo infectado, servindo como um guia para aplicações terapêuticas. O HBsAg é secretado no sangue em grandes quantidades como característica da replicação ativa do HBV, mas sua produção é vinculada tanto à atividade viral como também ao DNA do hospedeiro pós integração e, este último, pode implicar na dificuldade terapêutica de alcançar a perda da expressão de HBsAg, contribuindo para a persistência da infecção (Yuen et al. 2018).

O HBV é um vírus de longa história evolutiva junto ao seu hospedeiro humano (Paraskevis et al. 2013b), na maioria dos casos, a infecção viral e a subsequente replicação dentro do hepatócito não produzem efeito citopático. Entretanto, as consequências do processo de infecção por HBV vão depender da interação entre fatores virais e do hospedeiro (Chisari and Ferrari 1995; Schinzari et al. 2015; Gómez-Moreno and Garaigorta 2017). Enquanto o dano hepático associado à progressão da doença ocorre principalmente por efeito das tentativas da resposta imune do hospedeiro de eliminar o HBV dos hepatócitos infectados, como resposta, acabam ocorrendo diversas mutações no HBV, o que gera uma

seleção de mutações de escape à resposta imunológica (Locarnini 2005). Com isso, os desfechos clínicos da doença são altamente variáveis.

1.3 Classificação Genotípica

A falta de atividade de correção de erro da polimerase do HBV torna o vírus altamente suscetível a adquirir mutações durante cada ciclo de replicação (Steinhauer and Holland 1986). Essa alta taxa mutacional contribui para a formação de variantes virais. A classificação dessas variantes é realizada através de análises filogenéticas do genoma completo do vírus de modo que diferenças de 7.5% ou mais entre os genomas comparados configuram um genótipo de HBV, enquanto variações de 4% ou mais em um mesmo genótipo passam classificá-lo em diferentes subgenótipos (Kramvis 2014). Atualmente existem 10 genótipos descritos para HBV os quais são denominados de A-J (Huy et al. 2008; Tatematsu et al. 2009; Kramvis 2014). Os genótipos de HBV apresentam uma forte estruturação geográfica e, embora o padrão de distribuição esteja mudando especialmente em regiões com elevadas taxas migratórias, suas distribuições globais estão assim caracterizadas (Kurbanov et al. 2010; Sunbul 2014): o genótipo A está presente no mundo todo, mas é primariamente descrito na África Subsaariana, Norte da Europa e América do Norte; genótipos B e C são mais frequentemente encontrados no leste e sudeste asiático e na Oceania; o genótipo D também apresenta uma distribuição global e é especialmente característico da Europa, Oriente Médio e países da região do Mediterrâneo; o genótipo E é endêmico da África Ocidental e raramente registrado fora do continente africano; genótipos F e H representam as linhagens autóctones da América, com ocorrência bastante restrita às Américas do Sul e Central; o genótipo G não foi associado a uma região específica e tem sido detectado em amostras dos Estados Unidos, México, França, Alemanha, Turquia, Brasil e Japão; os mais recentemente descritos genótipos I e J foram detectados apenas em indivíduos do Vietnam e Laos, e Japão, respectivamente.

Essa associação entre diferentes genótipos de HBV e regiões ou etnias específicas fazem deste vírus um bom marcador dos processos históricos vinculados às populações, uma vez que reflete, especialmente, os padrões migratórios que compuseram a população do local. Exemplos disso podem ser vistos na distribuição dos genótipos de HBV na América Latina, como discutido a seguir.

A América Latina apresenta uma distribuição diferenciada das demais regiões, com ocorrência do genótipo H especialmente no México e predomínio do genótipo F na maior parte dos outros países de língua espanhola (Alvarado-Mora and Pinho 2013). Os genótipos F e H são considerados autóctones das Américas e tem sido relacionado às populações Nativas Americanas (Devesa and Pujol 2007; Roman et al. 2010). Em contraste a esta distribuição (Lampe et al. 2017)(Fig. 2), no Brasil o genótipo mais frequente (A, subgenótipo A1) representa uma linhagem associada à origem africana, o que tem sido relacionado à massiva imigração forçada de escravos africanos para este país entre os séculos XVI e XIX (Araujo et al. 2004; Moura et al. 2013). Por fim, na região Sul do Brasil ocorre um marcado predomínio do genótipo D, característico da região do Mediterrâneo e que tem sido associado à ancestralidade europeia. O sul do Brasil recebeu diversas ondas de colonizadores europeus durante a segunda metade do século XIX, o que pode explicar a alta frequência desse genótipo nessa região (Bertolini et al. 2012; Moura et al. 2013; Gusatti et al. 2015).

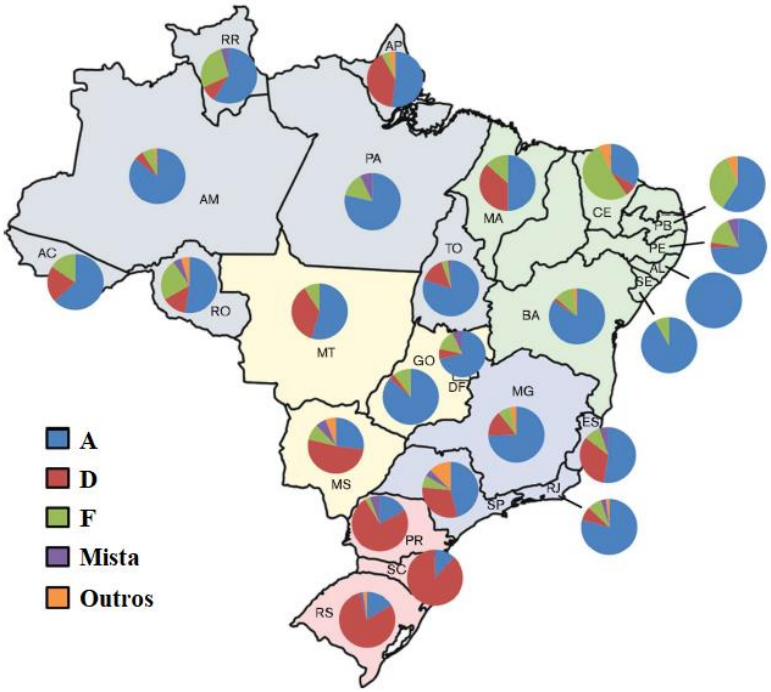


Figura 2: Distribuição dos genótipos de HBV no Brasil: enquanto o genótipo A (azul) – relacionado à origem africana – predomina em quase todos os estados, na região Sul ocorre maior representatividade do genótipo D (vermelho) – relacionado com origem europeia. Em comparação, o genótipo F (verde) - considerado autóctone da América do Sul e relacionado às populações Nativas Americanas - apresenta menor frequência (Lampe et al. 2017).

1.4 História Evolutiva

Todas estas características associadas ao HBV reforçam a ideia de uma origem viral antiga e um longo período de segregação com o hospedeiro humano (Paraskevis et al. 2013). Recentemente, HBV foi isolado de amostras antigas datadas de ~800 a 4500 anos atrás (Mühlemann et al. 2018), e também em múmias do século XVI (Bar-Gal et al. 2012; Ross et al. 2018). Todas as linhagens se mostraram proximamente relacionadas às que ocorrem atualmente, evidenciando uma divergência genotípica bastante antiga. Embora essa idade antiga associada ao vírus pareça plausível, não existe um consenso científico sobre a sua história de origem e dispersão. Dentre as hipóteses sugeridas está a que propõe que o 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 et al. 2009); ou ainda 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 and Norder 1995; Mühlemann et al. 2018).

Um dos motivos mais intrigantes em relação à essa história evolutiva reside na marcada divergência que os genótipos americanos F e H têm dos demais. Esse fato tem levantado hipóteses controversas, como a que sugere uma origem do HBV no Novo Mundo e posterior disseminação para os demais continentes há apenas 400 anos (Simmonds 2001). Adicionalmente, os processos que levaram ao acúmulo de tantas diferenças no ramo filogenético destas linhagens também não são completamente entendidos, e alguns estudos têm proposto a ocorrência de uma aceleração na taxa evolutiva desses genótipos influenciada por seleção positiva (Paraskevis et al. 2015). Desta forma, investigar os processos evolutivos envolvidos com as linhagens de HBV representa um desafio que pode auxiliar não apenas no melhor entendimento dos processos históricos das populações, mas também do patógeno em questão.

1.5 Objetivos

1.5.1 Objetivo Geral

O objetivo desta tese é contribuir para um conhecimento mais aprofundado sobre o HBV circulante na América Latina, relacionando aspectos genéticos e evolutivos de ambos, vírus e hospedeiro.

1.5.2 Objetivos Específicos

a) Testar se a inferência de uma taxa evolutiva acelerada nos genótipos F e H é robusta em relação à diversidade de hipóteses filogenéticas plausíveis para o HBV, e elucidar o papel da seleção positiva sobre essa possível aceleração;

b) Avaliar a relação entre as linhagens de HBV do genótipo D circulantes na América Latina e na Europa, relacionando com aspectos históricos da imigração europeia na região bem como com os padrões de dispersão do HBV;

c) Investigar fatores genéticos do vírus e suas populações hospedeiras relacionados à suscetibilidade de infecção ao HBV em uma amostra de Nativos Americanos.

CAPÍTULO II

"HEPATITIS B VIRUS: ALTERNATIVE PHYLOGENETIC HYPOTHESES AND ITS IMPACT ON MOLECULAR EVOLUTION INFERENCES"

Manuscrito submetido à revista *Molecular Phylogenetics and Evolution.*

Hepatitis B Virus: alternative phylogenetic hypotheses and its impact on molecular evolution inferences

Bibiane A. Godoy[1], Nelson J. R. Fagundes[1,2]*

[1]Postgraduation Program in Genetics and Molecular Biology, Federal University of Rio Grande do Sul, Porto Alegre, RS

[2]Department of Genetics, Institute of Biosciences, Federal University of Rio Grande do Sul, Porto Alegre, RS, Brazil

*Corresponding author.

E-mail address: nelson.fagundes@ufrgs.br (N. J. R. Fagundes)

Declarations of interest: none.

Abstract

Characterizing molecular evolution patterns of the Hepatitis B Virus (HBV) is important for a better understanding of the natural history of this infection. However, several molecular evolution estimates are conditional on a given tree topology. There is no consensus about the phylogenetic relationships of HBV genotypes, and different studies often find alternative topologies. While most studies consider HBV genotypes F and H as sister to all other human genotypes, a recent study suggested an alternative HBV phylogeny that indicates an accelerated substitution rate for HBV-F/H partially driven by positive selection. In this study, we evaluate the impact of alternative HBV topologies on inferences of HBV phylogeny, rate acceleration, and positive selection on the HBV-F/H branch. Our results indicate that under certain methodological assumptions alternative HBV topologies are equally likely, even though some topologies may be preferred under certain methods. Considering phylogenetic uncertainty, there is no evidence that HBV-F/H had an accelerated substitution rate, even though inferences of positive selection are robust to alternative background topologies. Understanding the full range of likely topologies will be crucial for elaborating, testing, and refining hypothesis about the evolutionary HBV origins in our species.

Highlights

- Alternative methods result in result in different tree topologies for HBV
- These alternative topologies are equally likely under a maximum likelihood approach
- Genotypes F/H are sister to all other genotypes under the Bayesian criterion
- Evidence for HBV-F/H rate acceleration was conditional on the topology assumed
- Evidence for codons under positive selection were robust to the topology assumed

Key words

Hepatitis B Virus, Phylogenetic analysis, Molecular clock, Maximum Likelihood, Bayesian inference

1. Introduction

Hepatitis B Virus (HBV) is an infectious agent that affects hundreds of millions of people worldwide, reaching high prevalence in parts of Southeast Asia and Africa (WHO, 2017). Even though he overall rates of infection seems to have declined due to vaccination (Schweitzer et al., 2015) there are still 257 million HBV chronic carriers who may develop severe outcomes such as cirrhosis and hepatocellular carcinoma, causing a high number of annual deaths (WHO, 2017).

HBV belongs to the Hepadnaviridae family – a group of enveloped viruses characterized by marked hepatocytes tropism and a common structural organization. These viruses share a common genomic organization that comprises a partially double-stranded circular DNA with ~3,200bp. The human HBV has a compact genome organized with four partially overlapping open reading frames (ORFs) which encode the typical core (C), polymerase (P) and surface (S) proteins, and, in viruses infecting mammals, the X protein (Locarnini et al., 2013; Dill et al., 2016). This genomic structure, in which most of the genome is encoded in overlapping ORF, result in varying substitution rates between overlapping and non-overlapping genome regions in consequence of different selective constraints over different parts of genome (Simmonds, 2001).

Hepadnaviruses are classified into two genera: *Avihepadnavirus*, infecting birds, and *Orthohepadnavirus*, infecting mammals. However, the recent isolation of Hepadnaviridae virus among fishes and amphibians indicates an ability to infect all vertebrate classes (Locarnini et al., 2013; Hahn et al., 2015; Dill et al., 2016). Among apes, HBV has been isolated from chimpanzees, gibbons, gorillas and orangutans, whose lineages form clades nested within the human phylogenetic diversity of HBV genotypes (e.g. Mühlemann et al.,

2018). HBV has also been isolated from New World monkeys: the woolly (*Lagothrix lagotricha*) and capuchin (*Sapajus xanthosternos*) monkeys (Lanford et al., 1998; Souza et al., 2018, respectively). These lineages are sister to the "human/ape clade" (Souza et al., 2018). Such wide range of hosts suggest an ancient evolutionary history in primates involving ancient origin, host switching events, and virus/host coevolution (Dill et al., 2016; Geoghegan et al., 2017, Souza et al., 2018).

Based on complete viral genome sequencing, ten HBV genotypes (A to J) have been identified in modern human populations, showing a strong geographical structuring that may reflect specific ethnic profiles and human migration patterns (Tatematsu et al., 2009; Paraskevis et al., 2013). Ancient human samples suggest at least one more genotype, which is likely extinct in current human populations (Mühlemann et al., 2018). Although the geographical distribution of HBV genotypes and molecular clock estimates are suggestive of an ancient origin and a longtime coevolution with human populations, there is no consensus among the hypotheses addressing the evolutionary history of HBV (Locarnini et al., 2013; Paraskevis et al., 2013; Paraskevis et al., 2015; Mühlemann et al., 2018; Souza et al., 2018). HBV genotypes F and H (HBV-F/H) are considered autochthonous of the Americas (e.g. Locarnini et al., 2013). These genotypes are closely related and very divergent from other human genotypes. In most analyses, especially those assuming the molecular clock, this "divergence" result in HBV-F/H being inferred as sister to all other human genotypes (e.g. Godoy et al., 2013; Paraskevis et al., 2013; Mühlemann et al., 2018; Souza et al., 2018).

However, a phylogenetic study without assuming the molecular clock raised an alternative hypothesis (Paraskevis et al., 2015). In this study, the most likely topology indicates HBV-B as sister to other human genotypes, and places HBV-F/H nested within the

remaining human HBV variation, with its long branch reflecting an accelerated substitution rate driven (at least in part) by positive selection. However, because several molecular evolution estimates are conditional on a given tree topology, it is important to elucidate the robust of such estimates when alternative phylogenetic hypotheses exist. In this study, we reanalyze the dataset gathered by Paraskevis et al. (2015) to evaluate the impact of two alternative sets of methodological assumptions on a) estimating the best-supported HBV phylogeny; b) explicitly comparing alternative HBV topologies; c) testing for HBV-F/H rate acceleration; and d) testing for positive selection on HBV-F/H branch.

2. Material and Methods

2.1.1 Dataset and partitions

We used the same dataset analyzed by Paraskevis et al. (2015). Sequences were retrieved from the GenBank and included 105 complete genome sequences from all major human HBV genotypes and isolates from five non-human primate species (chimpanzee, gibbon, gorilla, orangutan and the woolly monkey). We excluded three sequences from the original dataset (AB493845, AY902768 and FJ692556) due to the occurrence of frameshift mutations.

To account for possible variation in substitution rates across sites, the alignment was partitioned into overlapping and non-overlapping genome regions, as in Paraskevis et al. (2015). We used GenBank entry X02496 as a reference for determining the precise boundaries among partitions. Overlapping regions included positions 1-837, 1376-1625, 1816-1840, 2309-2454, and 2850-3182. Non-overlapping regions included positions 1626-

1815 (X gene), 1841-2308 (C gene), 838-1375 and 2455-2849 (P gene; P1 and P2 regions, respectively).

2.2 Phylogenetic analyses and comparison of HBV topologies

We restricted our comparison to two of the most frequently used methods for estimating HBV phylogenies. First, we used an analysis based on Maximum Likelihood (ML) in the RAxML v.8.2.10 package (Stamatakis, 2014). We assumed no molecular clock, GTR + gamma as the evolutionary model, 10 independent runs, and 1,000 bootstrap replicates to assess clade support. This was the method used by Paraskevis et al. (2015). Alternatively, we used a Bayesian inference (BI) method implemented in Beast v.1.8.4 (Drummond and Rambaut, 2007), assuming a relaxed molecular clock (Drummond et al., 2006) using the rates reported by Paraskevis et al. (2015). We used GTR + gamma as the evolutionary model and the Yule tree-prior. We ran the Markov Chain Monte Carlo (MCMC) for 10 million states, sampling every 1,000 state, discarding the first 1 million states as burn-in. Both analyses included the Woolly Monkey HBV sample (AF046996), which was used to root the ML tree. In the case of the BI analysis, since the use of a molecular clock results in an ultrametric tree, there is no need to define a priori an outgroup taxon to root the phylogeny.

The best trees under the ML and BI criteria (ML-tree, BI-tree, respectively), were considered as alternative topological hypotheses for HBV history, and were compared under ML and BI approaches. Under the ML approach, we enforced a series of topological restrictions at the clade level (e.g. HBV-F/H assumed to be monophyletic without imposing any restrictions on its internal relationships) to reproduce the BI-tree. We used RAxML to with the same assumptions as before to estimate the ML score of BI-tree, and to compare

the topologies using the Shimodaira-Hasegawa test in Consel (Shimodaira & Hasegawa, 2001). Under the BI approach, we also enforced a series of topological restrictions at the clade level to generate three alternative hypotheses: a) BI-tree, b) ML-tree, and c) ML(relax)-tree, in which we allowed the internal relationship among clades to vary while maintaining the alternative ML-tree root. This third hypothesis was included to guarantee that BI-tree was tested against the best possible topology (under the BI approach) having the same root as ML-tree, as we ultimately wanted to compare alternative rooting hypotheses. Alternative hypotheses were estimated in Beast v.1.8.4 using the same assumptions as mentioned above and were compared using the Bayes Factors, which are ratios of marginal likelihoods (Kass & Raftery, 1995). Bayes Factors are usually expressed as pairwise differences between the log-transformed marginal likelihood of Model 1 vs. Model 0 ($lnBF_{10}$ = $lnLkModel_1$ – $lnLkModel_0$). $2lnBF_{10}$ values >10 are considered decisive support in favor of Model 1 (Kass & Raftery, 1995). Marginal likelihoods were estimated in Beast v.1.8.4 based on the stepping stone estimator (Baele et al., 2012) using 100 sampling steps in which each step is estimated from a MCMC of 1 million states. To test the effect of including the wooly monkey sequence in the analysis, we repeated all Bayesian analyses excluding this entry from the dataset.

2.3 Estimates of substitution rates and test of positive selection

Estimates of the evolutionary rate of HBV-F/H branch were performed in the MCMCTree module of the PAML v.4.7 package (Yang, 2007) using both ML-tree and BI-tree as alternative background topologies. This evolutionary model assumes independent rate variation among branches, allowing the identification of lineages showing evidence for increased (accelerated) evolutionary rate. Following Paraskevis et al. (2015), which presents the full reasoning for each of these events, we placed time calibrations in six notes along the

tree: HBV root (13,000 – 200,000 years (y)); $A5_{Haiti}$ (<500y); B6 (1,000 – 6,000y); C3 (5,100 – 12,000y); D4 (5,100 – 12,000y); F/H (13,000 – 20,000y). We used uniform priors for all calibrations except $A5_{Haiti}$, for which we used an upper bound calibration. All rate estimates are given in substitutions per site per year (s/s/y). For both hypotheses, we ran the MCMC for 1.5 million states, sampling every 100 states, and discarding the first 500,000 states as burnin.

Finally, the impact of positive selection under both ML-tree and BI-tree were estimated using the branch-site test available in the CODEML module of the PAML v.4.7 package (Yang, 2007). This analysis compares the rate of non-synonymous to synonymous substitutions to test for positive selection in specific (foreground) branches of the phylogeny compared to the remaining "background" branches. Codons under positive selection were inferred based on Bayes empirical Bayes estimates (Yang et al., 2005). For both hypotheses, we used the path leading to HBV-F/H as the foreground branch.

3. Results and Discussion

The topologies obtained under ML and BI approaches had marked differences regarding the relationship among HBV genotypes, especially regarding the most basal splits (Fig. 1). The best topology under the ML approach (ML-tree) placed HBV-B as sister to all other genotypes (Fig. 1a), while the best topology under the BI approach (BI-tree) placed HBV-F/H as sister to all other genotypes (Fig. 1b). Under the ML criterion, there was no statistical difference between ML-tree and BI-tree hypotheses (Shimodaira-Hasegawa test, P=0.142). On the other hand, under the BI criterion BI-tree received decisive support versus

both ML-tree ($2lnBF_{10}$ = 43.01) and ML(relax)-tree ($2lnBF_{10}$ = 21.17). These results were robust to the exclusion of the wooly monkey sequence from the dataset (Table 1).

Recent advances on molecular clock methods, including the introduction of relaxed clock models, have allowed robust phylogenetic inference in face of rate variation (Kumar & Hedges, 2016; Reis et al., 2016). In addition, other studies based on alternative sets of assumptions under ML criteria have recovered HBV-F/H as sister to all other human/ape genotypes, similarly to the BI-tree (e.g. Souza et al., 2018). These arguments seem to advocate in favor of a basal position for HBV-F/H. However, there are other topological disagreements among HBV phylogenies, even when they agree on HBV-F/H position (see Souza et al., 2018 vs. Mühlemann et al., 2018 for a recent example). This issue will be firmly settled only after a thorough comparison of phylogenetic methods including alternative molecular-clock methods and search strategies in both ML and BI frameworks.

Inferences of HBV-F/H rate acceleration were very sensitive to the background phylogeny used to guide the estimate. Under ML-tree there was a clear indication that the branch leading to HBV-F/H accelerated, having a mean substitution rate of 0.079s/s/y versus 0.030s/s/y for the other branches (Fig. 2a). Indeed, under this topology, the branch leading to HBV-F/H is the fastest-evolving branch in HBV history, in agreement with Paraskevis et al. (2015). However, when BI-tree is considered, the branch leading to HBV-F/H shows an ordinary rate of 0.040s/s/y, close to the average of 0.038s/s/y for the remaining branches (Fig. 2b), and slower than the fastest evolving branch (0.063s/s/y for a tip branch within HBV-B1). If a topology similar to BI-tree is more likely (see above), the conclusion of a faster substitution rate for his lineage doen not hold.

On the other hand, estimates for positive selection in HBV-F/H lineage were robust to the background phylogenetic hypothesis (Table 2). Evidence for codons under positive

selection were found in genes P (codons 69 and 75 in the P2 region), and X (codon 18). The P gene encodes the viral polymerase, which have an important role in DNA synthesis and RNA encapsidation, while X gene encodes the X-protein that is related to virus replication, spread and pathogenicity (Dandri and Locarnini, 2012). Li et al. (2017) have also suggested an important role for positive natural selection during genotype divergence. These authors developed an alternative method that allowed them to take into account the full HBV genome (including positions in overlapping ORFs), and inferred that positive selection affected genes P, X, and, mostly, S (affecting Pre-S1 protein). These results seem to corroborate that positive selection in HBV genome does not require an accelerated substitution rate.

How does the uncertainty over the exact HBV phylogeny impact hypothesis for HBV origins? Paraskevis et al. (2013; 2015) Souza et al. (2018), and Mühlemann et al. (2018) seem to agree on a general evolutionary scenario for HBV even though their phylogenetic reconstruction is not identical. Indeed, irrespective of topological details, recognition that HBV has a relatively slow evolutionary rate seems to be only compatible with a long coevolution history between HBV and humans. A HBV origin in the Pleistocene, together with the fact that the African HBV-E genotype is relatively recent among virtually all alternative topologies, inevitably leads to a Eurasian origin of HBV. We agree with this view. However, as highlighted by Mühlemann et al. (2018), the recent recognition of ancient, extinct, HBV genotypes in humans raises the possibility that we underestimate the phylogenetic depth of human HBV lineages.

If BI-tree (or any topology in which HBV-F/H is sister to all other genotypes) represents the most likely HBV phylogenetic hypothesis, this mean that the most ancient split in HBV history is between New World and Old World genotypes. It is this node rather than the divergence between HBV-F and HBV-H (Paraskevis et al., 2015) that may be

associated with the Peopling of the Americas ~16,000y ago (Raghavan et al., 2005). This is fully compatible with the date of ~15,600y (13,700 – 17,800y) for the human HBV estimated by Mühlemann et al. (2018), and may be seen as further evidence of a short (<8,000y) Beringian standstill (Raghavan et al., 2015) that is also the most likely explanation for the high genetic divergence shown by HBV-F/H genotypes. Souza et al. (2018) have raised a similar point, even though we do not see any need to evoke recent Polynesian migrations to account for HBV-F/H variation in the Americas.

In summary, our study demonstrates the impact of phylogenetic background hypotheses on further molecular evolution features of HBV. HBV-F/H most likely represents the sister lineage to all other human/ape HBV genotypes. Taking into account phylogenetic uncertainty, there is no strong evidence that HBV-F/H has an accelerated substitution rate, even though inferences of positive selection were robust to alternative background topologies. Understanding the full range of likely topologies will be crucial for elaborating, testing, and refining hypothesis about the evolutionary HBV origins in our species.

Acknowledgements

We thank João R. R. Pinho, Michele S. Gomes-Gouvêa, Felipe G. Grazziotin and Jomar P. Laurino for early discussions on HBV phylogeny. We also thank André L. Zani for comments on an early draft of the manuscript. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), and by the Postgraduation Program in Genetics and Molecular Biology (PPGBM-UFRGS).

References

Baele, G., Lemey, P., Bedford, T., Rambaut, A., Suchard, M. A., Alekseyenko, A. V., 2012. Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. Mol. Biol. Evol. 29, 2157–2167. doi:10.1093/molbev/mss084

Dandri, M., Locarnini, S., 2012. New insight in the pathobiology of hepatitis B virus infection. Gut. doi:10.1136/gutjnl-2012-302056

Dill, J.A., Camus, A.C., Leary, J.H., Di Giallonardo, F., Holmes, E.C., Ng, T.F.F., 2016. Distinct Viral Lineages from Fish and Amphibians Reveal the Complex Evolutionary History of Hepadnaviruses. J. Virol. 90, 7920–7933. doi:10.1128/JVI.00832-16

Drummond, A.J., Ho, S.Y.W., Phillips, M.J., Rambaut, A., 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol. 4, 699–710. doi:10.1371/journal.pbio.0040088

Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7, 214. doi:10.1186/1471-2148-7-214

Geoghegan, J.L., Duchêne, S., Holmes, E.C., 2017. Comparative analysis estimates the relative frequencies of co-divergence and cross-species transmission within viral families. PLoS Pathog. 13, e1006215. doi:10.1371/journal.ppat.1006215

Godoy, B.A., Alvarado-Mora, M. V, Gomes-Gouvêa, M.S., Rebello Pinho, J.R., Fagundes, N.J.R., 2013. Origin of HBV and its arrival in the Americas - The importance of natural selection on time estimates. Antivir. Ther. 18, 505–512. doi:10.3851/IMP2600

Hahn, C.M., Iwanowicz, L.R., Cornman, R.S., Conway, C.M., Winton, J.R., Blazer, V.S., 2015. Characterization of a Novel Hepadnavirus in the White Sucker (Catostomus commersonii) from the Great Lakes Region of the United States. J. Virol. 89, 11801–11811. doi:10.1128/JVI.01278-15

Kass, R., Raftery, A., 1995. Bayes factors. J. Am. Stat. Assoc. 90, 773–795.

Kumar, S., Hedges, S.B., 2016. Advances in Time Estimation Methods for Molecular Data. Mol Biol Evol. 33, 863–869. doi:10.1093/molbev/msw026

Lanford, R.E., Chavez, D., Brasky, K.M., Burns, R.B., Rico-Hesse, R., Rico-Hesse, R., 1998. Isolation of a hepadnavirus from the woolly monkey, a New World primate. Proc. Natl. Acad. Sci. U. S. A. 95, 5757–61.

Li, S., Wang, Z., Li, Y., Ding, G., 2017. Adaptive evolution of proteins in hepatitis B virus during divergence of genotypes. Sci. Rep. 7, 1990. doi:10.1038/s41598-017-02012-8

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

Mühlemann, B., Jones, T.C., Damgaard, P.B., Allentoft, M.E., Shevnina, I., Logvin, A., Usmanova, E., Panyushkina, I.P., Boldgiv, B., Bazartseren, T., Tashbaeva, K., Merz, V., Lau, N., Smroka, V., Voyakin, D., Kitov, E., Epimakhov, A., Pokutta, D., Vicze, M., Price, T.D., Moiseyev, V., Hansen, A.J., Orlando, L., Rasmussen, S., Sikora, M., Vinner, L., Osterhaus, A.D.M.E., Smith, D.J., Glebe, D., Fouchier, R.A.M., Drosten, C., Sjogren, K.G., Kristiansen, K., Willerslev, E., 2018. Ancient hepatitis B viruses from the Bronze Age to the Medieval period. Nature 557, 418-423. doi:10.1038/s41586-018-0097-z

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

Paraskevis, D., Angelis, K., Magiorkinis, G., Kostaki, E., Ho, S.Y.W., Hatzakis, A., 2015. Dating the origin of hepatitis B virus reveals higher substitution rate and adaptation on

the branch leading to F/H genotypes. Mol. Phylogenet. Evol. 93, 44–54. doi:10.1016/j.ympev.2015.07.010

Paraskevis, D., Magiorkinis, G., Magiorkinis, E., Ho, S.Y.W., Belshaw, R., Allain, J.P., Hatzakis, A., 2013. Dating the origin and dispersal of hepatitis B virus infection in humans and primates. Hepatology 57, 908–916. doi:10.1002/hep.26079

Raghavan, M., Steinrücken, M., Harris, K., Schiffels, S., Rasmussen, S., DeGiorgio, M., Albrechtsen, A., Valdiosera, C., Ávila-Arcos, M.C., Malaspinas, A.S., Eriksson, A., Moltke, I., Metspalu, M., Homburger, J.R., Wall, J., Cornejo, O.E., Moreno-Mayar, J.V., Korneliussen, T.S., Pierre, T., Rasmussen, M., Campos, P.F., de Barros Damgaard, P., Allentoft, M.E., Lindo, J., Metspalu, E., Rodríguez-Varela, R., Mansilla, J., Henrickson, C., Seguin-Orlando, A., Malmström, H., Stafford Jr, T., Shringarpure, S.S., Moreno-Estrada, A., Karmin, M., Tambets, K., Bergström, A., Xue, Y., Warmuth,V., Friend, A.D., Singarayer, J., Valdes, P., Balloux, F., Leboreiro, I., Vera, J.L., Rangel-Villalobos, H., Pettener, D., Luiselli, D., Davis, L.G., Heyer, E., Zollikofer, C.P.E., Ponce de León, M.S., Smith, C.I., Grimes, V., Pike, K.A., Deal, M., Fuller, B.T., Arriaza, B., Standen, V., Luz, M.F., Ricaut, F., Guidon, N., Osipova, L., Voevoda, M.I., Posukh, O.L., Balanovsky, O., Lavryashina, M., Bogunov, Y., Khusnutdinova, E., Gubina, M., Balanovska, E., Fedorova, S., Litvinov, S., Malyarchuk, B., Derenko, M., Mosher, M.J., Archer, D., Cybulski, J., Petzelt, B., Mitchell, J., Worl, R., Norman, P.J., Parham, P., Kemp, B.M., Kivisild, T., Tyler-Smith, C., Sandhu, M.S., Crawford, M., Villems, R., Smith, D.G., Waters, M.R., Goebel, T., Johnson, J.R., Malhi, R.S., Jakobsson, M., Meltzer, D.J., Manica, A., Durbin, R., Bustamante, C.D., Song, Y.S., Nielsen, R., Willerslev, E., 2015. Genomic evidence for the Pleistocene and recent

population history of Native Americans. Science 349, aab3884. doi:10.1126/science.aab3884

Reis, M., Donoghue, P.C., Yang, Z., 2016. Bayesian molecular clock dating of species divergences in the genomics era. Nat. Rev. Genet. 17, 71-80. doi:10.1038/nrg.2015.8

Schweitzer, A., Horn, J., Mikolajczyk, R.T., Krause, G., Ott, J.J., 2015. Estimations of worldwide prevalence of chronic hepatitis B virus infection: A systematic review of data published between 1965 and 2013. Lancet 386, 1546–1555. doi:10.1016/S0140-6736(15)61412-X

Shimodaira, H., Hasegawa, M., 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics. 17, 1246-1247.

Simmonds P., 2001. Reconstructing the origins of human hepatitis viruses. Philos. Trans. R. Soc. Lond. B Biol. Sci. 356, 1013-1026. doi:10.1098/rstb.2001.0890

Souza, B.F.C.D., Konig, A., Rasche, A., de Oliveira Carneiro, I., Stephan, N., Corman, V.M., Roppert, P.L., Goldmann, N., Kepper, R., Müller, S.F., Völker, C, de Souza, A.J.S., Gomes-Gouvêa, M.S., Moreira-Soto, A., Stöcker, A., Nassal, M., Franke, C.R., Rebello Pinho. J.R., Soares, M.D.C.P., Geyer, J., Lemey, P., Drosten, C., Netto, E.M., Glebe, D., Drexler, J.F., 2018. A novel hepatitis B virus species discovered in capuchin monkeys sheds new light on the evolution of primate hepadnaviruses. J. Hepatol. 68, 1114-1122. doi:10.1016/j.jhep.2018.01.029

Stamatakis, A., 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313. doi:10.1093/bioinformatics/btu033

Tatematsu, K., Tanaka, Y., Kurbanov, F., Sugauchi, F., Mano, S., Maeshiro, T., Nakayoshi, T., Wakuta, M., Miyakawa, Y., Mizokami, M., 2009. A Genetic Variant of Hepatitis B

Virus Divergent from Known Human and Ape Genotypes Isolated from a Japanese Patient and Provisionally Assigned to New Genotype J. J. Virol. 83, 10538–10547. doi:10.1128/JVI.00462-09

WHO, 2017. Global Hepatitis Report, 2017, World Health Organization. doi:ISBN 978-92-4-156545-5

Yang, Z., 2007. PAML 4: Phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24, 1586–1591. doi:10.1093/molbev/msm088

Yang, Z., Wong, W.S.W., Nielsen, R., 2005. Bayes empirical Bayes inference of amino acid sites under positive selection. Mol. Biol. Evol. 22: 1107-1118. doi:10.1093/molbev/msi097

Table 1

Marginal likelihood and Bayes factor comparison among HBV topologies estimated using Bayesian inference (BI).

Topological Hypothesis	**Marginal Likelihood[1]**	**Bayes Factor**		
		ML-tree	**ML(relax)-tree**	**BI-tree**
Including the Woolly Monkey isolate				
ML-tree	-49200.18	-	-21.84	-43.01
ML(relax)-tree	-49189.26	21.84	-	-21.17
BI-tree	**-49178.67**	**43.01**	**21.17**	**-**
Not including the Woolly Monkey isolate				
ML-tree	-47153.71	-	-7.87	-53.54
ML(relax)-tree	-47149.78	7.87	-	-45.67
BI-tree	**-47126.94**	**53.54**	**45.67**	**-**

[1]Log-Likelihood. The best-supported topological hypothesis is shown in bold.

Table 2

Positive selection analyses of non-overlapping HBV genome regions based on ML-tree and BI-tree alternative hypotheses.

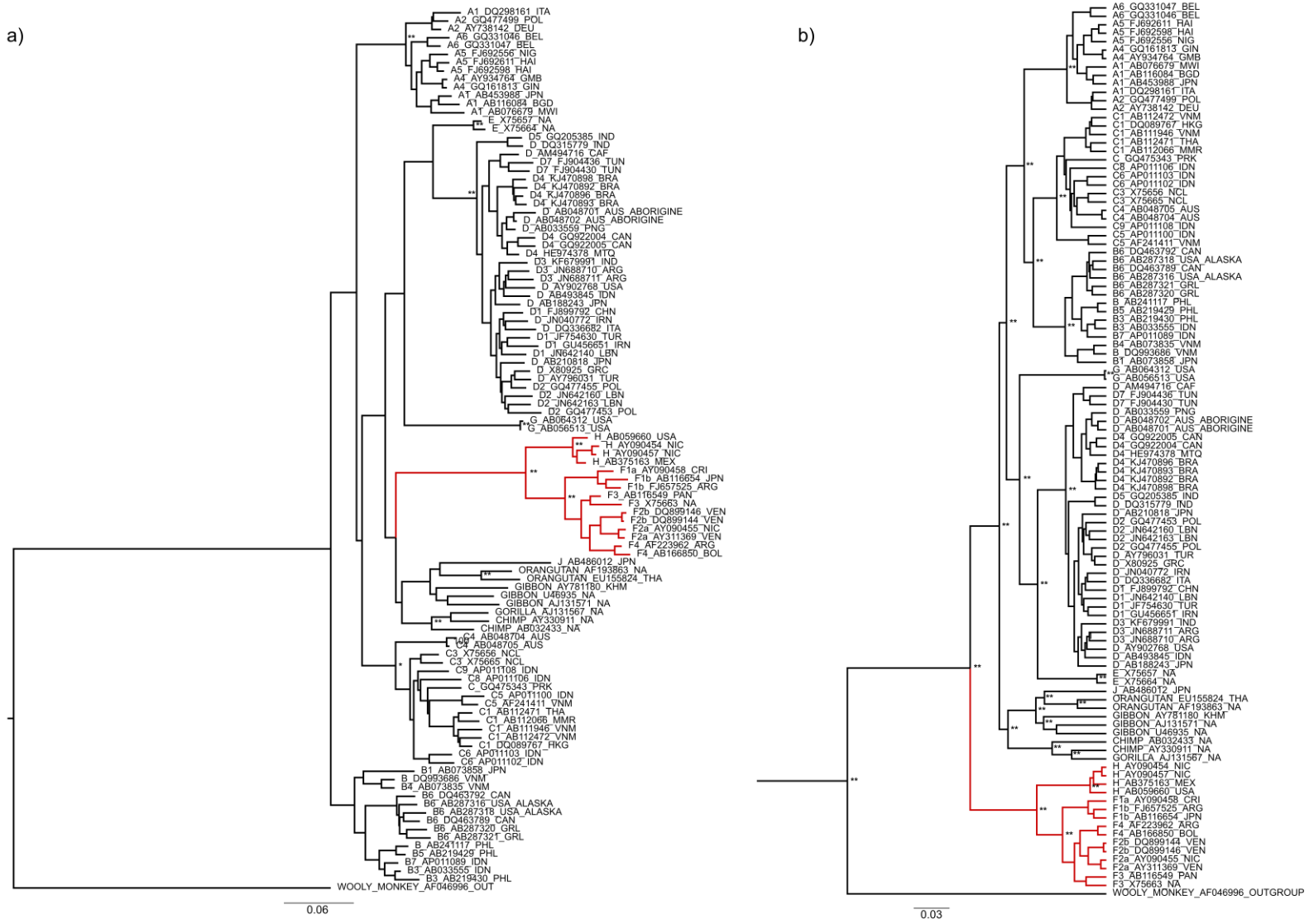
	Genome Partition	LRT[1]	*P*-value	Codons under positive selection[2]
ML-tree	P1	1.125	0.289	-
	P2	8.580	**0.003**	**69, 75**
	C	1.073	0.300	-
	X	7.348	**0.007**	**18**
BI-tree	P1	0.000	1.000	-
	P2	8.670	**0.003**	**69, 75**
	C	0.000	1.000	-
	X	8.800	**0.003**	**18**

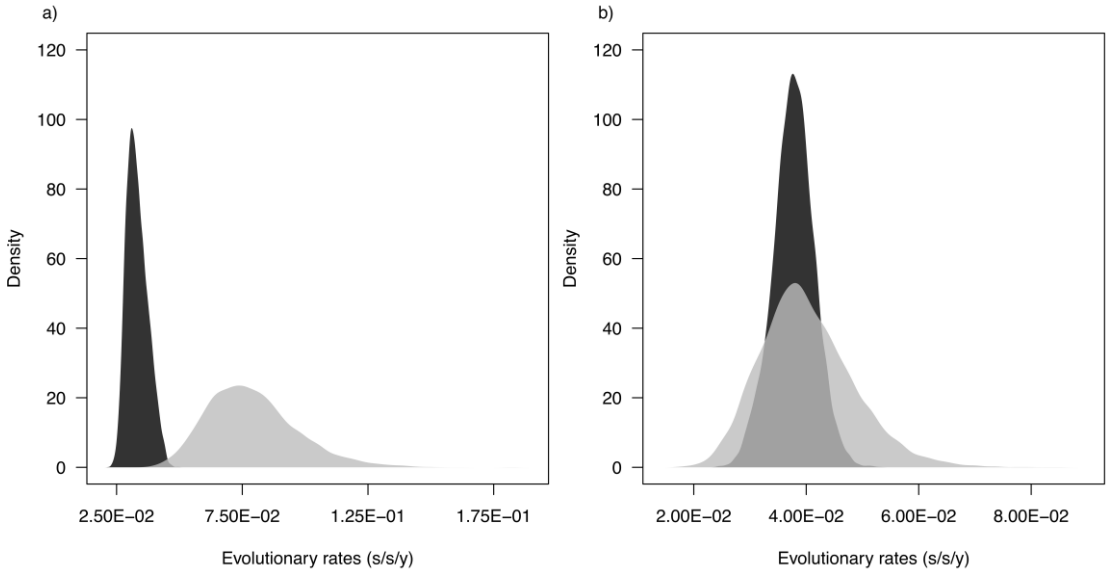
[1]Likelihood-ratio test, [2]Based on BEB Posterior Probability > 0.95

Figure Legends

Fig. 1. The best-supported topology for HBV phylogeny under a) a ML criterion without a molecular clock (ML-tree), or b) a BI criterion assuming a relaxed molecular clock (BI-tree). HBV genotypes are broadly indicated in terminals. Node support (bootstrap support (BS), in the case of ML, posterior probabilities (PP), in the case BI) are indicated for major clades as follows: *BS≥90 or PP≥0.90; **BS≥95, PP≥0.95. Unlabeled nodes had BS<90 or PP<0.90.

Fig. 2. Posterior density of evolutionary rates (s/s/y) estimated for the HBV-F/H branch (black line) and for background branches (gray line) assuming the a) ML-tree or b) BI-tree as the true topology.


a)
A1_DQ298161_ITA
A2_GQ477499_POL
A2_AY738142_DEU
A6_GQ331046_BEL
A6_GQ331047_BEL
A5_FJ692556_NIG
A5_FJ692611_HAI
A5_FJ692598_HAI
A4_AY934764_GMB
A4_GQ161813_GIN
AT_AB453988_JPN
A1_AB116084_BGD
AT_AB076679_MWI
E_X75657_NA
E_X75664_NA
D5_GQ205385_IND
D_DQ315779_IND
D_AM494716_CAF
D7_FJ904436_TUN
D7_FJ904430_TUN
D4_KJ470898_BRA
D4_KJ470892_BRA
D4_KJ470896_BRA
D4_KJ470893_BRA
D_AB048701_AUS_ABORIGINE
D_AB048702_AUS_ABORIGINE
D_AB033559_PNG
D4_GQ922004_CAN
D4_GQ922005_CAN
D4_HE974378_MTQ
D3_KF679991_IND
D3_JN688710_ARG
D3_JN688711_ARG
D_AY902768_USA
D_AB493845_IDN
D_AB188243_JPN
D1_FJ899792_CHN
D_JN040772_IRN
D_DQ336682_ITA
D1_JF754630_TUR
D1_GU456651_IRN
D1_JN642140_LBN
D_AB210818_JPN
D_X80925_GRC
D1_AY796031_TUR
D2_GQ477455_POL
D2_JN642160_LBN
D2_JN642163_LBN
D2_GQ477453_POL
G_AB064312_USA
G_AB056513_USA
H_AB059660_USA
H_AY090454_NIC
H_AY090457_NIC
H_AB375163_MEX
F1a_AY090458_CRI
F1b_AB116654_JPN
F1b_FJ657525_ARG
F3_AB116549_PAN
F3_X75663_NA
F2b_DQ899146_VEN
F2b_DQ899144_VEN
F2a_AY090455_NIC
F2a_AY311369_VEN
F4_AF223962_ARG
F4_AB166850_BOL
J_AB486012_JPN
ORANGUTAN_AF193863_NA
ORANGUTAN_EU155824_THA
GIBBON_AY781180_KHM
GIBBON_U46935_NA
GIBBON_AJ131571_NA
GORILLA_AJ131567_NA
CHIMP_AY330911_NA
CHIMP_AB032433_NA
C4_AB048704_AUS
C4_AB048705_AUS
C3_X75656_NCL
C3_X75665_NCL
C9_AP011108_IDN
C8_AP011106_IDN
C_GQ475343_PRK
C5_AP011100_IDN
C5_AF241411_VNM
C1_AB112471_THA
C1_AB112066_MMR
C1_AB111946_VNM
C1_AB112472_VNM
C1_DQ089767_HKG
C6_AP011103_IDN
C6_AP011102_IDN
B1_AB073858_JPN
B_DQ993686_VNM
B4_AB073835_VNM
B6_DQ463792_CAN
B6_AB287316_USA_ALASKA
B6_AB287318_USA_ALASKA
B6_DQ463789_CAN
B6_AB287320_GRL
B6_AB287321_GRL
B_AB241117_PHL
B5_AB219429_PHL
B7_AP011089_IDN
B3_AB033555_IDN
B3_AB219430_PHL
WOOLY_MONKEY_AF046996_OUT
0.06
b)
A6_GQ331047_BEL
A6_GQ331046_BEL
A5_FJ692611_HAI
A5_FJ692598_HAI
A5_FJ692556_NIG
A4_GQ161813_GIN
A4_AY934764_GMB
A1_AB076679_MWI
A1_AB116084_BGD
A1_AB453988_JPN
A1_DQ298161_ITA
A2_GQ477499_POL
A2_AY738142_DEU
C1_AB112472_VNM
C1_DQ089767_HKG
C1_AB111946_VNM
C1_AB112471_THA
C1_AB112066_MMR
C_GQ475343_PRK
C8_AP011106_IDN
C6_AP011103_IDN
C6_AP011102_IDN
C3_X75656_NCL
C3_X75665_NCL
C4_AB048705_AUS
C4_AB048704_AUS
C9_AP011108_IDN
C5_AP011100_IDN
C5_AF241411_VNM
B6_DQ463792_CAN
B6_AB287318_USA_ALASKA
B6_DQ463789_CAN
B6_AB287316_USA_ALASKA
B6_AB287321_GRL
B6_AB287320_GRL
B_AB241117_PHL
B5_AB219429_PHL
B3_AB219430_PHL
B3_AB033555_IDN
B7_AP011089_IDN
B4_AB073835_VNM
B_DQ993686_VNM
B1_AB073858_JPN
G_AB064312_USA
G_AB056513_USA
D_AM494716_CAF
D7_FJ904436_TUN
D7_FJ904430_TUN
D_DQ315779_IND
D_AB033559_PNG
D_AB048702_AUS_ABORIGINE
D_AB048701_AUS_ABORIGINE
D4_GQ922005_CAN
D4_GQ922004_CAN
D4_HE974378_MTQ
D4_KJ470896_BRA
D4_KJ470893_BRA
D4_KJ470892_BRA
D4_KJ470898_BRA
D5_GQ205385_IND
D_DQ315779_IND
D_AB210818_JPN
D2_GQ477453_POL
D2_JN642160_LBN
D2_JN642163_LBN
D2_GQ477455_POL
D_AY796031_TUR
D_X80925_GRC
D_JN040772_IRN
D_DQ336682_ITA
D1_FJ899792_CHN
D1_JN642140_LBN
D1_JF754630_TUR
D1_GU456651_IRN
D3_KF679991_IND
D3_JN688711_ARG
D3_JN688710_ARG
D_AY902768_USA
D_AB493845_IDN
D_AB188243_JPN
E_X75657_NA
E_X75664_NA
J_AB486012_JPN
ORANGUTAN_EU155824_THA
ORANGUTAN_AF193863_NA
GIBBON_AY781180_KHM
GIBBON_AJ131571_NA
GIBBON_U46935_NA
CHIMP_AB032433_NA
CHIMP_AY330911_NA
GORILLA_AJ131567_NA
H_AY090454_NIC
H_AY090457_NIC
H_AB375163_MEX
H_AB059660_USA
F1a_AY090458_CRI
F1b_FJ657525_ARG
F1b_AB116654_JPN
F4_AF223962_ARG
F4_AB166850_BOL
F2b_DQ899144_VEN
F2b_DQ899146_VEN
F2a_AY090455_NIC
F2a_AY311369_VEN
F3_AB116549_PAN
F3_X75663_NA
WOOLY_MONKEY_AF046996_OUTGROUP
0.03


a)
b)
Density
Evolutionary rates (s/s/y)
0
20
40
60
80
100
120
2.50E−02
7.50E−02
1.25E−01
1.75E−01
2.00E−02
4.00E−02
6.00E−02
8.00E−02

CAPÍTULO III

"EVOLUTIONARY RELATIONSHIPS AMONG HEPATITIS B VIRUS GENOTYPE D IN LATIN AMERICA AND EUROPE"

Manuscrito em preparação para submissão no *Journal of General Virology.*

Evolutionary relationships among Hepatitis B Virus genotype D in Latin America and Europe

Bibiane A. Godoy[1], Michele S. Gomes-Gouvêa[2], Romina Salpini[3], Valentina Svicher[3], Francesca Ceccherini-Silberstein[3], Carlo F. Perno[3], João R.R. Pinho[2,4], Nelson J. R. Fagundes[1,5*]

[1]Postgraduation Program in Genetics and Molecular Biology, Federal University of Rio Grande do Sul, Porto Alegre, RS, Brazil

[2]Institute of Tropical Medicine and School of Medicine, Department of Gastroenterology, University of São Paulo School of Medicine, São Paulo, SP, Brazil

[3]Department of Experimental Medicine and Surgery, University of Rome Tor Vergata, Rome, Italy

[4]Hospital Israelita Albert Einstein, São Paulo, SP, Brazil

[5]Department of Genetics, Institute of Biosciences, Federal University of Rio Grande do Sul, Porto Alegre, RS, Brazil

*Correspondence: Nelson J. R. Fagundes nelson.fagundes@ufrgs.br

Keywords: HBV; Genetic population structure, Population admixture, Bayesian Skyline plot

Abbreviations: ARG, Argentina; BR-BA, Brazilian state of Bahia; BR-MA, Brazilian state of Maranhão; BR-S, Southern Brazil; BR-SE, Southeastern Brazil; CUB, Cuba; DEU+POL, Germany and Poland; ESP+FRA, Spain and France; HBV, Hepatitis B Virus; HBV/A, HBV genotype A; HBV/A1, HBV subgenotype A1; HBV/D, HBV genotype D; HBV/D1, HBV subgenotype D1; HBV/D2, HBV subgenotype D2; HBV/D3, HBV subgenotype D3; HBV/D4, HBV subgenotype D4; HBV/F, HBV genotype F; HBV/H, HBV genotype H; HTI, Haiti; IND, India; ITA, Italia; LBN+TUR, Lebanon and Turkey; ML, Maximum likelihood; NLD, Netherlands; ORF, open reading frames.

Abstract

Hepatitis B Virus (HBV) is a hepatotropic virus that infects 7 to 12 million people in Latin America. HBV genotype D (HBV/D) has a worldwide distribution and, in Latin America, is associated to European migrations. Indeed, some Latin America regions where HBV/D reaches higher frequencies, such as Southern Brazil, had received an important flow of Italian migrants in the late 19th century. In this study, we used phylogenetic and population genetics approaches to compare HBV/D isolates from Europe and Latin America and establish evolutionary relationships between viral distribution and historical events associated to HBV spread. Our results showed a predominance of HBV/D3 towards southern South America with subgenotype distributions mirroring those present in Italy. On the other hand, the Caribbean and Northeastern Brazil showed a distinct subgenotype distribution, with HBV/D4 reaching high frequencies. At the subgenotypic level, HBV/D3 was more related to Italian migration in Southeast Brazil (BR-SE), and evidenced an older population expansion shared by both populations, which may suggest that BR-SE acted as a source population in Latin America. In addition, we found similarities between Germany and Poland HBV/D3 and other Latin America populations, and also between Germany, Poland and BR-SE for HBV/D2. Finally, HBV/D4 showed a specific population dynamics, with marked genetic differences among localities. This study highlighted a complex history for HBV/D dispersal across Latin America, in which multiple sources populations contributed to the current HBV/D gene pool in this region.

INTRODUCTION

Hepatitis B Virus (HBV) is a hepatotropic virus transmitted by contact with blood or other body fluids from infected persons, whose contagion results in an acute or chronic infection with clinical outcomes that may range from asymptomatic hepatitis to fulminant liver failure [1]. The general prevalence of HBV is quite variable and reaches the highest rates in Western Pacific and African regions, where more than 6% of the adult population is infected [2]. The increase of vaccination coverage represents an effective way to prevent new HBV infections and has shown significant impact in reducing rates in many countries [3, 4]. However, even in areas of low overall HBV prevalence, more vulnerable groups such as migrants [5–8] and people living under low socio-economic or marginalized conditions may represent focus of HBV infection [9–12]. Currently, there are around 257 million people worldwide presenting the incurable HBV chronic infection, causing a high number of annual deaths due to complications of the disease, keeping HBV as a major global health problem [2].

HBV presents an enveloped structure that contains a nucleocapsid formed by the core protein, which harbors the partially double-stranded circular DNA genome. HBV genome is compact and encodes four partially overlapping open reading frames (ORFs) [13]: C (core protein), X (regulatory HBx protein), P (polymerase), and S (surface proteins), which is the most commonly used region for characterization of different viral strains via DNA sequencing [14, 15]. The comparison between complete HBV genomes has led to the classification into ten strains affecting humans (A-J) designated as HBV genotypes [16–18]. Genotypes present DNA sequence divergence above 7.5% and are further divided into subgenotypes based on intragroup divergences between 4 and 7.5% along with robust phylogenetic support [19]. HBV genotypes and subgenotypes show distinct geographical

distributions that, in general, reflect ethnic and/or historical profiles, even though recent migratory waves may blur these patterns to some extent [20, 21].

A recent study identified HBV DNA in human remains dated between ~800-4,500 years old and showed that the sequences are closely related to modern HBV genotypes [22]. This ancient coevolution between HBV and human populations reinforces the geographical structure shown by HBV genotype distribution, and allows to associate human population history to viral epidemiology [21, 23]. HBV/D has been identified in ancient samples from Central Asia [22], and is currently classified into at least six subgenotypes (D1, D2, D3/D6, D4, D5, D7/D8) [24], which are related to Eurasian ethnicity. The HBV/D occurrence are predominant in the Mediterranean basin, but are currently distributed worldwide due to intense European migration and colonization [25, 26].

There are 7 to 12 million people infected by HBV in Latin America and, although the overall prevalence rate is low, the Caribbean and the Amazon Basin show higher rates, accounting for the vast majority of HBV infections in the Americas [27, 28]. Originally settled by Native Americans, an increasing influx of European migrants followed the European invasion in the early 16th century. In addition, the African slave trade until the 19th century and new European migration waves during the 19th and 20th centuries resulted in a long period of miscegenation [29, 30]. This complex demographic history produced a mixture of "genetic ancestries" that is also reflected in the distribution of HBV genotypes in the region. HBV/F and HBV/H are considered indigenous of the Americas and is usually related to Native American ancestry, being the most prevalent HBV genotypes in Mexico and other Spanish-speaking countries [31, 32]. In contrast, HBV/A genotype is the most frequent HBV genotype in Brazil (especially HBV/A1), and has a high prevalence in Haiti. This HBV genotype has been associated to African populations and to the slave trade that

affected both countries during colonial times [33–35]. HBV/D represents the most common genotype in Southern Brazil, where European genetic ancestry is high [36], and has been associated with more recent European migrations to the region during late 19th century, especially from Italy [37–39]. In this study, we use phylogenetic and population genetics approaches to compare HBV/D isolates from Europe and Latin America and establish evolutionary relationships between viral distribution and historical events associated to HBV spread.

METHODS

Data Assembly

This study involved a total of 1,390 HBV/D sequences from 13 countries (Fig. 1; Table 1; Table S1). Initially, we included 219 HBV sequences covering 1,306 base pairs (bp) in the S and P (S/Pol) genome region, which were collected between 2006 and 2011 from three Brazilian regions (South, Southeast and Northeast) [40]. To this dataset, we added 79 HBV/D sequences of 678bp collected in clinical centers from Italy between 2007 and 2014 [41]. To allow a thorough comparison between Latin America and European HBV/D isolates, we retrieved from the GenBank HBV/D sequences isolated from European (or Middle Eastern) countries prioritizing both a known historical link with Latin America (Lebanon and Turkey, for example), and alignment overlap.

We considered as inclusion criterion the number (at least 30 isolated by country) and length (678–1306 bp) for the available HBV/D sequences. In some cases, isolates from different countries were merged to allow a larger dataset within a specific cultural or geographical region: 1) We merged Spanish and French strains (ESP+FRA) to have a "Western Latin" population. 2) We merged German and Polish strains (DEU+POL) to have

a “Northeastern European” population. 3) We merged Lebanese and Turkish (LBN+TUR) to have a “Middle Eastern” population. We also included Dutch (NLD) and Italian (ITA) samples in the dataset. Strains isolated from India (IND) were included as “outgroup”. The locations abbreviations used in this study are in accordance with ISO 3166 [42].

Since Brazil is a large country with vary distinct regional histories, we divided the Brazilian sample into South (BR-S) and Southeast (BR-SE) regions. The Northeastern states of Maranhão (BR-MA) and Bahia (BR-BA) were considered as independent populations due to a marked difference in HBV/D subgenotype distributions. Other Latin American populations included in our analyses were from Argentina (ARG), Cuba (CUB) and Haiti (HTI). We confirmed the classification of all sequences as HBV/D using phylogenetic methods (data not shown). Subgenotype identity was confirmed using BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi), since sequence length was too short for reliable identification using phylogenetic methods (data not shown).

Alignment, recombination analysis and phylogenetic inference

We included 20 reference sequences from other HBV genotypes to our HBV/D dataset, and performed the sequence alignment using the ClustalW algorithm implemented in MEGA v.7 [43]. All HBV/D sequences were checked for recombination using the RDP, GENECOV and MAXCHI methods implemented in RDP v.4 [44]. All sequences showing evidence for recombination by at least two methods were excluded from further analyses.

To improve phylogenetic signal and optimize computational time, we removed identical sequences and sequences smaller than 920bp to build a final alignment of 914 sequences (Table S1). We used MEGA v.7 [43] to infer the best-fitting nucleotide substitution model (GTR+G). A maximum likelihood (ML) HBV phylogeny was estimated with RAxML

v8.2.10 [45] using ten independent searches, and 1,000 bootstrap replicates to evaluate clade support. The estimated phylogeny was visualized and edited in FigTree v1.4.3 (available online: http://tree.bio.ed.ac.uk/).

Population comparisons

We used the fixation index (Φ_{ST}) to quantify the genetic structure among populations. Φ_{ST} is analogous to the classical F_{ST} measure, which can be interpreted as a measure of identity by descent among populations [46], but it incorporates the mutational distance among haplotypes in the calculation [47]. In order to avoid the impact of missing data, we used a 678bp fragment common to the whole HBV/D sample. The appropriate parameters for the best-fitting nucleotide substitution model (GTR+G) were estimated in MEGA v.7 [43].

Pairwise Φ_{ST}s were calculated in Arlequin v3.5.2 [48] assuming the Tamura-Nei [49] +G substitution model, as Arlequin does not allow computation based on the GTR model [47]. The statistical significance of estimates was assessed by 10,000 permutations and adjusted using the Bonferroni correction for multiple-comparisons. We performed Φ_{ST} comparisons for the whole dataset and for each subgenotype (HBV/D1-D4). For HBV/D4, we merged samples from Spain, France and Italy in a "Southern Europe" population, in order to allow enough samples for comparison.

Population dynamics of HBV/D and its subgenotypes

Variation of effective population size through time was estimated using the Bayesian Skyline method [50]. The full HBV/D alignment (1,306 bp) was divided by locations and the substitution nucleotide models for each set were estimated with MEGA v.7 [43]. Bayesian phylogenetic analyzes were performed in BEAST v1.8.4 [51] using 50,000,000 states of Markov Chain Monte Carlo, sampling every 1,000 steps, and discarding the first 5,000 samples as burnin. We assumed a strict molecular clock using a normal prior distribution with a mean of $1.0x10^{-5}$ substitutions per site per year (s/s/y), and a standard deviation of $5.0x10^{-6}$ s/s/y, truncated at $2.0x10^{-5}$ and $6.0x10^{-6}$ s/s/y, which consider a range of likely rates discussed in previous studies [52]. The Bayesian Skyline plots resulting from these analyses were generated in Tracer v1.7 [51].

RESULTS

Distribution and frequency of HBV/D and its subgenotypes

The geographical distribution and the subgenotype classification of the 1,390 HBV/D samples included in this study is shown in both Fig. 1 and Table 1. ITA showed a predominance of HBV/D3, followed by HBV/D2 and HBV/D1. The same pattern was observed in BR-S and BR-SE, and was similar in ARG, except for a higher HBV/D1 frequency in this country. HBV/D1 presents a low prevalence in Latin America, but was the most common subgenotype in LBN+TUR (where it reached its highest frequency), NLD, and DEU+POL, being also quite prevalent in IND. Central America and Northeastern Brazil showed a peculiar HBV/D subgenotype distribution, with a high prevalence of HBV/D4. HBV/D5 occurred only in IND, while HBV/D7 occurred in CUB, BR-SE, ESP+FRA, NLD and ITA, though always in relatively low frequency.

Phylogenetic relations among HBV/D isolates

The ML tree for HBV genotypes support a monophyletic clade for HBV/D (Fig. 2, Fig. S1), even though not all subgenotypes could be recovered as a well-supported clade. HBV/D4 showed certain geographical structuring within Latin America, with BR-BA + BR-MA isolates clustering close to each other, CUB + HTI clustering close to each other (and to a pair of ESP+FRA samples), and BR-SE scattered along the tree. Concerning HBV/D7, most Latin American isolates (CUB) were more closely related to ESP+FRA and distant to the isolate from BR-SE (Fig. 2). However, even if some patterns seem suggestive of a shared evolutionary history, determining parental-ancestral population relationships in the tree was far from trivial.

For HBV/D1, HBV/D2, and HBV/D3 the picture was even more complex (Fig. 3a, Fig. 3b, Fig. S2, Fig. S3). The distribution of internal nodes of these clades was quite chaotic, showing low geographical structuring. There were no coherent "geographic clades" with few exceptions, frequently involving IND isolates, but also in the internal relationships within HBV/D2. Nonetheless, as it was the case for HBV/D4, it was not possible to infer direct migration fluxes based on tree topology alone.

Pairwise population comparison using Φ_{ST}s

Considering HBV/D as a whole, pairwise Φ_{ST} estimates showed low values between several Latin America x European population pairs (Fig. 4) which may be indicative of recent migration. Although statistically significant, ARG, BR-S and BR-SE were very similar to ITA. BR-BA also has a low value vs. ITA, even though this larger value could reflect the higher prevalence of HBV/D4 in this population. On the other hand, BR-MA, CUB, and HTI showed large values against all European samples used in our study.

Comparisons at the subgenotype level allow us to refine the inference of population relationships, but come at the expense of statistical power (see Discussion). Below, we emphasize the closest genetic relationship indicated by our data, even though in some cases, a given Latin American population may have had non-significant values against several European populations. Estimates for HBV/D1 showed that while Brazilian populations (BR-S, BR-SE) were closest to NLD, ARG and CUB were closest to LBN+TUR. For HBV/D2, ARG was closest to ITA, while BR-S and BR-SE were closest to DEU+POL. For HBV/D3, all Latin American populations were closest to ITA. Finally, for HBV/D4, all Latin American populations were closest to Southern European (ESP+FRA+ITA), even though only BR-SE showed a low Φ_{ST} value. Brazilian Northeastern populations (BR-BA and BR-MA) showed intermediate values, while Caribbean populations showed large Φ_{ST} values.

Population dynamics of HBV/D

The Bayesian Skyline plots indicated a wide variation in viral population dynamics (Fig. 5). Among European populations, ITA had the most ancient phylogenetic history for HBV/D, with an ancient pulse of viral epidemics (i.e. viral population growth) ~2,000 years ago. In contrast, other European populations showed signal of viral epidemics around or more recent than 1,000 years ago. Latin American populations, on the other hand, showed different patterns. ARG and HTI had no evidence of viral epidemics, while CUB showed a gradual and modest increase in viral population size. On the other hand, all Brazilian populations had a stronger signal of viral population growth. However, dating estimates suggest that these expansions, as well as the phylogenetic depth of HBV/D in all Latin American populations predate the arrival of HBV/D in the Americas. These results possibly

indicate that a substantial amount of phylogenetic diversity within HBV/D entered Latin America following the European invasion.

DISCUSSION

This study compared circulating HBV/D sequences in Latin America and Europe using phylogenetic and population genetic approaches to unravel the arrival and spread of this genotype in Latin America. HBV/D is the third most prevalent HBV genotype in Latin America after HBV/F and HBV/A [53], and has been associated with the European colonization [54, 55]. While Iberian colonizers (Spanish and Portuguese) established early settlements in Latin America since the 16th century, it was only in the beginning of 19th century that European migrants from other nationalities arrive in higher numbers [30].

To take the Brazilian case [56], it is at this time that the first Germany colonies were formed in the BR-S, spreading to ARG and BR-SE. From 1870 a massive Italian migration occurs encouraged by the Brazilian government, which was initially targeted to BR-SE in order to replace the labor force in coffee plantations in the imminence of slavery abolition. From 1889, Italian immigrants were increasingly directed to BR-S. However, due to poor working conditions, the Italian government stopped accepting the subsidy offered by the Brazilian government, which reduced the flow of immigrants to the country [56]. Argentina also started receiving Italian immigrants since 1870, similarly to Brazil [57]. Indeed, Brazil and Argentina accounted for the largest influx of Italian immigrants in Latin America [29]. In these countries, high frequencies of HBV/D are coincident with regions under an important influence of Italian immigration [35, 37, 38][58, 59].

Our results are in agreement with this historical account. ARG, BR-S and BR-SE have low Φ_{ST} values compared to ITA, indicating that most HBV/D genotypes in these

regions result from this population flow (Fig. 4). BR-BA also had a relatively low Φ_{ST} value compared to ITA. However, differently from ARG, BR-S, and BR-SE, genotype distribution differs strikingly from ITA, with an important frequency of HBV/D4 in BR-BA (Fig. 1; Table 1). On the other hand, BR-MA, CUB and HTI showed large Φ_{ST} values against any European population, indicating a distinct and undefined population relationship concerning HBV/D isolates (Fig 4).

This general picture is obviously insufficient to account for the whole complexity of European immigration history in Latin America. One possibility to refine our historical inferences is analyzing each HBV/D subgenotype separately. However, these results must be taken with caution, since this strategy reduces sample sizes for the comparisons, therefore affecting statistical power. Having these caveats in mind, some population patterns deserve further discussion. The general similarity between ARG, BR-S, BR-SE and ITA is highly influenced by their genetic similarities in HBV/D3 (Fig.4), which was the most prevalent HBV/D subgenotype in these populations and in BR-BA. Indeed, all HBV/D3 from Latin American populations, even for BR-MA and HTI, were closest to ITA than to any other European population (Fig. 4).

HBV/D1 and HBV/D2 reveal differences between ARG, on the one hand, and BR-S and BR-SE, on the other, with respect to their closest European population. Brazilian samples were closest to NLD (ITA being the second closest population for both) considering HBV/D1, and to DEU+POL considering HBV/D2, while ARG was closest to LIB+TUR and ITA for HBV/D1 and HBV/D2, respectively (Fig. 4). While the link between Brazilian population and DEU+POL is supported by immigration history (see above), the close relationship between BR-S, BR-SE and NDL is surprising, as Dutch incursions during the colonial period occurred in Northeastern Brazil [60] Further studies will be necessary to

evaluate if this is a spurious signal or not. While both Argentina and Brazil received immigrants from the Middle East, the Lebanese immigration was much stronger in the former [61] which may explain the distinct evolutionary relationships between these populations. On its turn, HBV/D2 similarities between ARG and ITA reinforces the possible impact of Italian immigration on HBV/D genotypes in Latin America.

The history of HBV/D4 is more elusive. Our phylogenetic analysis (Fig. 2) suggests three "genetic groups" for HBV/D4 in Latin America: a) BR-SE isolates, which are not clustered in the phylogeny; b) BR-MA and BR-BA isolates, which are very closely related; and c) CUB and HTI isolates (including a couple of isolates from ESP+FRA). This latter group could reflect the Spanish and French influence in the Caribbean [30, 62]. However, when a larger dataset of Caribbean isolates is considered in the Φ_{ST} analysis, this close relationship disappears (Fig. 4). Indeed, the Φ_{ST} analysis suggests a significant distinction between any Caribbean and European pairwise comparison, while Brazilian HBV/D4 is not significantly different from the Southern European (ESP+FRA+ITA) population, even though BR-MA had a somewhat larger Φ_{ST} value (Fig. 4).

HBV/D4 can be found as a rare subgenotype in some African countries including Morocco [63], Ghana [64], Kenya [65], Rwanda and Somalia [66]. Its relatively widespread distribution in Africa compared to its paucity in Europe has raised the hypothesis of an African origin for this strain with a dispersion to Caribbean region and Brazil following slave trade [67, 68]. This is an intriguing possibility, which is compatible with our results especially for the Caribbean populations.

HBV/D population dynamics estimated in the Bayesian Skyline plots (Fig. 5) reveal two important patterns: 1) among European populations, ITA has the most ancient history

and the most ancient signal of a viral population expansion, which may indicate a central role for Italy in the HBV/D spread across Europe. 2) all Latin American populations have a viral demographic history more ancient than 500 years ago, thus predating its arrival in the Americas. In other words, evolutionary demographic signals shown by these populations reflect events in the source population(s) harboring these isolates. A similar pattern was found by Godoy et al. [10] or HBV/A1 viral population dynamics in Brazil. This finding is not compatible with a strong founder effect in the arrival of HBV/D in the Americas, and is corroborated by the general lack of coherent geographical clades in HBV/D phylogenies, and by the fact that different HBV/D subgenotypes in a single derived (Latin American) population may have come from different ancestral populations.

CONCLUSIONS

In summary, our study suggests a complex history for HBV/D in Latin America. Even though there is evidence for the influence of Italian immigration on populations like ARG, BR-S, and BR-SE, this effect seems to be more pronounced on HBV/D3. Other subgenotypes reveal closer affinities with other putative source populations such as German, Polish, or Libanese, indicating a mixture of HBV/D ancestries across Latin America. The origins of HBV/D4 was more elusive, and the hypothesis of an African origin connected to slave trade to the Caribbean cannot be refuted based on present analyses.

Funding information

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, by the Conselho Nacional de

Desenvolvimento Científico e Tecnológico (CNPq), and by the Postgraduation Program in Genetics and Molecular Biology (PPGBM-UFRGS).

References

1. **Trépo C, Chan HLY, Lok A**. Hepatitis B virus infection. *Lancet* 2014;384:2053–2063.
2. **World Health Organization**. Global Hepatitis Report, 2017. *Who*. Epub ahead of print 2017. DOI: ISBN 978-92-4-156545-5.
3. **Ropero Álvarez AM, Pérez-Vilar S, Pacis-Tirso C, Contreras M, El Omeiri N, *et al.*** Progress in vaccination towards hepatitis B control and elimination in the Region of the Americas. *BMC Public Health*;17. Epub ahead of print 2017. DOI: 10.1186/s12889-017-4227-6.
4. **Schweitzer A, Horn J, Mikolajczyk RT, Krause G, Ott JJ**. Estimations of worldwide prevalence of chronic hepatitis B virus infection: A systematic review of data published between 1965 and 2013. *Lancet* 2015;386:1546–1555.
5. **Coppola N, Alessio L, Gualdieri L, Pisaturo M, Sagnelli C, *et al.*** Hepatitis B virus infection in undocumented immigrants and refugees in Southern Italy: Demographic, virological, and clinical features. *Infect Dis Poverty*;6. Epub ahead of print 2017. DOI: 10.1186/s40249-016-0228-4.
6. **Richter C, Beest G Ter, Gisolf EH, Van Bentum P, Waegemaekers C, *et al.*** Screening for chronic hepatitis B and C in migrants from Afghanistan, Iran, Iraq, the former Soviet Republics, and Vietnam in the Arnhem region, the Netherlands. *Epidemiol Infect* 2014;142:2140–2146.
7. **McCarthy AE, Weld LH, Barnett ED, So H, Coyle C, *et al.*** Spectrum of illness in international migrants seen at geosentinel clinics in 1997-2009, part 2: Migrants resettled internationally and evaluated for specific health concerns. *Clinical Infectious Diseases* 2013;56:925–933.

8. **Jazwa A, Coleman MS, Gazmararian J, Wingate LT, Maskery B, *et al.*** Cost–benefit comparison of two proposed overseas programs for reducing chronic Hepatitis B infection among refugees: Is screening essential? *Vaccine* 2015;33:1393–1399.

9. **Alvarado-Mora M V, Botelho L, Gomes-Gouvêa MS, De Souza VF, Nascimento MC, *et al.*** Detection of Hepatitis B virus subgenotype A1 in a Quilombo community from Maranhão, Brazil. *Virol J* 2011;8:415.

10. **Godoy BA, Gomes-Gouvêa MS, Zagonel-Oliveira M, Alvarado-Mora M V., Salzano FM, *et al.*** High prevalence of HBV/A1 subgenotype in native south Americans may be explained by recent economic developments in the Amazon. *Infect Genet Evol* 2016;43:354–363.

11. **Tsachouridou O, Georgiou A, Naoum S, Vasdeki D, Papagianni M, *et al.*** Factors associated with poor adherence to vaccination against hepatitis viruses, *streptococcus pneumoniae* and seasonal influenza in HIV-infected adults. *Hum Vaccin Immunother* 2018;21645515.2018.1509644.

12. **Carvalho PMR dos S, Matos MA de, Martins RMB, Pinheiro RS, Caetano KAA, *et al.*** Prevalence, risk factors and hepatitis B immunization: helping fill the gap on hepatitis B epidemiology among homeless people, Goiânia, Central Brazil. *Cad Saude Publica*;33. Epub ahead of print 2017. DOI: 10.1590/0102-311x00109216.

13. **Dandri M, Locarnini S**. New insight in the pathobiology of hepatitis B virus infection. *Gut* 2012;61:i6–i17.

14. **Guirgis BSS, Abbas RO, Azzazy HME**. Hepatitis B virus genotyping: Current methods and clinical implications. *International Journal of Infectious Diseases* 2010;14:e941–e953.

15. **Norder H, Hammas B, Lofdahl S, Courouce AM, Magnius LO**. Comparison of

the amino acid sequences of nine different serotypes of hepatitis B surface antigen and genomic classification of the corresponding hepatitis B virus strains. *J Gen Virol* 1992;73:1201–1208.

16. **Huy TTT, Ngoc TT, Abe K**. New Complex Recombinant Genotype of Hepatitis B Virus Identified in Vietnam. *J Virol* 2008;82:5657–5663.
17. **Tatematsu K, Tanaka Y, Kurbanov F, Sugauchi F, Mano S, *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.
18. **Kramvis A**. Genotypes and genetic variability of hepatitis B virus. *Intervirology* 2014;57:141–150.
19. **Kramvis A, Arakawa K, Yu MC, Nogueira R, Stram DO, *et al.*** Relationship of serological subtype, basic core promoter and precore mutations to genotypes/subgenotypes of hepatitis B virus. *J Med Virol* 2008;80:27–46.
20. **Sunbul M**. Hepatitis B virus genotypes: Global distribution and clinical importance. *World J Gastroenterol* 2014;20:5427–5434.
21. **Norder H, Couroucé AM, Coursaget P, Echevarria JM, Lee SD, *et al.*** Genetic diversity of hepatitis B virus strains derived worldwide: Genotypes, subgenotypes, and HBsAg subtypes. *Intervirology* 2004;47:289–309.
22. **Mühlemann B, Jones TC, De Barros Damgaard P, Allentoft ME, Shevnina I, *et al.*** Ancient hepatitis B viruses from the Bronze Age to the Medieval period. *Nature* 2018;557:418–423.
23. **Paraskevis D, Magiorkinis G, Magiorkinis E, Ho SYW, Belshaw R, *et al.*** Dating the origin and dispersal of hepatitis B virus infection in humans and primates.

Hepatology 2013;57:908–916.

24. **Yousif M, Kramvis A**. Genotype D of hepatitis B virus and its subgenotypes: An update. *Hepatol Res* 2013;43:355–364.

25. **Norder H, Couroucé A-M, Coursaget P, Echevarria JM, Lee S-D, *et al.*** Genetic Diversity of Hepatitis B Virus Strains Derived Worldwide: Genotypes, Subgenotypes, and HB_sAg Subtypes. *Intervirology* 2004;47:289–309.

26. **Hadziyannis SJ**. Natural history of chronic hepatitis B in Euro-Mediterranean and African Countries. *Journal of Hepatology* 2011;55:183–191.

27. **Alvarado-Mora M V, Rebello Pinho JR**. Epidemiological update of hepatitis B, C and delta in Latin America. *Antiviral Therapy* 2013;18:429–433.

28. **PAHO**. Hepatitis B and C in the Spotlight. http://iris.paho.org/xmlui/handle/123456789/31449 (2016, accessed 21 September 2018).

29. **Finch H, Klein HS, Adelman J, Hu-DeHart E, Díaz LM, *et al.*** Latin and Central American migration. In: Cohen R (editor). *The Cambridge Survey of World Migration*. Cambridge: Cambridge University Press. pp. 203–232.

30. **Salzano FM, Sans M**. Interethnic admixture and the evolution of Latin American populations. *Genetics and Molecular Biology* 2014;37:151–170.

31. **Panduro A, Maldonado-Gonzalez M, Fierro NA, Roman S**. Distribution of HBV genotypes F and H in Mexico and Central America. *Antiviral Therapy* 2013;18:475–484.

32. **Roman S, Jose-Abrego A, Fierro NA, Escobedo-Melendez G, Ojeda-Granados C, *et al.*** Hepatitis B virus infection in Latin America: A genomic medicine approach. *World J Gastroenterol* 2014;20:7181–7196.

33. **Kramvis A, Paraskevis D**. Subgenotype A1 of HBV - Tracing human migrations in and out of Africa. *Antivir Ther* 2013;18:513–521.

34. **Andernach IE, Nolte C, Pape JW, Muller CP**. Slave trade and hepatitis B virus genotypes and subgenotypes in Haiti and Africa. *Emerg Infect Dis* 2009;15:1222–1228.

35. **Lampe E, Mello FCA, do Espírito-Santo MP, Oliveira CMC, Bertolini DA, *et al.*** Nationwide overview of the distribution of hepatitis B virus genotypes in Brazil: A 1000-sample multicentre study. *J Gen Virol* 2017;98:1389–1398.

36. **Pena SDJ, Pietro G Di, Fuchshuber-Moraes M, Genro JP, Hutz MH, *et al.*** The Genomic Ancestry of Individuals from Different Geographical Regions of Brazil Is More Uniform Than Expected. *PLoS One* 2011;6:e17063.

37. **Bertolini DA, Gomes-Gouvêa MS, Carvalho-Mello IMVG de, Saraceni CP, Sitnik R, *et al.*** Hepatitis B virus genotypes from European origin explains the high endemicity found in some areas from southern Brazil. *Infect Genet Evol* 2012;12:1295–1304.

38. **Gusatti CS, Costi C, Halon ML, Grandi T, Medeiros AFR, *et al.*** Hepatitis B virus genotype D isolates circulating in Chapecó, Southern Brazil, originate from Italy. *PLoS One* 2015;10:e0135816.

39. **Chachá SGF, Gomes-Gouvêa MS, Malta F de M, Ferreira S da C, Villanova MG, *et al.*** Distribution of HBV subgenotypes in Ribeirão Preto, Southeastern Brazil: a region with history of intense Italian immigration. *Brazilian J Infect Dis* 2017;21:424–432.

40. **Gomes-Gouvêa MS, Ferreira AC, Teixeira R, Andrade JR, Ferreira AS, *et al.*** HBV carrying drug-resistance mutations in chronically infected treatment-naive

patients. *Antivir Ther* 2015;20:387–395.

41. **Salpini R, Colagrossi L, Bellocchi MC, Surdo M, Becker C, *et al.*** Hepatitis B surface antigen genetic elements critical for immune escape correlate with hepatitis B virus reactivation upon immunosuppression. *Hepatology* 2015;61:823–833.
42. **International Organization for Standardization (ISO).** ISO 3166 Country Codes. *ISO Standards*. http://www.iso.org/iso/country_codes.html (2018, accessed 30 September 2018).
43. **Kumar S, Stecher G, Tamura K**. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol* 2016;33:1870–1874.
44. **Martin DP, Murrell B, Golden M, Khoosal A, Muhire B**. RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evol* 2015;1:vev003.
45. **Stamatakis A**. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 2014;30:1312–1313.
46. **Hartl DL, Clark AG**. *Principles of population genetics*. Sinauer Associates; 2007.
47. **Excoffier L, Smouse PE, Quattro JM**. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 1992;131:479–91.
48. **Excoffier L, Lischer HEL**. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* 2010;10:564–567.
49. **Tamura K, Nei M**. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol* 1993;10:512–26.
50. **Drummond AJ, Rambaut A, Shapiro B, Pybus OG**. Bayesian coalescent inference

of past population dynamics from molecular sequences. *Mol Biol Evol* 2005;22:1185–1192.

51. **Drummond AJ, Rambaut A**. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 2007;7:214.

52. **Godoy BA, Alvarado-Mora M V, Gomes-Gouvêa MS, Pinho JRR, Fagundes NJR**. Origin of HBV and its arrival in the Americas - The importance of natural selection on time estimates. *Antivir Ther* 2013;18:505–512.

53. **Alvarado-Mora M V, Rebello Pinho JR**. Distribution of HBV genotypes in Latin America. *Antiviral Therapy* 2013;18:459–465.

54. **Bertolini DA, Gomes-Gouvêa MS, Carvalho-Mello IMVG de, Saraceni CP, Sitnik R, *et al.*** Hepatitis B virus genotypes from European origin explains the high endemicity found in some areas from southern Brazil. *Infect Genet Evol* 2012;12:1295–1304.

55. **Chachá SGF, Gomes-Gouvêa MS, Malta F de M, Ferreira S da C, Villanova MG, *et al.*** Distribution of HBV subgenotypes in Ribeirão Preto, Southeastern Brazil: a region with history of intense Italian immigration. *Brazilian J Infect Dis* 2017;21:424–432.

56. **IBGE**. Brasil:500 anos de povoamento. *IBGE* 2007;226.

57. **Klein HS**. The Integration of Italian Immigrants into the United States and Argentina: A Comparative Analysis. *Am Hist Rev* 1983;88:306.

58. **Piñeiro y Leone FG, Pezzano SC, Torres C, Rodríguez CE, Eugenia Garay M, *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.

59. **Mojsiejczuk LN, Torres C, Sevic I, Badano I, Malan R, *et al.*** Molecular epidemiology of hepatitis B virus in Misiones, Argentina. *Infect Genet Evol* 2016;44:34–42.

60. **Mello EC**. Imagens do Brasil holandês 1630-1654. *ARS (São Paulo)* 2009;7:160–171.

61. **Boos T**. The Arab Diaspora in Latin America. 2017;10–11.

62. **Central Intelligence Agency Cia, Cia CIA**. Central Intelligence Agency: The World Factbook. *Available online at httpswwwciagovlibrarypublicationstheworldfactbookgeosmlhtml*. https://www.cia.gov/library/publications/the-world-factbook/geos/ha.html (2007, accessed 8 October 2018).

63. **Baha W, Ennaji MM, Lazar F, Melloul M, El Fahime E, *et al.*** HBV genotypes prevalence, precore and basal core mutants in Morocco. *Infect Genet Evol* 2012;12:1157–1162.

64. **Candotti D, Opare-Sem O, Rezvan H, Sarkodie F, Allain J-P**. Molecular and serological characterization of hepatitis B virus in deferred Ghanaian blood donors with and without elevated alanine aminotransferase. *J Viral Hepat* 2006;13:715–724.

65. **Kwange SO, Budambula NLM, Kiptoo MK, Okoth F, Ochwoto M, *et al.*** Hepatitis B virus subgenotype A1, occurrence of subgenotype D4, and S gene mutations among voluntary blood donors in Kenya. *Virus Genes* 2013;47:448–455.

66. **Andernach IE, Nolte C, Pape JW, Muller CP**. Slave trade and hepatitis B virus genotypes and subgenotypes in Haiti and Africa. *Emerg Infect Dis* 2009;15:1222–1228.

67. **Barros LMF, Gomes-Gouvêa MS, Kramvis A, Mendes-Corrêa MCJ, dos Santos**

A, *et al.* High prevalence of hepatitis B virus subgenotypes A1 and D4 in Maranhão state, Northeast Brazil. *Infect Genet Evol* 2014;24:68–75.

68. **Banerjee P, Mondal RK, Nandi M, Ghosh S, Khatun M, *et al.*** A rare HBV subgenotype D4 with unique genomic signatures identified in north-eastern india-an emerging clinical challenge? *PLoS One*;9. Epub ahead of print 2014. DOI: 10.1371/journal.pone.0109425.

Table 1: Subgenotype distributions and frequencies of 1390 HBV sequences belonging to D genotype.

Locations	HBV/D Subgenotype frequencies (%)						
	D1	**D2**	**D3**	**D4**	**D5**	**D7**	**Total**
ARG	10 (30,3)	5 (15,2)	18 (54,5)	-	-	-	33
BR-BA	1 (5,3)	-	11 (57,9)	7 (36,8)	-	-	19
BR-MA	-	1 (1,1)	3 (3,3)	86 (95,6)	-	-	90
BR-S	5 (5,1)	23 (23,5)	69 (70,4)	1 (1,0)	-	-	98
BR-SE	14 (9,5)	21 (14,3)	101 (68,7)	10 (6,8)	-	1 (0,7)	147
CUB	10 (14,7)	-	-	48 (70,6)	-	10 (14,7)	68
DEU+POL	52 (44,1)	38 (32,2)	27 (22,9)	1 (0,8)	-	-	118
ESP+FRA	13 (19,4)	26 (38,8)	19 (28,4)	3 (4,5)	-	6 (8,9)	67
HTI	-	-	5 (13,5)	32 (86,5)	-	-	37
IND	46 (36,0)	14 (10,9)	36 (28,1)	12 (9,4)	20 (15,6)	-	128
ITA	46 (15,7)	49 (16,7)	194 (66,2)	2 (0,7)	-	2 (0,7)	293
LBN+TUR	115 (92,0)	10 (8,0)	-	-	-	-	125
NLD	113 (67,7)	15 (9,0)	18 (10,8)	12 (7,2)	-	9 (5,3)	167
TOTAL	425 (30,7)	202 (14,5)	501 (36,0)	214 (15,4)	20 (1,4)	28 (2,0)	**1390**

Figure legends:

Figure 1: Geographical distribution of the 1390 HBV sequences belonging to D genotype that were included in this study. Subgenotype frequencies are described in Table 1. Colors indicate the different HBV/D subgenotypes (D1-D5, D7).

Figure 2: Maximum likelihood tree estimated for 914 HBV sequences. Genotypes and subgenotypes are indicated as well as the bootstrap values for the major nodes. The sequences belonging to D4, D5 and D7 genotypes were colored according to their geographical regions.

Figure 3: The phylogenetic clades collapsed in Figure 2 are represented as midpoint rooted subtree for HBV/D subgenotypes D1 and D2 (a), and D3 (b). Colors represent different geographical regions of the strains.

Figure 4: Pairwise F_{ST} values between HBV sequences from different locations, according to D genotype as a whole, as well as their subgenotypes. The intensity of color corresponds to the F_{ST} value, indicating genetic differentiation between pairs. The highlight cells contain the lowest calculated value for each analysis and the number is shown in italic. P value after Bonferroni correction is indicated for each analysis and the statistically significant values are shown in bold.

Figure 5: Bayesian Skyline Plots of HBV/D samples and their subgenotypes from Southeastern Brazil in comparison to their most similar locations, according to the F_{ST} analysis.

Figure 1:

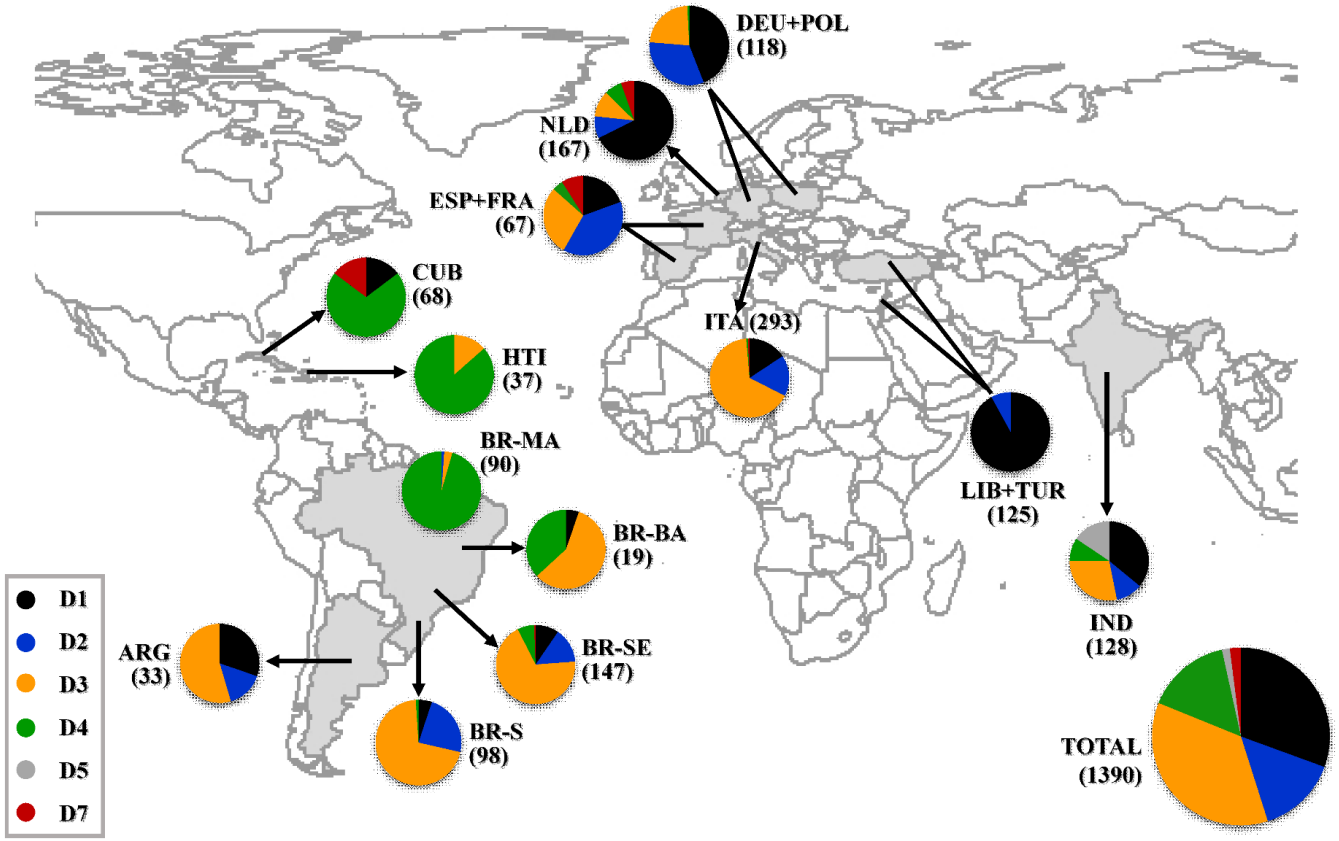

DEU+POL
(118)
NLD
(167)
ESP+FRA
(67)
CUB
(68)
HTI
(37)
ITA (293)
BR-MA
(90)
BR-BA
(19)
LIB+TUR
(125)
IND
(128)
ARG
(33)
BR-SE
(147)
BR-S
(98)
TOTAL
(1390)
D1
D2
D3
D4
D5
D7

Figure 2:

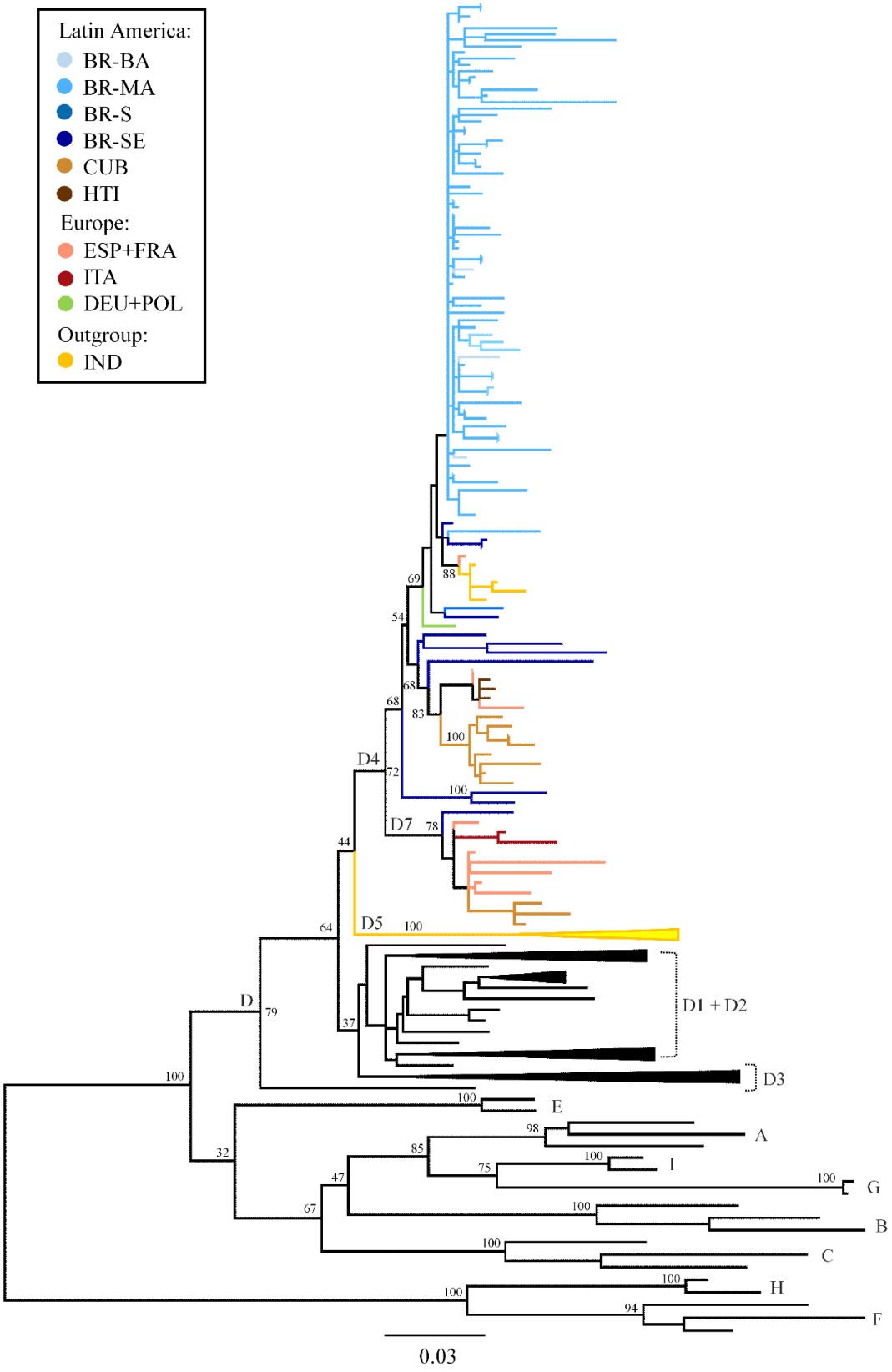

Latin America:
BR-BA
BR-MA
BR-S
BR-SE
CUB
HTI
Europe:
ESP+FRA
ITA
DEU+POL
Outgroup:
IND
D4
D7
D5
D
D1 + D2
D3
E
A
I
G
B
C
H
F
0.03

Figure 3:

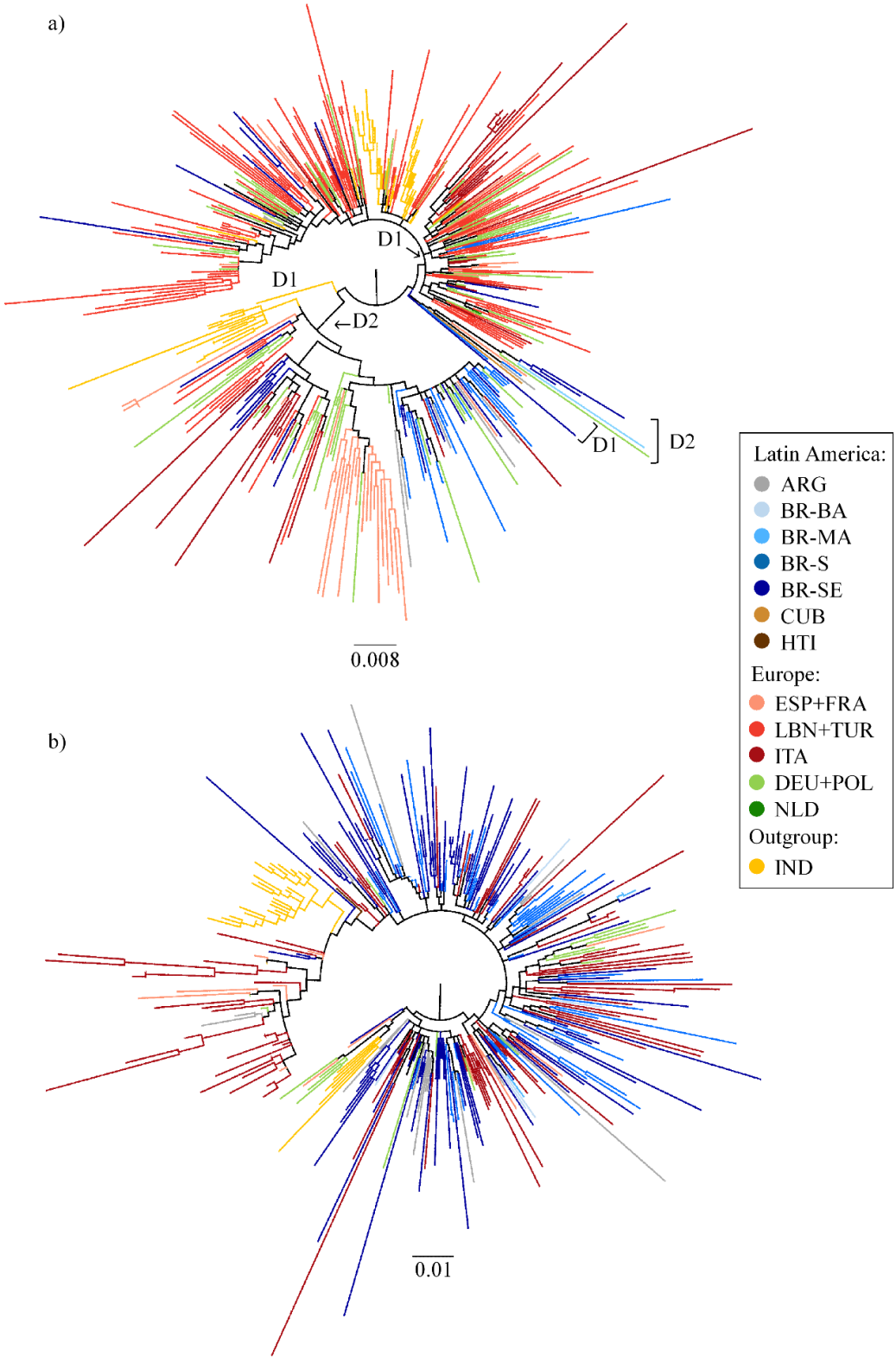

a)
D1
D1
←D2
D1
D2
0.008
b)
0.01
Latin America:
ARG
BR-BA
BR-MA
BR-S
BR-SE
CUB
HTI
Europe:
ESP+FRA
LBN+TUR
ITA
DEU+POL
NLD
Outgroup:
IND

Figure 4:

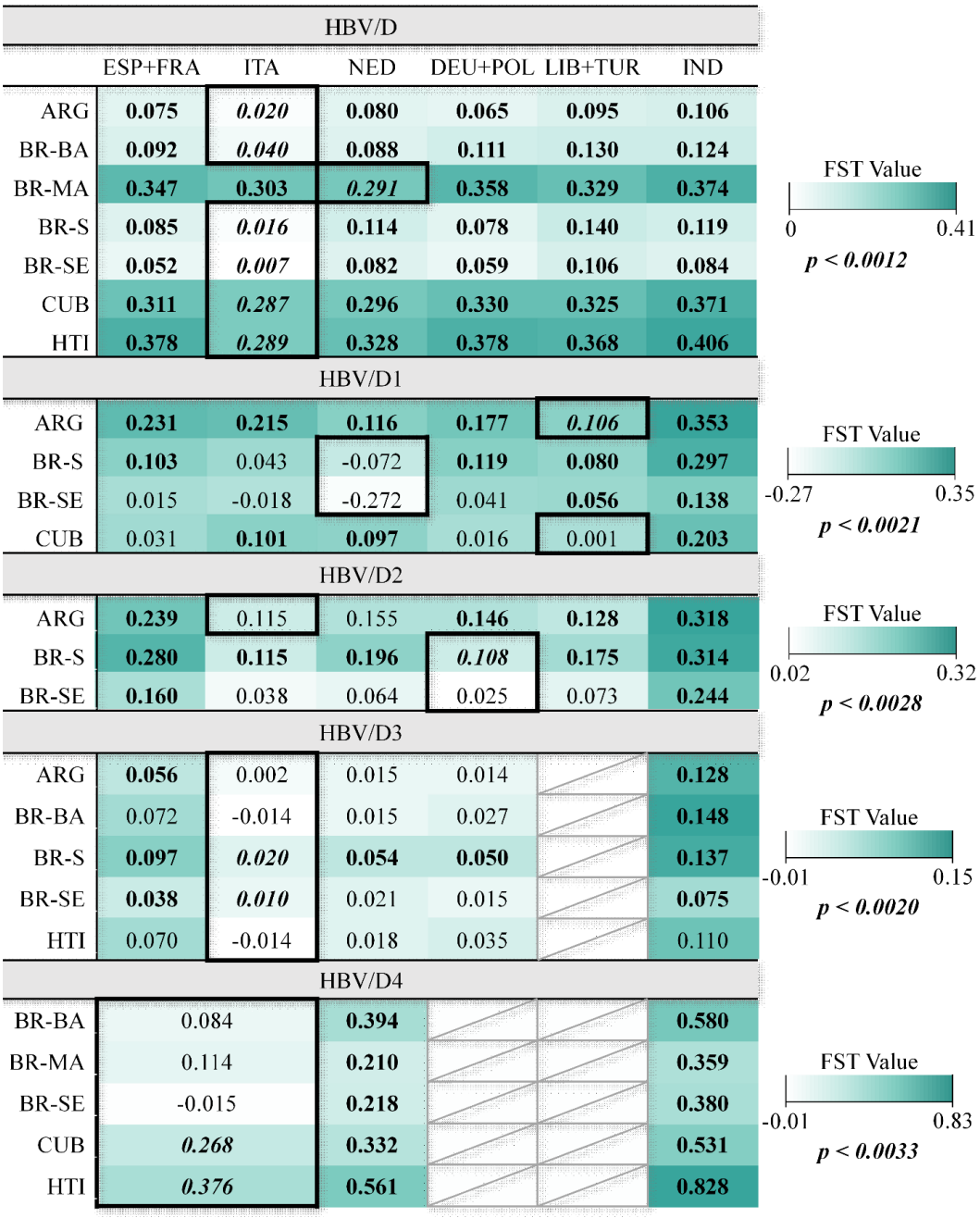
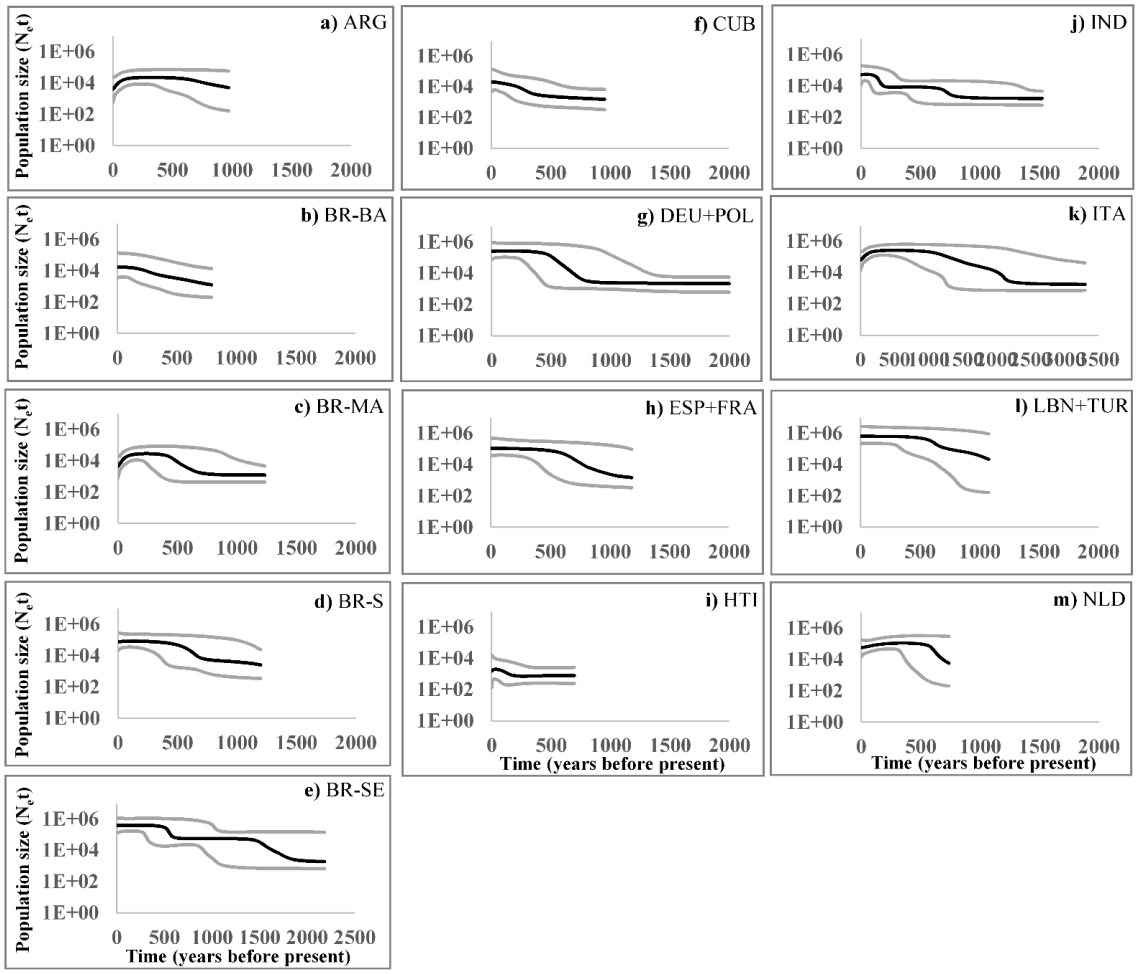

HBV/D
ESP+FRA ITA NED DEU+POL LIB+TUR IND
ARG 0.075 0.020 0.080 0.065 0.095 0.106
BR-BA 0.092 0.040 0.088 0.111 0.130 0.124
BR-MA 0.347 0.303 0.291 0.358 0.329 0.374
BR-S 0.085 0.016 0.114 0.078 0.140 0.119
BR-SE 0.052 0.007 0.082 0.059 0.106 0.084
CUB 0.311 0.287 0.296 0.330 0.325 0.371
HTI 0.378 0.289 0.328 0.378 0.368 0.406
FST Value
0 0.41
p < 0.0012
HBV/D1
ARG 0.231 0.215 0.116 0.177 0.106 0.353
BR-S 0.103 0.043 -0.072 0.119 0.080 0.297
BR-SE 0.015 -0.018 -0.272 0.041 0.056 0.138
CUB 0.031 0.101 0.097 0.016 0.001 0.203
FST Value
-0.27 0.35
p < 0.0021
HBV/D2
ARG 0.239 0.115 0.155 0.146 0.128 0.318
BR-S 0.280 0.115 0.196 0.108 0.175 0.314
BR-SE 0.160 0.038 0.064 0.025 0.073 0.244
FST Value
0.02 0.32
p < 0.0028
HBV/D3
ARG 0.056 0.002 0.015 0.014 0.128
BR-BA 0.072 -0.014 0.015 0.027 0.148
BR-S 0.097 0.020 0.054 0.050 0.137
BR-SE 0.038 0.010 0.021 0.015 0.075
HTI 0.070 -0.014 0.018 0.035 0.110
FST Value
-0.01 0.15
p < 0.0020
HBV/D4
BR-BA 0.084 0.394 0.580
BR-MA 0.114 0.210 0.359
BR-SE -0.015 0.218 0.380
CUB 0.268 0.332 0.531
HTI 0.376 0.561 0.828
FST Value
-0.01 0.83
p < 0.0033

Figure 5:


a) ARG
b) BR-BA
c) BR-MA
d) BR-S
e) BR-SE
f) CUB
g) DEU+POL
h) ESP+FRA
i) HTI
j) IND
k) ITA
l) LBN+TUR
m) NLD
Population size (N_et)
1E+06
1E+04
1E+02
1E+00
0 500 1000 1500 2000
0 500 1000 1500 2000 2500
Time (years before present)

Supplementary Material

Table S1: HBV sequences included in this study. All HBV/D samples were used in the population analyzes, and those samples included in the Maximum Likelihood tree are indicated as well as the reference sequences to the other HBV genotypes. Haplotypes represents equal sequences represented in the phylogeny.

ID	Accession Number	Population	Subgenotype	Phylogeny
_	KF584158	ARG	D1	✓
_	KF584159	ARG	D1	✓
_	KF584160	ARG	D1	✓
_	KF584161	ARG	D1	
_	KF584162	ARG	D1	✓ (D1_Hap1_ARG)
_	KF584163	ARG	D1	✓ (D1_Hap1_ARG)
_	KF584164	ARG	D1	✓ (D1_Hap1_ARG)
_	KF584165	ARG	D1	✓ (D1_Hap1_ARG)
_	KF584166	ARG	D1	✓ (D1_Hap1_ARG)
_	JN688695	ARG	D1	✓
_	KJ647350	ARG	D2	✓
_	KJ647352	ARG	D2	✓
_	KJ647356	ARG	D2	✓
_	JN688717	ARG	D2	✓
_	KJ647351	ARG	D2	✓
_	KJ647349	ARG	D3	✓
_	KJ647353	ARG	D3	✓
_	KJ647355	ARG	D3	✓
_	KJ843187	ARG	D3	✓
_	JN688678	ARG	D3	✓
_	JN688679	ARG	D3	✓
_	JN688683	ARG	D3	✓
_	JN688685	ARG	D3	✓
_	JN688689	ARG	D3	✓
_	JN688708	ARG	D3	✓
_	JN688710	ARG	D3	✓
_	JN688711	ARG	D3	✓
_	JN688712	ARG	D3	✓
_	JN688713	ARG	D3	✓
_	JN688715	ARG	D3	✓
_	JN688716	ARG	D3	✓
_	JN688722	ARG	D3	✓
_	KC012652	ARG	D3	✓

_	KU847509	BR_BA	D1	
_	KU847510	BR_BA	D3	✓
_	KU847513	BR_BA	D3	
_	KU847519	BR_BA	D3	✓
_	KU847532	BR_BA	D3	✓
_	KU847561	BR_BA	D3	✓
_	KU847571	BR_BA	D3	
_	KU847472	BR_BA	D3	
_	KU847574	BR_BA	D3	
_	KU847579	BR_BA	D3	
_	KU847592	BR_BA	D3	✓
_	KU847593	BR_BA	D3	✓
_	KU847488	BR_BA	D4	✓
_	KU847493	BR_BA	D4	
_	KU847535	BR_BA	D4	✓
_	KU847552	BR_BA	D4	✓
_	KU847491	BR_BA	D4	✓ (D4_Hap1_BA_BRA)
_	KU847518	BR_BA	D4	✓ (D4_Hap1_BA_BRA)
_	KU847521	BR_BA	D4	✓ (D4_Hap1_BA_BRA)
8104	*	BR_MA	D2	
8052	*	BR_MA	D3	✓
8075	*	BR_MA	D3	✓
8096	*	BR_MA	D3	✓
8003	*	BR_MA	D4	✓
8016	*	BR_MA	D4	✓
8022	*	BR_MA	D4	✓
8030	*	BR_MA	D4	✓
8032	*	BR_MA	D4	✓
8033	*	BR_MA	D4	✓ (D4_Hap1_MA_BRA)
8034	*	BR_MA	D4	✓ (D4_Hap1_MA_BRA)
8035	*	BR_MA	D4	✓ (D4_Hap1_MA_BRA)
8037	*	BR_MA	D4	✓
8041	*	BR_MA	D4	✓
8042	*	BR_MA	D4	✓
8064	*	BR_MA	D4	✓
8066	*	BR_MA	D4	✓
8079	*	BR_MA	D4	✓
8103	*	BR_MA	D4	✓
8111	*	BR_MA	D4	✓
8117	*	BR_MA	D4	✓ (D4_Hap2_MA_BRA)
8118	*	BR_MA	D4	✓ (D4_Hap3_MA_BRA)

8120	*	BR_MA	D4	✓
8122	*	BR_MA	D4	✓
8127	*	BR_MA	D4	✓
8134	*	BR_MA	D4	
8136	*	BR_MA	D4	✓
8142	*	BR_MA	D4	✓
8143	*	BR_MA	D4	✓
8144	*	BR_MA	D4	✓
8145	*	BR_MA	D4	✓ (D4_Hap4_MA_BRA)
8146	*	BR_MA	D4	✓ (D4_Hap4_MA_BRA)
8149	*	BR_MA	D4	✓ (D4_Hap4_MA_BRA)
_	KJ470884	BR_MA	D4	✓ (D4_Hap4_MA_BRA)
_	KJ470885	BR_MA	D4	✓
_	KJ470886	BR_MA	D4	✓ (D4_Hap2_MA_BRA)
_	KJ470887	BR_MA	D4	✓
_	KJ470888	BR_MA	D4	✓ (D4_Hap1_MA_BRA)
_	KJ470889	BR_MA	D4	✓
_	KJ470890	BR_MA	D4	✓ (D4_Hap1_MA_BRA)
_	KJ470891	BR_MA	D4	✓
_	KJ470892	BR_MA	D4	✓
_	KJ470893	BR_MA	D4	✓ (D4_Hap3_MA_BRA)
_	KJ470894	BR_MA	D4	✓
_	KJ470895	BR_MA	D4	✓
_	KJ470896	BR_MA	D4	✓
_	KJ470897	BR_MA	D4	✓
_	KJ470898	BR_MA	D4	✓
_	KX302085	BR_MA	D4	✓
_	KX302086	BR_MA	D4	✓ (D4_Hap5_MA_BRA)
_	KX302087	BR_MA	D4	✓
_	KX302088	BR_MA	D4	✓
_	KX302089	BR_MA	D4	✓
_	KX302090	BR_MA	D4	✓
_	KX302091	BR_MA	D4	✓
_	KX302092	BR_MA	D4	✓
_	KX302093	BR_MA	D4	✓
_	KX302094	BR_MA	D4	✓ (D4_Hap5_MA_BRA)
_	KX302096	BR_MA	D4	✓
_	KX302097	BR_MA	D4	✓
_	KX302098	BR_MA	D4	✓ (D4_Hap6_MA_BRA)
_	KX302099	BR_MA	D4	✓
_	KX302100	BR_MA	D4	✓

_	KX302101	BR_MA	D4	✓
_	KX302102	BR_MA	D4	✓
_	KX302103	BR_MA	D4	✓
_	KX302104	BR_MA	D4	✓ (D4_Hap6_MA_BRA)
_	KX302107	BR_MA	D4	✓
_	KX302108	BR_MA	D4	✓
_	KX302109	BR_MA	D4	✓
_	KX302110	BR_MA	D4	✓ (D4_Hap6_MA_BRA)
_	KX302111	BR_MA	D4	✓
_	KX302112	BR_MA	D4	✓
_	KX302114	BR_MA	D4	✓
_	KX302115	BR_MA	D4	✓
_	KX302116	BR_MA	D4	✓ (D4_Hap5_MA_BRA)
_	KX302117	BR_MA	D4	✓
_	KX302118	BR_MA	D4	✓
_	KX302120	BR_MA	D4	✓
_	KX302121	BR_MA	D4	✓
_	KX302122	BR_MA	D4	✓ (D4_Hap6_MA_BRA)
_	KX302123	BR_MA	D4	✓
_	KX302124	BR_MA	D4	✓
_	KX302125	BR_MA	D4	✓
_	KX302126	BR_MA	D4	✓ (D4_Hap6_MA_BRA)
_	KX302127	BR_MA	D4	✓
_	KX302128	BR_MA	D4	✓
_	KX302131	BR_MA	D4	✓
_	KX302132	BR_MA	D4	✓
_	KX302119	BR_MA	D4	✓
RS_45	*	BR_S	D1	✓
RS_23	*	BR_S	D1	✓
_	EF690480	BR_S	D1	
_	JF815636	BR_S	D1	✓
_	JF815642	BR_S	D1	✓
RS_1	*	BR_S	D2	✓
RS_2	*	BR_S	D2	✓
RS_6	*	BR_S	D2	✓
RS_9	*	BR_S	D2	✓
RS_13	*	BR_S	D2	✓
RS_17	*	BR_S	D2	✓
RS_40	*	BR_S	D2	✓
RS_43	*	BR_S	D2	✓
RS_48	*	BR_S	D2	✓

RS_54	*	BR_S	D2	✓
RS_61	*	BR_S	D2	✓
_	EF690473	BR_S	D2	
_	EF690472	BR_S	D2	
_	JF815613	BR_S	D2	✓
_	JF815618	BR_S	D2	✓
_	JF815623	BR_S	D2	✓
_	JF815631	BR_S	D2	✓
_	JF815639	BR_S	D2	✓
_	JF815643	BR_S	D2	✓
_	JF815650	BR_S	D2	✓
_	JF815652	BR_S	D2	✓
_	JF815654	BR_S	D2	✓
_	JF815657	BR_S	D2	✓
RS_3	*	BR_S	D3	✓
RS_8	*	BR_S	D3	✓
RS_14	*	BR_S	D3	✓
RS_15	*	BR_S	D3	✓
RS_16	*	BR_S	D3	✓
RS_18	*	BR_S	D3	✓
RS_22	*	BR_S	D3	✓
RS_24	*	BR_S	D3	✓
RS_25	*	BR_S	D3	✓
RS_26	*	BR_S	D3	✓
RS_27	*	BR_S	D3	✓
RS_29	*	BR_S	D3	✓
RS_30	*	BR_S	D3	✓
RS_37	*	BR_S	D3	✓
RS_38	*	BR_S	D3	✓
RS_41	*	BR_S	D3	✓
RS_42	*	BR_S	D3	✓
RS_44	*	BR_S	D3	✓
RS_46	*	BR_S	D3	✓
RS_50	*	BR_S	D3	✓
RS_51	*	BR_S	D3	✓
RS_52	*	BR_S	D3	✓
RS_56	*	BR_S	D3	✓
RS_57	*	BR_S	D3	✓
RS_59	*	BR_S	D3	✓
RS_60	*	BR_S	D3	✓
RS_62	*	BR_S	D3	✓

RS_64	*	BR_S	D3	✓
_	EF690478	BR_S	D3	
_	JF815606	BR_S	D3	✓
_	JF815607	BR_S	D3	✓
_	JF815608	BR_S	D3	✓
_	JF815609	BR_S	D3	✓
_	JF815610	BR_S	D3	✓
_	JF815611	BR_S	D3	
_	JF815612	BR_S	D3	✓
_	JF815614	BR_S	D3	
_	JF815615	BR_S	D3	✓
_	JF815616	BR_S	D3	✓
_	JF815617	BR_S	D3	✓
_	JF815619	BR_S	D3	
_	JF815620	BR_S	D3	✓
_	JF815621	BR_S	D3	✓
_	JF815622	BR_S	D3	✓
_	JF815624	BR_S	D3	✓
_	JF815625	BR_S	D3	✓
_	JF815626	BR_S	D3	✓
_	JF815627	BR_S	D3	✓
_	JF815628	BR_S	D3	✓
_	JF815630	BR_S	D3	✓
_	JF815632	BR_S	D3	✓
_	JF815633	BR_S	D3	✓
_	JF815634	BR_S	D3	✓
_	JF815635	BR_S	D3	✓
_	JF815637	BR_S	D3	✓
_	JF815640	BR_S	D3	✓
_	JF815645	BR_S	D3	✓
_	JF815646	BR_S	D3	✓
_	JF815649	BR_S	D3	✓
_	JF815651	BR_S	D3	✓
_	JF815653	BR_S	D3	✓
_	JF815655	BR_S	D3	✓
_	JF815656	BR_S	D3	✓
_	JF815658	BR_S	D3	✓ (D3_Hap2_S_BRA)
_	JF815659	BR_S	D3	✓
_	JF815660	BR_S	D3	✓
_	JF815661	BR_S	D3	✓
_	JF815662	BR_S	D3	✓

_	JF815663	BR_S	D3	✓ (D3_Hap2_S_BRA)
_	JF815648	BR_S	D4	✓
RP_16	*	BR_SE	D1	✓
RP_67	*	BR_SE	D1	✓
RP_91	*	BR_SE	D1	✓
RP_101	*	BR_SE	D1	✓
RP_172	*	BR_SE	D1	✓
RP_178	*	BR_SE	D1	✓
RP_TN_19	*	BR_SE	D1	✓
RP_2	*	BR_SE	D1	✓
RP_184	*	BR_SE	D1	✓
RP_TN_29	*	BR_SE	D1	✓
SP_45	*	BR_SE	D2	
SP_47	*	BR_SE	D2	
SP_14	*	BR_SE	D2	
RP_13	*	BR_SE	D2	✓
RP_76	*	BR_SE	D2	✓
RP_84	*	BR_SE	D2	✓
RP_93	*	BR_SE	D2	✓
RP_108	*	BR_SE	D2	✓
RP_118	*	BR_SE	D2	✓
RP_153	*	BR_SE	D2	✓
RP_167	*	BR_SE	D2	✓
RP_170	*	BR_SE	D2	✓
RP_177	*	BR_SE	D2	✓
RP_180	*	BR_SE	D2	✓
RP_195	*	BR_SE	D2	✓
RP_191	*	BR_SE	D2	✓
SP_37	*	BR_SE	D2	✓
SP_94	*	BR_SE	D2	✓
SP_103	*	BR_SE	D2	✓
SP_120	*	BR_SE	D2	✓
SP_125	*	BR_SE	D2	✓
RP_TN_71	*	BR_SE	D2	✓
BH_50	*	BR_SE	D2	✓
BH_382	*	BR_SE	D2	✓
RP_3	*	BR_SE	D3	✓
RP_5	*	BR_SE	D3	✓
RP_6	*	BR_SE	D3	✓
RP_8	*	BR_SE	D3	✓
RP_9	*	BR_SE	D3	✓

RP_14	*	BR_SE	D3	✓
RP_15	*	BR_SE	D3	✓
RP_19	*	BR_SE	D3	✓
RP_22	*	BR_SE	D3	✓
RP_25	*	BR_SE	D3	✓
RP_26	*	BR_SE	D3	✓
RP_29	*	BR_SE	D3	✓
RP_30	*	BR_SE	D3	✓
RP_31	*	BR_SE	D3	✓
RP_32	*	BR_SE	D3	✓
RP_36	*	BR_SE	D3	✓
RP_37	*	BR_SE	D3	✓
RP_40	*	BR_SE	D3	✓
RP_41	*	BR_SE	D3	✓
RP_42	*	BR_SE	D3	✓
RP_43	*	BR_SE	D3	✓
RP_46	*	BR_SE	D3	✓
RP_51	*	BR_SE	D3	✓
RP_54	*	BR_SE	D3	✓
RP_55	*	BR_SE	D3	✓ (D3_Hap1_SE_BRA)
RP_56	*	BR_SE	D3	✓
RP_57	*	BR_SE	D3	✓
RP_58	*	BR_SE	D3	✓
RP_59	*	BR_SE	D3	✓
RP_61	*	BR_SE	D3	✓
RP_62	*	BR_SE	D3	✓
RP_65	*	BR_SE	D3	✓
RP_66	*	BR_SE	D3	✓
RP_70	*	BR_SE	D3	✓
RP_71	*	BR_SE	D3	✓
RP_72	*	BR_SE	D3	✓
RP_77	*	BR_SE	D3	✓
RP_79	*	BR_SE	D3	✓
RP_81	*	BR_SE	D3	✓
RP_94	*	BR_SE	D3	✓
RP_99	*	BR_SE	D3	✓
RP_110	*	BR_SE	D3	✓
RP_112	*	BR_SE	D3	✓
RP_120	*	BR_SE	D3	✓
RP_135	*	BR_SE	D3	✓
RP_138	*	BR_SE	D3	✓

RP_140	*	BR_SE	D3	✓
RP_141	*	BR_SE	D3	✓
RP_143	*	BR_SE	D3	✓
RP_146	*	BR_SE	D3	✓
RP_147	*	BR_SE	D3	✓
RP_149	*	BR_SE	D3	✓
RP_151	*	BR_SE	D3	✓
RP_156	*	BR_SE	D3	✓ (D3_Hap1_SE_BRA)
RP_158	*	BR_SE	D3	✓
RP_160	*	BR_SE	D3	✓
RP_162	*	BR_SE	D3	✓
RP_164	*	BR_SE	D3	✓
RP_168	*	BR_SE	D3	✓
RP_171	*	BR_SE	D3	
RP_173	*	BR_SE	D3	✓
RP_179	*	BR_SE	D3	✓
RP_187	*	BR_SE	D3	✓
RP_192	*	BR_SE	D3	✓
SP_7	*	BR_SE	D3	✓
SP_9	*	BR_SE	D3	✓
SP_10	*	BR_SE	D3	✓
SP_17	*	BR_SE	D3	✓
SP_23	*	BR_SE	D3	✓
SP_25	*	BR_SE	D3	✓
SP_27	*	BR_SE	D3	✓
SP_29	*	BR_SE	D3	✓
SP_31	*	BR_SE	D3	✓
SP_40	*	BR_SE	D3	✓
SP_43	*	BR_SE	D3	✓
SP_44	*	BR_SE	D3	✓
SP_46	*	BR_SE	D3	✓
SP_59	*	BR_SE	D3	✓
SP_63	*	BR_SE	D3	✓
SP_66	*	BR_SE	D3	✓
SP_73	*	BR_SE	D3	✓
SP_77	*	BR_SE	D3	✓
SP_93	*	BR_SE	D3	✓
SP_101	*	BR_SE	D3	✓ (D3_Hap2_SE_BRA)
SP_104	*	BR_SE	D3	✓ (D3_Hap2_SE_BRA)
SP_106	*	BR_SE	D3	✓ (D3_Hap2_SE_BRA)
RP_TN_1	*	BR_SE	D3	✓

RP_TN_16	*	BR_SE	D3	✓
RP_TN_18	*	BR_SE	D3	✓
RP_TN_24	*	BR_SE	D3	✓
RP_TN_40	*	BR_SE	D3	✓
RP_TN_41	*	BR_SE	D3	✓
RP_TN_49	*	BR_SE	D3	✓
RP_TN_54	*	BR_SE	D3	✓
RP_TN_55	*	BR_SE	D3	✓
RP_TN_72	*	BR_SE	D3	✓
BH_135	*	BR_SE	D3	✓
BH_202	*	BR_SE	D3	✓
BH_346	*	BR_SE	D3	✓
_	EF690475	BR_SE	D3	
_	EF690477	BR_SE	D3	
RP_74	*	BR_SE	D4	✓
RP_119	*	BR_SE	D4	✓
RP_131	*	BR_SE	D4	✓
RP_190	*	BR_SE	D4	✓
SP_122	*	BR_SE	D4	✓
SP_33	*	BR_SE	D4	✓
SP_53	*	BR_SE	D4	✓
RP_TN_19	*	BR_SE	D4	✓
BH_34	*	BR_SE	D4	✓
BH_176	*	BR_SE	D4	✓
RP_88	*	BR_SE	D8	✓
_	KM606702	CUB	D1	✓
_	KM606753	CUB	D1	✓
_	KP165769	CUB	D1	
_	KP165689	CUB	D1	
_	KP165686	CUB	D1	
_	KP165636	CUB	D1	
_	KM606747	CUB	D1	✓
_	KM606898	CUB	D1	
_	KM606932	CUB	D1	
_	KM606853	CUB	D1	
_	KM606744	CUB	D4	✓
_	KM606740	CUB	D4	✓
_	KM606709	CUB	D4	✓
_	KM606712	CUB	D4	✓
_	KM606752	CUB	D4	✓
_	KM606754	CUB	D4	✓

_	KM606755	CUB	D4	✓
_	KM606714	CUB	D4	✓
_	KM606968	CUB	D4	
_	KM606967	CUB	D4	
_	KM606958	CUB	D4	
_	KM606959	CUB	D4	
_	KM606960	CUB	D4	
_	KM606961	CUB	D4	
_	KM606955	CUB	D4	
_	KM606953	CUB	D4	
_	KM606951	CUB	D4	
_	KM606944	CUB	D4	
_	KM606935	CUB	D4	
_	KM606931	CUB	D4	
_	KM606922	CUB	D4	
_	KM606902	CUB	D4	
_	KM606884	CUB	D4	
_	KM606883	CUB	D4	
_	KM606867	CUB	D4	
_	KM606861	CUB	D4	
_	KM606851	CUB	D4	
_	KM606848	CUB	D4	
_	KM606763	CUB	D4	
_	KM606842	CUB	D4	
_	KM606840	CUB	D4	
_	KM606839	CUB	D4	
_	KM606838	CUB	D4	
_	KM606837	CUB	D4	
_	KM606836	CUB	D4	
_	KM606835	CUB	D4	
_	KM606824	CUB	D4	
_	KM606823	CUB	D4	
_	KM606818	CUB	D4	
_	KM606812	CUB	D4	
_	KM606801	CUB	D4	
_	KM606796	CUB	D4	
_	KM606790	CUB	D4	
_	KM606776	CUB	D4	
_	KP165816	CUB	D4	
_	KP165792	CUB	D4	
_	KP165787	CUB	D4	

_	KP165746	CUB	D4	
_	KM606741	CUB	D7	✓
_	KM606750	CUB	D7	✓
_	KM606751	CUB	D7	✓
_	KM606897	CUB	D7	
_	KM606892	CUB	D7	
_	KP165798	CUB	D7	
_	KP165743	CUB	D7	
_	KP165725	CUB	D7	
_	KP165713	CUB	D7	
_	KP165622	CUB	D7	
_	AF208872	DEU	D1	
_	AF209391	DEU	D1	
_	AF209398	DEU	D1	
_	GQ486045	DEU	D1	✓
_	GQ486049	DEU	D1	✓
_	GQ486059	DEU	D1	✓
_	GQ486109	DEU	D1	✓
_	GQ486704	DEU	D1	✓
_	GQ486729	DEU	D1	✓
_	GQ486706	DEU	D1	✓
_	GQ486709	DEU	D1	✓
_	GQ486110	DEU	D1	✓
_	GQ486167	DEU	D1	✓
_	GQ486248	DEU	D1	✓
_	GQ486251	DEU	D1	✓
_	GQ486274	DEU	D1	✓
_	GQ486316	DEU	D1	✓
_	GQ486352	DEU	D1	✓
_	GQ486376	DEU	D1	✓
_	GQ486386	DEU	D1	✓
_	GQ486434	DEU	D1	✓
_	GQ486446	DEU	D1	✓
_	GQ486502	DEU	D1	✓
_	GQ486504	DEU	D1	✓
_	GQ486545	DEU	D1	✓
_	GQ486551	DEU	D1	✓
_	GQ486554	DEU	D1	✓
_	GQ486662	DEU	D1	✓
_	GQ486667	DEU	D1	✓
_	GQ486670	DEU	D1	✓

_	GQ486680	DEU	D1	✓
_	GQ486715	DEU	D1	✓
_	GQ486671	DEU	D1	✓
_	GQ486672	DEU	D1	✓
_	GQ486673	DEU	D1	✓
_	GQ486674	DEU	D1	✓
_	GQ486675	DEU	D1	✓
_	GQ486682	DEU	D1	✓
_	GQ486696	DEU	D1	✓
_	GQ486711	DEU	D1	✓
_	GQ486716	DEU	D1	✓
_	GQ486730	DEU	D1	✓
_	GQ486731	DEU	D1	✓
_	GQ486735	DEU	D1	✓
_	GQ486738	DEU	D1	✓
_	GQ486768	DEU	D1	✓
_	DQ399006	DEU	D2	
_	GQ486717	DEU	D2	✓
_	AF065110	DEU	D2	
_	AF065117	DEU	D2	
_	AF065119	DEU	D2	
_	AF208870	DEU	D2	
_	AF208874	DEU	D2	
_	AF209401	DEU	D2	
_	DQ459085	DEU	D2	
_	GQ486067	DEU	D2	✓
_	GQ486157	DEU	D2	✓
_	GQ486212	DEU	D2	✓
_	GQ486286	DEU	D2	✓
_	GQ486294	DEU	D2	✓
_	GQ486321	DEU	D2	✓
_	GQ486428	DEU	D2	✓
_	GQ486681	DEU	D2	✓
_	GQ486705	DEU	D2	✓
_	GQ486708	DEU	D2	✓
_	GQ486767	DEU	D2	✓
_	X72702	DEU	D2	✓
_	AF061523	DEU	D3	
_	AF065112	DEU	D3	
_	AF065118	DEU	D3	
_	AF209394	DEU	D3	

_	AF209396	DEU	D3	
_	AJ131956	DEU	D3	✓
_	AF208877	DEU	D3	
_	GQ486039	DEU	D3	✓
_	GQ486136	DEU	D3	✓
_	GQ486257	DEU	D3	✓
_	GQ486323	DEU	D3	✓
_	GQ486344	DEU	D3	✓
_	GQ486355	DEU	D3	✓
_	GQ486405	DEU	D3	✓
_	GQ486517	DEU	D3	✓
_	GQ486698	DEU	D3	✓
_	GQ486732	DEU	D3	✓
_	GQ486769	DEU	D3	✓
_	GQ486604	DEU	D4	✓
_	GQ477458	POL	D1	✓
_	GQ477459	POL	D1	✓
_	GQ477505	POL	D1	
_	GQ477513	POL	D1	
_	GQ486261	POL	D1	✓
_	GQ477506	POL	D2	
_	GQ477507	POL	D2	
_	GQ477508	POL	D2	
_	GQ477510	POL	D2	
_	GQ477512	POL	D2	
_	GQ477452	POL	D2	✓
_	GQ477453	POL	D2	✓
_	GQ477454	POL	D2	✓
_	GQ477455	POL	D2	✓
_	GQ477456	POL	D2	✓
_	GQ477457	POL	D2	✓
_	GQ486075	POL	D2	✓
_	GQ486262	POL	D2	✓
_	GQ486302	POL	D2	✓
_	Z35716	POL	D2	✓
_	GQ486473	POL	D2	✓
_	GQ486488	POL	D2	✓
_	GQ486525	POL	D2	✓
_	GQ486073	POL	D3	✓
_	GQ486122	POL	D3	✓
_	GQ486144	POL	D3	✓

_	GQ486214	POL	D3	✓
_	GQ486338	POL	D3	✓
_	GQ477511	POL	D3	
_	GQ477509	POL	D3	
_	GQ486585	POL	D3	✓
_	GQ486586	POL	D3	✓
_	AJ627224	ESP	D1	✓
_	GQ486191	ESP	D1	✓
_	GQ486771	ESP	D1	✓
_	AJ627215	ESP	D2	✓
_	AJ627216	ESP	D2	✓
_	AJ627218	ESP	D2	✓
_	AJ627220	ESP	D2	✓
_	AJ627222	ESP	D2	✓
_	AJ627223	ESP	D2	✓
_	AY090452	ESP	D2	✓
_	GQ486060	ESP	D2	✓
_	GQ486125	ESP	D2	✓
_	GQ486137	ESP	D2	✓
_	GQ486184	ESP	D2	✓
_	GQ486193	ESP	D2	✓
_	GQ486199	ESP	D2	✓
_	GQ486200	ESP	D2	✓
_	GQ486206	ESP	D2	✓
_	GQ486207	ESP	D2	✓
_	GQ486303	ESP	D2	✓
_	GQ486313	ESP	D2	✓
_	GQ486324	ESP	D2	✓
_	GQ486325	ESP	D2	✓
_	GQ486755	ESP	D2	✓
_	GQ486785	ESP	D2	✓
_	GQ486827	ESP	D2	✓
_	GQ486846	ESP	D2	✓
_	AJ627217	ESP	D3	✓
_	AJ627221	ESP	D3	✓
_	GQ486088	ESP	D3	✓
_	GQ486266	ESP	D3	✓
_	GQ486322	ESP	D3	✓
_	GQ486678	ESP	D3	✓
_	AJ627219	ESP	D4	✓
_	GQ486596	ESP	D4	✓

_	GQ486842	ESP	D7	✓
_	AJ344116	FRA	D1	✓
_	AJ517792	FRA	D1	
_	GQ486085	FRA	D1	✓
_	GQ486145	FRA	D1	✓
_	GQ486403	FRA	D1	✓
_	GQ486414	FRA	D1	✓
_	GQ486602	FRA	D1	✓
_	GQ486676	FRA	D1	✓
_	GQ486700	FRA	D1	✓
_	GQ486719	FRA	D1	✓
_	GQ486489	FRA	D2	✓
_	KY697841	FRA	D2	
_	AJ344117	FRA	D3	✓
_	AM422939	FRA	D3	✓
_	AY738864	FRA	D3	
_	AY738878	FRA	D3	
_	AY738880	FRA	D3	
_	AY738881	FRA	D3	
_	GQ486722	FRA	D3	✓
_	GQ486726	FRA	D3	✓
_	GQ486250	FRA	D3	✓
_	GQ486759	FRA	D3	✓
_	GQ486297	FRA	D3	✓
_	GQ486298	FRA	D3	✓
_	V01460	FRA	D3	✓
_	GQ486699	FRA	D4	✓
_	GQ486263	FRA	D7	✓
_	GQ486404	FRA	D7	✓
_	GQ486420	FRA	D7	✓
_	GQ486746	FRA	D7	✓
_	GQ486748	FRA	D7	✓
_	FJ692502	HTI	D3	
_	FJ692505	HTI	D3	
_	FJ692506	HTI	D3	✓ (D3_Hap1_HTI)
_	FJ692507	HTI	D3	✓ (D3_Hap1_HTI)
_	FJ692504	HTI	D3	
_	FJ692539	HTI	D4	
_	FJ692508	HTI	D4	
_	FJ692509	HTI	D4	
_	FJ692510	HTI	D4	

_	FJ692511	HTI	D4	
_	FJ692512	HTI	D4	
_	FJ692513	HTI	D4	
_	FJ692514	HTI	D4	
_	FJ692515	HTI	D4	
_	FJ692516	HTI	D4	
_	FJ692517	HTI	D4	
_	FJ692518	HTI	D4	
_	FJ692519	HTI	D4	
_	FJ692520	HTI	D4	
_	FJ692521	HTI	D4	
_	FJ692522	HTI	D4	
_	FJ692523	HTI	D4	
_	FJ692524	HTI	D4	
_	FJ692525	HTI	D4	
_	FJ692526	HTI	D4	
_	FJ692527	HTI	D4	
_	FJ692528	HTI	D4	
_	FJ692529	HTI	D4	
_	FJ692530	HTI	D4	
_	FJ692531	HTI	D4	
_	FJ692532	HTI	D4	✓
_	FJ692533	HTI	D4	✓
_	FJ692534	HTI	D4	
_	FJ692535	HTI	D4	
_	FJ692536	HTI	D4	✓
_	FJ692537	HTI	D4	
_	FJ692538	HTI	D4	
_	AB246347	IND	D1	✓
_	AY161158	IND	D1	✓
_	AY161159	IND	D1	✓
_	AY945307	IND	D1	✓
_	DQ315778	IND	D1	✓
_	KT366512	IND	D1	
_	KT366513	IND	D1	
_	KT366514	IND	D1	
_	KT366515	IND	D1	
_	KT366520	IND	D1	
_	KC752138	IND	D1	
_	KC875309	IND	D1	✓
_	KC875308	IND	D1	✓

_	KC875307	IND	D1	✓
_	KC875306	IND	D1	✓
_	KC875305	IND	D1	✓
_	KC875304	IND	D1	✓
_	KC875303	IND	D1	✓
_	KC875302	IND	D1	✓
_	KC875301	IND	D1	✓
_	KC875300	IND	D1	✓
_	KC875299	IND	D1	✓
_	KC875298	IND	D1	✓
_	KC875297	IND	D1	✓
_	KC875296	IND	D1	✓
_	KC875295	IND	D1	✓
_	KC875294	IND	D1	✓
_	KC875293	IND	D1	✓
_	KC875292	IND	D1	✓
_	KC875291	IND	D1	✓
_	KC875290	IND	D1	✓
_	KC875289	IND	D1	✓
_	KC875288	IND	D1	✓
_	KC875287	IND	D1	✓
_	KC875286	IND	D1	✓
_	KC875285	IND	D1	✓
_	KC875284	IND	D1	✓
_	KC875283	IND	D1	✓
_	KC875282	IND	D1	✓
_	KC875281	IND	D1	✓
_	KC875280	IND	D1	✓
_	KC875279	IND	D1	✓
_	KC875278	IND	D1	✓
_	KC875277	IND	D1	✓
_	KC875276	IND	D1	✓
_	KC875275	IND	D1	✓
_	AB090268	IND	D2	✓
_	KM590905	IND	D2	✓
_	KM590906	IND	D2	✓
_	KM590913	IND	D2	✓
_	KM590922	IND	D2	✓
_	KM590907	IND	D2	✓
_	KM590912	IND	D2	✓
_	KM590914	IND	D2	✓

_	KM590918	IND	D2	✓
_	KM590920	IND	D2	✓
_	KT366518	IND	D2	
_	KC875312	IND	D2	✓
_	KC875311	IND	D2	✓
_	KC875310	IND	D2	✓
_	DQ315776	IND	D3	✓
_	KT366519	IND	D3	
_	DQ315777	IND	D3	✓
_	KC752145	IND	D3	
_	KF679988	IND	D3	✓
_	KF679989	IND	D3	✓
_	KF679990	IND	D3	✓
_	KF679991	IND	D3	✓
_	KF679992	IND	D3	✓ (D3_Hap2_IND)
_	KF679993	IND	D3	✓ (D3_Hap2_IND)
_	KP322602	IND	D3	✓
_	KC875337	IND	D3	✓
_	KC875336	IND	D3	✓
_	KC875335	IND	D3	✓
_	KC875334	IND	D3	✓
_	KC875333	IND	D3	✓
_	KC875332	IND	D3	✓
_	KC875331	IND	D3	✓
_	KC875330	IND	D3	✓
_	KC875329	IND	D3	✓
_	KC875328	IND	D3	✓
_	KC875327	IND	D3	✓
_	KC875326	IND	D3	✓
_	KC875325	IND	D3	✓
_	KC875324	IND	D3	✓
_	KC875323	IND	D3	✓ (D3_Hap6_IND)
_	KC875322	IND	D3	✓ (D3_Hap6_IND)
_	KC875321	IND	D3	✓
_	KC875320	IND	D3	✓
_	KC875319	IND	D3	✓
_	KC875318	IND	D3	✓
_	KC875317	IND	D3	✓
_	KC875316	IND	D3	✓
_	KC875315	IND	D3	✓
_	KC875314	IND	D3	✓

_	KC875313	IND	D3	✓
_	KF192830	IND	D4	✓ (D4_Hap3_IND)
_	KF192841	IND	D4	✓ (D4_Hap3_IND)
_	KF192831	IND	D4	✓ (D4_Hap4_IND)
_	KF192832	IND	D4	✓ (D4_Hap4_IND)
_	KF192833	IND	D4	✓ (D4_Hap4_IND)
_	KF192836	IND	D4	✓ (D4_Hap4_IND)
_	KF192837	IND	D4	✓ (D4_Hap4_IND)
_	KF192838	IND	D4	✓ (D4_Hap4_IND)
_	KF192839	IND	D4	✓ (D4_Hap4_IND)
_	KF192834	IND	D4	✓
_	KF192835	IND	D4	✓
_	KF192840	IND	D4	✓
_	DQ315779	IND	D5	✓
_	GQ205377	IND	D5	✓
_	GQ205378	IND	D5	✓ (D5_Hap5_IND)
_	GQ205386	IND	D5	✓ (D5_Hap5_IND)
_	GQ205387	IND	D5	✓ (D5_Hap5_IND)
_	GQ205388	IND	D5	✓ (D5_Hap5_IND)
_	GQ205379	IND	D5	✓
_	DQ315780	IND	D5	✓
_	KP322603	IND	D5	✓
_	GQ205381	IND	D5	✓
_	GQ205382	IND	D5	✓
_	GQ205383	IND	D5	✓
_	GQ205384	IND	D5	✓
_	GQ205385	IND	D5	✓
_	GQ205389	IND	D5	✓
_	KC875342	IND	D5	✓
_	KC875341	IND	D5	✓
_	KC875340	IND	D5	✓
_	KC875339	IND	D5	✓
_	KC875338	IND	D5	✓
C6	**	ITA	D1	
C9	**	ITA	D1	
C10	**	ITA	D1	
C16	**	ITA	D1	
C35	**	ITA	D1	
C39	**	ITA	D1	
_	JN226086	ITA	D1	✓
C41	**	ITA	D1	

C48	**	ITA	D1	
C50	**	ITA	D1	
C55	**	ITA	D1	
C86	**	ITA	D1	
C87	**	ITA	D1	
C104	**	ITA	D1	
C122	**	ITA	D1	
C135	**	ITA	D1	
_	EF514291	ITA	D1	✓
_	EF514296	ITA	D1	✓
_	EF514301	ITA	D1	✓
_	EF514313	ITA	D1	✓
_	EF514314	ITA	D1	✓
_	EU908812	ITA	D1	✓
_	EU908822	ITA	D1	✓
_	GQ486142	ITA	D1	✓
_	GQ486244	ITA	D1	✓
_	GQ486301	ITA	D1	✓
_	JN225971	ITA	D1	
_	JN225983	ITA	D1	
_	JN225996	ITA	D1	
_	JN226004	ITA	D1	
_	JN226017	ITA	D1	
_	JN226032	ITA	D1	
_	JN226073	ITA	D1	✓
_	JN226088	ITA	D1	✓
_	JN226089	ITA	D1	✓
_	JN226090	ITA	D1	✓
_	JN226099	ITA	D1	✓
_	JX849599	ITA	D1	✓
_	DQ304551	ITA	D1	✓
_	DQ304550	ITA	D1	✓
_	DQ304549	ITA	D1	✓
_	DQ304548	ITA	D1	✓
_	DQ304547	ITA	D1	✓
_	DQ486023	ITA	D1	✓
C57	**	ITA	D1	
_	JN225982	ITA	D1	
C2	**	ITA	D2	
C5	**	ITA	D2	
C7	**	ITA	D2	

C25	**	ITA	D2	
C30	**	ITA	D2	
C37	**	ITA	D2	
_	JN226096	ITA	D2	✓
C38	**	ITA	D2	
C56	**	ITA	D2	
_	EF514321	ITA	D2	✓ (D2_Hap3_ITA)
_	EF514325	ITA	D2	✓ (D2_Hap3_ITA)
_	EF514327	ITA	D2	✓
C67	**	ITA	D2	
C85	**	ITA	D2	
C90	**	ITA	D2	
C92	**	ITA	D2	
C47	**	ITA	D2	
C94	**	ITA	D2	
C132	**	ITA	D2	
C134	**	ITA	D2	
C137	**	ITA	D2	
_	EF514285	ITA	D2	✓
_	EF514297	ITA	D2	✓
_	EF514320	ITA	D2	✓
_	EF514322	ITA	D2	
_	EF514326	ITA	D2	
_	EU908819	ITA	D2	✓
_	EU908834	ITA	D2	✓
_	JN225977	ITA	D2	
_	JN225980	ITA	D2	
_	JN225981	ITA	D2	
_	JN226003	ITA	D2	
_	JN226012	ITA	D2	
_	JN226013	ITA	D2	
_	JN226018	ITA	D2	
_	JN226019	ITA	D2	
_	JN226022	ITA	D2	
_	JN226033	ITA	D2	
_	JN226034	ITA	D2	
_	JN226038	ITA	D2	
_	JN226075	ITA	D2	✓
_	JN226076	ITA	D2	✓
_	JN226077	ITA	D2	✓
_	JN226079	ITA	D2	✓

_	JN226080	ITA	D2	✓
_	JN226083	ITA	D2	✓
_	JN226094	ITA	D2	✓
_	JX849540	ITA	D2	✓
_	JX849605	ITA	D2	✓
_	EU908827	ITA	D3	✓
C59	**	ITA	D3	
C1	**	ITA	D3	
C3	**	ITA	D3	
_	JN226009	ITA	D3	
C4	**	ITA	D3	
C8	**	ITA	D3	
C11	**	ITA	D3	
C12	**	ITA	D3	
C13	**	ITA	D3	
C14	**	ITA	D3	
C15	**	ITA	D3	
C17	**	ITA	D3	
C18	**	ITA	D3	
C19	**	ITA	D3	
C20	**	ITA	D3	
C22	**	ITA	D3	
C23	**	ITA	D3	
C24	**	ITA	D3	
C26	**	ITA	D3	
C27	**	ITA	D3	
C28	**	ITA	D3	
C29	**	ITA	D3	
C31	**	ITA	D3	
C32	**	ITA	D3	
C34	**	ITA	D3	
C36	**	ITA	D3	
C40	**	ITA	D3	
_	JN226097	ITA	D3	✓
C42	**	ITA	D3	
C43	**	ITA	D3	
C46	**	ITA	D3	
C49	**	ITA	D3	
C52	**	ITA	D3	
C54	**	ITA	D3	
C60	**	ITA	D3	

C61	**	ITA	D3	
C63	**	ITA	D3	
C71	**	ITA	D3	
C81	**	ITA	D3	
C93	**	ITA	D3	
C97	**	ITA	D3	
C99	**	ITA	D3	
C110	**	ITA	D3	
C116	**	ITA	D3	
C124	**	ITA	D3	
_	DQ486021	ITA	D3	✓
_	DQ486022	ITA	D3	✓
_	EF514283	ITA	D3	✓
_	EF514284	ITA	D3	✓
_	EF514289	ITA	D3	✓
_	EF514290	ITA	D3	✓
_	EF514286	ITA	D3	✓
_	EF514287	ITA	D3	✓
_	EF514288	ITA	D3	✓
_	EF514292	ITA	D3	✓
_	EF514293	ITA	D3	✓
_	EF514294	ITA	D3	✓
_	EF514295	ITA	D3	✓
_	EF514298	ITA	D3	✓
_	EF514299	ITA	D3	✓ (D3_Hap4_ITA)
_	EF514300	ITA	D3	✓
_	EF514311	ITA	D3	✓ (D3_Hap4_ITA)
_	EF514302	ITA	D3	✓
_	EF514304	ITA	D3	✓
_	EF514305	ITA	D3	✓
_	EF514306	ITA	D3	✓
_	EF514308	ITA	D3	✓
_	EF514309	ITA	D3	✓
_	EF514310	ITA	D3	✓
_	EF514312	ITA	D3	✓
_	EF514315	ITA	D3	✓
_	EF514316	ITA	D3	✓
_	EF514317	ITA	D3	✓
_	EF514318	ITA	D3	✓
_	EF514319	ITA	D3	✓
_	EF514323	ITA	D3	✓

_	EF514324	ITA	D3	✓
_	EF514328	ITA	D3	✓
_	EF514329	ITA	D3	✓
_	EF514330	ITA	D3	✓
_	EF514331	ITA	D3	✓
_	EF514332	ITA	D3	✓
_	EF514333	ITA	D3	✓
_	EF514334	ITA	D3	✓
_	EF514335	ITA	D3	✓
_	EF514336	ITA	D3	✓
_	EF514337	ITA	D3	✓
_	EU908799	ITA	D3	✓ (D3_Hap5_ITA)
_	EU908803	ITA	D3	✓
_	EU908814	ITA	D3	✓
_	EU908837	ITA	D3	✓
_	EU908842	ITA	D3	✓ (D3_Hap5_ITA)
_	EU908800	ITA	D3	✓
_	EU908801	ITA	D3	✓
_	EU908802	ITA	D3	✓ (D3_Hap6_ITA)
_	EU908808	ITA	D3	✓ (D3_Hap6_ITA)
_	EU908817	ITA	D3	✓ (D3_Hap6_ITA)
_	EU908843	ITA	D3	✓ (D3_Hap6_ITA)
_	EU908804	ITA	D3	✓
_	EU908806	ITA	D3	✓
_	EU908807	ITA	D3	✓
_	EU908809	ITA	D3	✓
_	EU908821	ITA	D3	✓
_	EU908823	ITA	D3	✓
_	X65257	ITA	D3	✓
_	EU908810	ITA	D3	✓
_	EU908811	ITA	D3	✓
_	EU908813	ITA	D3	✓
_	EU908816	ITA	D3	✓
_	EU908835	ITA	D3	✓
_	JX849504	ITA	D3	✓
_	EU908815	ITA	D3	✓
_	EU908818	ITA	D3	✓
_	EU908820	ITA	D3	✓ (D3_Hap7_ITA)
_	EU908825	ITA	D3	✓ (D3_Hap7_ITA)
_	EU908824	ITA	D3	✓
_	EU908826	ITA	D3	✓

_	EU908828	ITA	D3	✓
_	EU908829	ITA	D3	✓
_	EU908830	ITA	D3	✓
_	EU908831	ITA	D3	✓
_	EU908832	ITA	D3	✓
_	EU908833	ITA	D3	✓
_	EU908836	ITA	D3	✓
_	EU908838	ITA	D3	✓
_	EU908839	ITA	D3	✓
_	EU908840	ITA	D3	✓
_	EU908841	ITA	D3	✓
_	EU908844	ITA	D3	✓
_	GQ486051	ITA	D3	✓
_	GQ486138	ITA	D3	✓
_	GQ486332	ITA	D3	✓
_	JN225963	ITA	D3	
_	JN225964	ITA	D3	
_	JN225965	ITA	D3	
_	JN225966	ITA	D3	
_	JN225967	ITA	D3	
_	JN225968	ITA	D3	
_	JN225969	ITA	D3	
_	JN225972	ITA	D3	
_	JN225973	ITA	D3	
_	JN225975	ITA	D3	
_	JN225978	ITA	D3	
_	JN225979	ITA	D3	
_	JN225984	ITA	D3	
_	JN225985	ITA	D3	
_	JN225986	ITA	D3	
_	JN225987	ITA	D3	
_	JN225989	ITA	D3	
_	JN225990	ITA	D3	
_	JN225991	ITA	D3	
_	JN225992	ITA	D3	
_	JN225993	ITA	D3	
_	JN225994	ITA	D3	
_	JN225995	ITA	D3	
_	JN225997	ITA	D3	
_	JN225999	ITA	D3	
_	JN226000	ITA	D3	

_	JN226001	ITA	D3	
_	JN226002	ITA	D3	
_	JN226005	ITA	D3	
_	JN226006	ITA	D3	
_	JN226007	ITA	D3	
_	JN226010	ITA	D3	
_	JN226011	ITA	D3	
_	JN226014	ITA	D3	
_	JN226015	ITA	D3	
_	JN226016	ITA	D3	
_	JN226020	ITA	D3	
_	JN226021	ITA	D3	
_	JN226023	ITA	D3	
_	JN226024	ITA	D3	
_	JN226025	ITA	D3	
_	JN226026	ITA	D3	
_	JN226027	ITA	D3	
_	JN226028	ITA	D3	
_	JN226029	ITA	D3	
_	JN226030	ITA	D3	
_	JN226031	ITA	D3	
_	JN226035	ITA	D3	
_	JN226037	ITA	D3	
_	JN226070	ITA	D3	✓
_	JN226071	ITA	D3	✓
_	JN226072	ITA	D3	✓
_	JN226082	ITA	D3	✓
_	JN226087	ITA	D3	✓
_	JN226098	ITA	D3	✓
_	JX849501	ITA	D3	✓
_	JX849545	ITA	D3	✓
_	MG585269	ITA	D3	✓
_	DQ329357	ITA	D3	✓
_	DQ329356	ITA	D3	✓
_	DQ486025	ITA	D3	✓
C77	**	ITA	D4	
C115	**	ITA	D4	
_	EF514303	ITA	D7	✓
_	EF514307	ITA	D7	✓
_	JN642126	LBN	D1	✓
_	JN642127	LBN	D1	✓

_	JN642128	LBN	D1	✓
_	JN642129	LBN	D1	✓
_	JN642130	LBN	D1	✓
_	JN642131	LBN	D1	✓
_	JN642132	LBN	D1	✓
_	JN642133	LBN	D1	✓
_	JN642134	LBN	D1	✓
_	JN642135	LBN	D1	✓
_	JN642136	LBN	D1	✓
_	JN642137	LBN	D1	✓
_	JN642138	LBN	D1	✓
_	JN642139	LBN	D1	✓
_	JN642140	LBN	D1	✓
_	JN642141	LBN	D1	✓
_	JN642142	LBN	D1	✓
_	JN642145	LBN	D1	✓
_	JN642146	LBN	D1	✓
_	JN642147	LBN	D1	✓
_	JN642149	LBN	D1	✓
_	JN642150	LBN	D1	✓
_	JN642151	LBN	D1	✓
_	JN642152	LBN	D1	✓
_	JN642153	LBN	D1	✓
_	JN642154	LBN	D1	✓
_	JN642155	LBN	D1	✓
_	JN642156	LBN	D1	✓
_	JN642157	LBN	D1	✓
_	JN642158	LBN	D1	✓
_	JN642161	LBN	D1	✓
_	JN642164	LBN	D1	✓
_	JN642165	LBN	D1	✓
_	JN642166	LBN	D1	✓
_	JN642167	LBN	D1	✓
_	JN642143	LBN	D2	✓
_	JN642144	LBN	D2	✓
_	JN642148	LBN	D2	✓
_	JN642159	LBN	D2	✓
_	JN642160	LBN	D2	✓
_	JN642162	LBN	D2	✓
_	JN642163	LBN	D2	✓
_	AY721612	TUR	D1	✓

_	AY721611	TUR	D1	✓
_	AY721610	TUR	D1	✓
_	AY721609	TUR	D1	✓
_	AY721608	TUR	D1	✓
_	AY721607	TUR	D1	✓
_	AY721606	TUR	D1	✓
_	AY721605	TUR	D1	✓
_	AY796032	TUR	D1	✓
_	AY796030	TUR	D1	✓
_	AB674436	TUR	D1	✓
_	AB674434	TUR	D1	✓
_	AB674433	TUR	D1	✓
_	AB674432	TUR	D1	✓
_	AB674431	TUR	D1	✓
_	AB674430	TUR	D1	✓
_	AB674429	TUR	D1	✓
_	AB674428	TUR	D1	✓
_	AB674427	TUR	D1	✓
_	AB674426	TUR	D1	✓
_	AB674425	TUR	D1	✓
_	AB674424	TUR	D1	✓
_	AB674423	TUR	D1	✓
_	AB674422	TUR	D1	✓
_	AB674421	TUR	D1	✓
_	AB674420	TUR	D1	✓
_	AB674419	TUR	D1	✓
_	AB674416	TUR	D1	✓
_	AB674411	TUR	D1	✓
_	AB674410	TUR	D1	✓
_	AB674409	TUR	D1	✓
_	AB674408	TUR	D1	✓
_	AB674407	TUR	D1	✓
_	AB674406	TUR	D1	✓
_	AB674405	TUR	D1	✓
_	AB674404	TUR	D1	✓
_	AB674403	TUR	D1	✓
_	JF754635	TUR	D1	✓
_	JF754634	TUR	D1	✓
_	JF754633	TUR	D1	✓
_	JF754632	TUR	D1	✓
_	JF754631	TUR	D1	✓

_	JF754630	TUR	D1	✓
_	JF754629	TUR	D1	✓
_	JF754628	TUR	D1	✓
_	JF754627	TUR	D1	✓
_	JF754626	TUR	D1	✓
_	JF754624	TUR	D1	✓
_	JF754623	TUR	D1	✓
_	JF754620	TUR	D1	✓
_	JF754619	TUR	D1	✓
_	JF754618	TUR	D1	✓
_	JF754617	TUR	D1	✓
_	JF754616	TUR	D1	✓
_	JF754615	TUR	D1	✓
_	JF754614	TUR	D1	✓
_	JF754613	TUR	D1	✓
_	JF754612	TUR	D1	✓
_	JF754611	TUR	D1	✓
_	JF754610	TUR	D1	✓
_	JF754609	TUR	D1	✓
_	JF754608	TUR	D1	✓
_	JF754607	TUR	D1	✓
_	JF754604	TUR	D1	✓
_	JF754603	TUR	D1	✓
_	JF754602	TUR	D1	✓
_	JF754601	TUR	D1	✓
_	JF754600	TUR	D1	✓
_	JF754599	TUR	D1	✓
_	JF754598	TUR	D1	✓
_	JF754597	TUR	D2	✓
_	JF754596	TUR	D1	✓
_	JF754595	TUR	D1	✓
_	JF754594	TUR	D1	✓
_	JF754593	TUR	D1	✓
_	JF754592	TUR	D1	✓
_	JF754591	TUR	D1	✓
_	JF754590	TUR	D1	✓
_	JF754589	TUR	D1	✓
_	JF754588	TUR	D1	✓
_	JF754587	TUR	D1	✓
_	AY796031	TUR	D2	✓
_	JF754621	TUR	D2	✓

_	DQ412277	NLD	D1
_	DQ412278	NLD	D1
_	DQ412279	NLD	D1
_	DQ412281	NLD	D1
_	DQ412333	NLD	D1
_	DQ412431	NLD	D1
_	DQ412432	NLD	D1
_	DQ412435	NLD	D1
_	DQ412282	NLD	D1
_	DQ412283	NLD	D1
_	DQ412284	NLD	D1
_	DQ412286	NLD	D1
_	DQ412287	NLD	D1
_	DQ412307	NLD	D1
_	DQ412289	NLD	D1
_	DQ412290	NLD	D1
_	DQ412292	NLD	D1
_	DQ412293	NLD	D1
_	DQ412295	NLD	D1
_	DQ412299	NLD	D1
_	DQ412317	NLD	D1
_	DQ412319	NLD	D1
_	DQ412351	NLD	D1
_	DQ412352	NLD	D1
_	DQ412417	NLD	D1
_	DQ412428	NLD	D1
_	DQ412429	NLD	D1
_	DQ412434	NLD	D1
_	DQ412301	NLD	D1
_	DQ412302	NLD	D1
_	DQ412303	NLD	D1
_	DQ412304	NLD	D1
_	DQ412305	NLD	D1
_	DQ412308	NLD	D1
_	DQ412310	NLD	D1
_	DQ412311	NLD	D1
_	DQ412312	NLD	D1
_	DQ412318	NLD	D1
_	DQ412323	NLD	D1
_	DQ412324	NLD	D1
_	DQ412325	NLD	D1

_	DQ412326	NLD	D1
_	DQ412327	NLD	D1
_	DQ412328	NLD	D1
_	DQ412330	NLD	D1
_	DQ412332	NLD	D1
_	DQ412334	NLD	D1
_	DQ412335	NLD	D1
_	DQ412336	NLD	D1
_	DQ412337	NLD	D1
_	DQ412338	NLD	D1
_	DQ412339	NLD	D1
_	DQ412340	NLD	D1
_	DQ412341	NLD	D1
_	DQ412342	NLD	D1
_	DQ412343	NLD	D1
_	DQ412353	NLD	D1
_	DQ412354	NLD	D1
_	DQ412346	NLD	D1
_	DQ412347	NLD	D1
_	DQ412348	NLD	D1
_	DQ412350	NLD	D1
_	DQ412355	NLD	D1
_	DQ412356	NLD	D1
_	DQ412357	NLD	D1
_	DQ412359	NLD	D1
_	DQ412360	NLD	D1
_	DQ412363	NLD	D1
_	DQ412365	NLD	D1
_	DQ412366	NLD	D1
_	DQ412369	NLD	D1
_	DQ412373	NLD	D1
_	DQ412374	NLD	D1
_	DQ412376	NLD	D1
_	DQ412379	NLD	D1
_	DQ412380	NLD	D1
_	DQ412381	NLD	D1
_	DQ412383	NLD	D1
_	DQ412389	NLD	D1
_	DQ412390	NLD	D1
_	DQ412391	NLD	D1
_	DQ412392	NLD	D1

_	DQ412393	NLD	D1	
_	DQ412398	NLD	D1	
_	DQ412399	NLD	D1	
_	DQ412400	NLD	D1	
_	DQ412401	NLD	D1	
_	DQ412402	NLD	D1	
_	DQ412403	NLD	D1	
_	DQ412404	NLD	D1	
_	DQ412405	NLD	D1	
_	DQ412407	NLD	D1	
_	DQ412410	NLD	D1	
_	DQ412411	NLD	D1	
_	DQ412412	NLD	D1	
_	DQ412414	NLD	D1	
_	DQ412415	NLD	D1	
_	DQ412418	NLD	D1	
_	DQ412419	NLD	D1	
_	DQ412421	NLD	D1	
_	DQ412422	NLD	D1	
_	DQ412423	NLD	D1	
_	DQ412426	NLD	D1	
_	DQ412427	NLD	D1	
_	DQ412430	NLD	D1	
_	DQ412433	NLD	D1	
_	DQ412436	NLD	D1	
_	DQ412438	NLD	D1	
_	DQ412440	NLD	D1	
_	DQ412441	NLD	D1	
_	DQ412442	NLD	D1	
_	GQ486080	NLD	D1	✓
_	DQ412298	NLD	D1	
_	DQ412362	NLD	D2	
_	DQ412294	NLD	D2	
_	DQ412300	NLD	D2	
_	DQ412320	NLD	D2	
_	DQ412322	NLD	D2	
_	DQ412329	NLD	D2	
_	DQ412367	NLD	D2	
_	DQ412382	NLD	D2	
_	DQ412408	NLD	D2	
_	DQ412437	NLD	D2	

_	DQ412368	NLD	D2
_	DQ412372	NLD	D2
_	DQ412378	NLD	D2
_	DQ412397	NLD	D2
_	DQ412424	NLD	D2
_	DQ412385	NLD	D3
_	DQ412280	NLD	D3
_	DQ412288	NLD	D3
_	DQ412291	NLD	D3
_	DQ412297	NLD	D3
_	DQ412306	NLD	D3
_	DQ412313	NLD	D3
_	DQ412314	NLD	D3
_	DQ412316	NLD	D3
_	DQ412331	NLD	D3
_	DQ412344	NLD	D3
_	DQ412345	NLD	D3
_	DQ412349	NLD	D3
_	DQ412370	NLD	D3
_	DQ412375	NLD	D3
_	DQ412384	NLD	D3
_	DQ412387	NLD	D3
_	DQ412388	NLD	D3
_	DQ412285	NLD	D4
_	DQ412358	NLD	D4
_	DQ412377	NLD	D4
_	DQ412386	NLD	D4
_	DQ412394	NLD	D4
_	DQ412406	NLD	D4
_	DQ412409	NLD	D4
_	DQ412413	NLD	D4
_	DQ412395	NLD	D4
_	DQ412416	NLD	D4
_	DQ412420	NLD	D4
_	DQ412425	NLD	D4
_	DQ412296	NLD	D7
_	DQ412309	NLD	D7
_	DQ412315	NLD	D7
_	DQ412321	NLD	D7
_	DQ412361	NLD	D7
_	DQ412364	NLD	D7

_	DQ412371	NLD	D7	
_	DQ412396	NLD	D7	
_	DQ412439	NLD	D7	
_	AY233278	ZAF	A1	✓ (outgroup)
_	AJ309369	FRA	A2	✓ (outgroup)
_	AB194951	CMR	A3	✓ (outgroup)
_	AB010292	JPN	B1	✓ (outgroup)
_	U87747	ZAF	B2	✓ (outgroup)
_	AB033555	IDN	B3	✓ (outgroup)
_	AY217375	CHN	C1	✓ (outgroup)
_	AB049609	-	C2	✓ (outgroup)
_	X75665	NCL	C3	✓ (outgroup)
_	DQ060823	AGO	E	✓ (outgroup)
_	AB091255	CIV	E	✓ (outgroup)
_	AF405706	DEU	G	✓ (outgroup)
_	AB064310	USA	G	✓ (outgroup)
_	AY090455	NIC	F2a	✓ (outgroup)
_	AB036909	VEN	F3	✓ (outgroup)
_	AB365446	BOL	F4	✓ (outgroup)
_	AB059661	USA	H	✓ (outgroup)
_	AY090454	NIC	H	✓ (outgroup)
_	FJ023661	LAO	I	✓ (outgroup)
_	AB231908	VNM	I	✓ (outgroup)

***** Pinho, J.R.R. *et al.* In preparation

****** Perno, C.F. *et al.* In preparation.

Figure S1: Maximum likelihood tree represented in Figure 2 with the accession numbers of the reference sequences of included in this study.

Figure S2: Subtree of HBV/D1 and D2 subgenotypes with the accession numbers of the sequences included in the analysis.

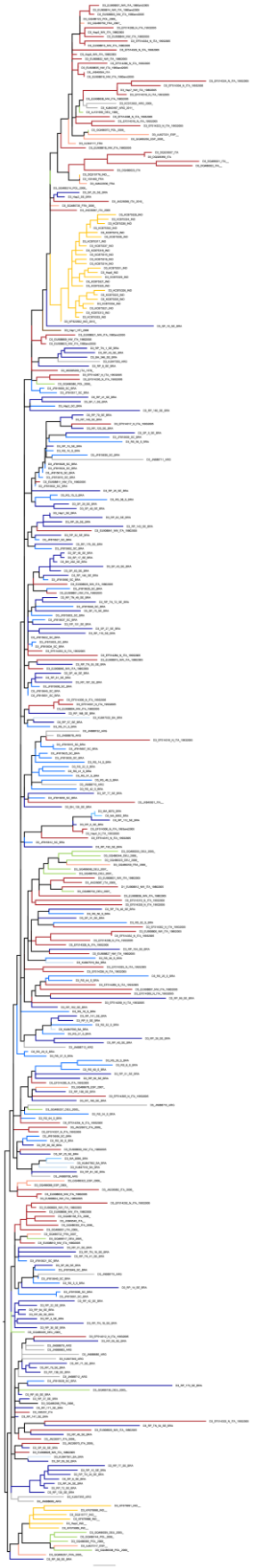
Figure S3: Subtree of HBV/D3 with the accession numbers of the sequences included in analysis.

Fig. S1

Nota à banca: durante a submissão, as figuras suplementares serão submetidas apenas como arquivo, não tendo seu tamanho limitado pelo tamanho da página. Entendemos, no entanto, que no formato de tese a legibilidade dos rótulos das árvores fique prejudicada.

Fig. S2

Fig. S3



CAPÍTULO IV

"GENETIC VARIANTS IN *RARB* AND ITS EFFECT ON HEPATITIS B VIRUS SUSCEPTIBILITY"

Manuscrito em preparação para submissão na *Liver International.*

Genetic variants in *RARB* and its effect on Hepatitis B Virus susceptibility

Bibiane A. Godoy[1], Cainã M. Couto-Silva[2], Tábita Hünemeier[2], Nelson J. R. Fagundes[1,3*]

[1]Postgraduation Program in Genetics and Molecular Biology, Federal University of Rio Grande do Sul, Porto Alegre, RS, Brazil

[2]Department of Genetics and Evolutionary Biology, Institute of Biosciences, University of São Paulo, São Paulo, SP, Brazil

[3]Department of Genetics, Institute of Biosciences, Federal University of Rio Grande do Sul, Porto Alegre, RS, Brazil

Running title: *RARB* effect on HBV susceptibility

***Corresponding author:**

Nelson J.R. Fagundes

Departamento de Genética, UFRGS

Caixa Postal 15053

91501-970, Porto Alegre, RS, Brazil

Tel: +55 (51) 33089820

e-mail: nelson.fagundes@ufrgs.br

Abstract

Hepatitis B Virus (HBV) affects hundreds of millions of people worldwide, causing a high number of deaths. Variation in the natural course of HBV infection results from an interplay between viral and host genetic factors. In this study, we characterize the genetic variation in a set of HBV strains and in their human host populations to evaluate its impact on HBV prevalence. We obtained genomewide SNP data for 48 Native Americans individuals previously characterized for its circulating HBV lineages. These individuals were arranged in two groups: HBV+ (N=29) and HBV- (N=19), and compared considering 640 SNPs in genes associated to HBV response. The analysis of HBV mutations did not reveal variants with known effects for drug resistance or vaccine escape. However, we found statistically significant differences between HBV+ and HBV- involving 18 SNPs, most of which in the *RARB* gene. *RARB* is an important regulator of *SLC10A1*, which codes for the membrane receptor used by HBV to infect hepatocytes. Host genetic factors have been suggested to play an important role in HBV persistence, disease progression, and vaccine-induced immunity. However, the involvement of *RARB* suggests a novel pathway by which genetic factors may affect susceptibility to HBV in human populations.

Key words: Host genetic factors; HBV internalization, protein interactions

1. INRODUCTION

Hepatitis B Virus (HBV) is an infectious pathogen that affects hundreds of millions of people worldwide, causing a high number of annual deaths due to severe outcomes such as fulminant hepatitis, cirrhosis, and hepatocellular carcinoma[1,2] HBV transmission may occur both vertically, from mother to child, or horizontally, due to sexual or parenteral contact[2]. HBV belongs to the Hepadnaviridae family – a group of enveloped viruses whose members are highly hepatotropic and present a small and partially double-stranded DNA genome. In the case of HBV, the ~3,200 base pair (bp) genome is organized in four overlapping open reading frames (ORF): polymerase (*P*) and HBx (*X*) genes which act in viral replication, the core gene (*C*) that produces the core protein (HBcAg) and the excretory e antigen (HBeAg), and the surface gene (*S*) that encodes the three surface antigens[2]. Based on phylogenetic methods, human HBV has been classified into ten genotypes (A-J) having different geographical distributions[3]. Together with results from molecular dating techniques, this suggests a long history of coevolution between HBV and the human host[4].

In the process of infection, HBV binds to a specific receptor on the surface of the hepatocyte, being endocytosed and then released into the cytoplasm. HBV DNA is transported to the nucleus, where it is converted by host enzymes into a covalently closed circular DNA (cccDNA) - a functional mini-chromosome that serve as a template for HBV replication and whose persistence represents the main obstacle for curing the disease[5]. A new HBV genome is produced from this template through the reverse transcriptase activity of HBV polymerase, in addition to all the proteins necessary for the formation of a functional viral particle that will be released from the hepatocyte[2]. In cases of acute hepatitis, which is the most frequent outcome of HBV infection, viral clearance occurs asymptomatically and

spontaneously. However, HBV may persist as an incurable chronic infection especially when the virus is acquired early in life, as in areas of high HBV prevalence[6].

Variation in the natural course of HBV infection results from the interplay among viral and host factors. Viral load, HBV genotype and the occurrence of *de novo* mutations may alter viral pathobiology and response to treatment[5]. On the other hand, the host genetic profile may affect immune and cellular response to the virus, thus favoring or not HBV infection and influencing disease progression[7], resulting in different clinical outcomes that ranges from asymptomatic to the acute liver failure, cirrhosis or hepatocellular carcinoma (HCC)[8]. The high number of deaths related to these complications have placed HBV alongside the major infectious diseases which represent a global health threat whose eradication requires concerted global actions[1,9].

In general, South America presents a low rate of HBV infection except in the Amazon region and for Native Americans populations, which show intermediate to high HBV prevalence facilitated by cultural practices[10] and a more vulnerable social condition[1]. Another concern in this region is the high endemicity of Hepatitis D Virus (HDV), which may lead to more severe outcomes after HBV infection[11]. In a previous study[12], we described circulating HBV lineages in a large sample of Native South American individuals and confirmed its high overall prevalence (~10%) in these populations. However, while some populations showed no HBV occurrence, other presented values as high as ~80%. In the present study, we investigate the role of genetic variation in the virus and its human host populations to test if there are genetic factors related the wide variation in prevalence estimates.

2. MATERIAL AND METHODS

2.1 Populations

This study included 48 Native Americans individuals from eight different populations in the Amazon region (six populations), Central Brazilian Plateau, and Southern Brazil. These populations are part of a repository of historic collections carried out between 1983 and 1990 with logistic support of a Brazilian governmental agency (National Indian Foundation — FUNAI). Original samples consisted of blood and saliva collected under Brazilian National Ethics Commission approval (CONEP Resolution no. 123/98). Participants collaborated voluntarily to the study and, regard to illiteracy, the oral consent was obtained from all subjects observing the Helsinki Declaration. The ethic committee approved the oral consent procedure as well as the use of these samples in population and evolutionary studies.

Previous studies characterized circulating HBV strains in these populations[12], as well as their genomewide genetic variation in Single Nucleotide Polymorphism (SNP) markers[13]. Eight populations were included in both studies, differences in sample size notwithstanding, with four of them showing evidence for HBV circulation[12]. Depending on the HBV status of its parental population[11], all individuals studied by Skoglund et al.[13] belonging to these eight populations were arranged in two groups: HBV+ and HBV-. HBV+ included Apalaí (N=4), Arara (N= 4), Guarani (N=17), and Karitiana, (N=4) individuals (N=29), while HBV- included Suruí (N=3), Urubu Kaapor (N=3), Xavante (N=11), and Zoró (n=1) individuals (N=19) (Table S1).

2.2 HBV mutations detection

We investigated all HBV isolates obtained from Apalaí, Arara and Guarani individuals[12] for HBV genetic variants with clinical relevance using HBV Geno2pheno v.2 (https://hbv.geno2pheno.org/index.php). This tool compares each HBV sequence against annotated references and evaluates all mutations in the reverse transcriptase domain of Polymerase and Surface genes for its predicted effect related to drug resistance (Lamivudine, Adefovir, Entecavir, Tenofovir and Telbivudine) and to vaccine and immunological escape, according to the literature.

2.3 Selection of host genes linked to HBV response

SNP data from the host populations were obtained using the Affymetrix Human Origins array for Native American populations[13] We used previous gene expression microarray datasets for HBV-associated acute liver failure[14] to select genes involved in the main altered processes leading to acute liver failure. We chose to use this specific outcome because it has a fast progression, is related to HBV/HDV co-infection[11], and show high morbidity and mortality[2], characteristics that may result on a higher impact over allele frequencies between groups.

We used two criteria to include SNPs on the dataset: 1) SNP is in differentially expressed genes, or 2) SNP is in genes coding for highly connected proteins in the protein-protein interaction network. In addition, we included SNPs in genes directly involved with HBV control or cellular infection process according to the Online Mendelian Inheritance in Man (OMIM, https://www.ncbi.nlm.nih.gov/omim), Kyoto Encyclopedia of Genes and Genomes (KEGG, https://www.kegg.jp/kegg/) and Ensembl

(https://www.ensembl.org/index.html) databases. A total of 640 SNPs belonging to 38 genes located along 15 different chromosomes were selected for analysis.

2.4 Statistical Analysis

The statistical analysis of SNP data was performed using PLINK v.1.07 (http://zzz.bwh.harvard.edu/plink)[15]. We tested for genetic differences between HBV+ and HBV- populations based on a case-control rationale, using the Fisher's exact test to compare populations. The statistical significance for population comparisons was adjusted for multiple comparisons based on the Benjamini–Hochberg false discovery rate (FDR-BH)[16] assuming corrected *P*-value threshold of 0.05. We investigated the possible functional effect for SNPs showing statistical differences between groups using the "Variant Effect Predictor" tool in Ensembl (https://www.ensembl.org/index.html).

2.5 STRING biological interaction validation

The functional protein association networks database STRING v10 (http://string-db.org/)[17] was used to identify biological interactions among gene products. We analyzed all genes used in the case-control study (see above) under the "multiple proteins" option, and their interactions were estimated based on "Experiments", "Databases" and "Co-expression" with the minimum required confidence score set to 0.4 (medium confidence). We evaluated functional enrichment using STRING-tool considering Gene Ontology (GO) biological processes. Because the gene showing the strongest signal of HBV+ association (*RARB*, see results) was not connected to any other node in the network, we performed a second analysis adding the remaining Retinoic Acid Receptors (RAR) genes (*RARA* and *RARG*) as well as *RXRA* and allowing a maximum number of 5 interactors.

3. RESULTS

The analysis of HBV mutations in these populations did not reveal variants with known effects for drug resistance or vaccine escape. However, we found various amino acid substitutions with unknown clinical relevance affecting *P* and *S* genes (Table S2).

We found significant genetic associations for HBV+/HBV-status for 18 SNPs after correcting for multiple comparisons (Table 1). Most SNPs occurred in *RARB* gene, followed by *IGF1*, *F5* and *ALDH6A1* genes. With the only exception of *F5*, the minor alleles presented higher frequencies in HBV- group (Table 1). For *RARB* (one), *IGF1* (two) and *ALDH6A1* (one), we found SNPs predicted to occur in regulatory regions. No SNPs occurred in coding regions or were associated to other functional effects (Table 1).

The interaction network for all genes included indicate a main cluster of highly connected genes associated to defense and inflammatory response, platelet-associated processes and response to stress (Fig. 1a). Two of the genes having HBV+ associated SNPs: *F5* and *IGF1* are connected to other proteins. On the other hand, *ALDH6A1* had low connectivity with the rest of network, and several genes appeared as isolated nodes. This includes *RARB*, which had the strongest association signal in the case-control approach. Including other RAR genes and allowing more connections resulted in a sub-network of RAR genes linked to *SRC*, *IGF1* and then to the highly connected cluster recovered in the previous analysis (Fig. 1b), highlighting an indirect connection between *RARB* and the processes mentioned above.

4. DISCUSSION

In this study, we evaluated the interplay between host and virus genetic variation in possibly influencing the success of HBV infection in a sample of Native American populations. Mutation screening in HBV strains isolated from HBV+ populations did not highlight a role of HBV mutations affecting virulence, infectivity, or severe outcomes from HBV infection (Table S1). Of course, mutation screening can only be performed in HBV+ populations, which complicates the comparison between HBV+ and HBV- groups. Anyway, our results suggest that HBV+ populations do not harbor an exceptional set of HBV lineages.

On the other hand, our study indicates important genetic differences between HBV+ and HBV- populations (Table 1; Fig. 1). The onset of HBV infection involves a low-affinity and reversible binding of HBV to heparan sulfate proteoglycans (HSPG) on the hepatocyte surface. Following this event, HBV interacts with high affinity with its specific receptor, the hepatic bile acid transporter co-transposing polypeptide (NTCP), triggering viral internalization. This receptor is expressed exclusively on the membrane of hepatocytes and is crucial for the host specificity of HBV[1]. However, genes coding for these receptors (*HSPG2* and *SLC10A1*, respectively) showed no differences between HBV+ and HBV- populations.

RARB, on its turn, is a member of the nuclear receptor superfamily of DNA-binding transcription factors[18]. Its product, RARB, interacts with other RARs (RARA, RARG) and RXRA (Fig. 1b) to become an important regulator of *SLC10A1*, thus affecting the production of the NTCP receptor[18]. Functional studies have shown that RARB antagonists produces a rapid decrease in NTCP levels, with strong anti-HBV activity[19,20]. RARB modulation have been shown to have higher impact on HBV infection susceptibility than others RARs[19,20]. The protagonism of *RARB* during HBV infection is corroborated by recent studies

associating RARs to outcomes such as HCC and cirrhosis[21–24], and by evidence of expression control of *RARB* mediated by HbX-induced methylation[22,25]. RARB-mediated NTCP regulation may be also important to protect from HDV infection, which also binds to NTCP during infection[26,27]. Differently from *RARB*, which is clearly associated to an important route for HBV infection, it is more difficult to find a plausible biological explanation for the association between HBV+/HBV- status and the other significant signals in our study (*IGF1*, *F5*, *ALDH6A1*). Nonetheless, it suggests that biological processes associated to HBV modulation involving inflammation, immune and stress response, or platelet processes (Fig. 1) may also affect HBV prevalence.

Our study has two important limitations: a relatively small sample size, and the fact that not all individuals in HBV+ populations carry HBV. However, the statistically significant differences between HBV+ and HBV- populations was robust to stringent *P*-value adjustments for multiple testing. Besides, if HBV prevalence is though as a population feature, then the genetics of population may be informative for risk factors associated to prevalence irrespective of the specific epidemiological status of their individuals. Indeed, our results contribute to a growing body of evidence suggesting that host genetic factors may play an important role in HBV persistence, disease progression, and vaccine-induced immunity[28,29]. Non-HLA genes (in addition to non-classical and classical HLA genes) have been associated with disease progression[28] and vaccine-induced immunity[29]. However, HBV persistence, which is the "phenotype" most correlated to the "HBV presence" studied by us, had been only associated to variation in classical HLA genes[28]. Our results suggest a novel pathway by which genetic factors may affect susceptibility to HBV in human populations, whose general relevance may be confirmed in further investigations.

AKNOWLEDGEMENTS

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), and by the Postgraduation Program in Genetics and Molecular Biology (PPGBM-UFRGS).

REFERENCES

1. WHO. Global Hepatitis Report, 2017. 2017. ISBN 978-92-4-156545-5.

2. Yuen M-F, Chen D-S, Dusheiko GM, et al. Hepatitis B virus infection. *Nat Rev Dis Prim*. 2018;4:18035.

3. Sunbul M. Hepatitis B virus genotypes: Global distribution and clinical importance. *World J Gastroenterol*. 2014;20:5427–5434.

4. Paraskevis D, Magiorkinis G, Magiorkinis E, et al. Dating the origin and dispersal of hepatitis B virus infection in humans and primates. *Hepatology* 2013;57:908-916.

5. Petersen J, Thompson AJ, Levrero M. Aiming for cure in HBV and HDV infection. *Journal of Hepatology*. 2016;65:835–848.

6. Stevens CE, Beasley RP, Tsui J, Lee W-C. Vertical Transmission of Hepatitis B Antigen in Taiwan. *N Engl J Med*. 1975;292:771–774.

7. Cheng X, Xia Y, Serti E, et al. Hepatitis B virus evades innate immunity of hepatocytes but activates cytokine production by macrophages. *Hepatology*. 2017;66:1779–1793.

8. Geoghegan JL, Duchêne S, Holmes EC. Comparative analysis estimates the relative frequencies of co-divergence and cross-species transmission within viral families. *PLoS Pathog*. 2017;13:e1006215.

9. Locarnini S, Chen DS, Shibuya K. No more excuses: Viral hepatitis can be eliminated.

The Lancet. 2016;387:1703–1704.

10. Coimbra Júnior CE, Santos RV, Yoshida CF, et al. Hepatitis B epidemiology and cultural practices in Amerindian populations of Amazonia: the Tupí-Mondé and the Xavánte from Brazil. *Soc Sci Med*. 1996;42:1735–43.
11. Gomes-Gouvêa MS, Soares MCP, Bensabath G, et al. Hepatitis B virus and hepatitis delta virus genotypes in outbreaks of fulminant hepatitis (Labrea black fever) in the western Brazilian Amazon region. *J Gen Virol*. 2009;90:2638–2643.
12. Godoy BA, Gomes-Gouvêa MS, Zagonel-Oliveira M, et al. High prevalence of HBV/A1 subgenotype in native south Americans may be explained by recent economic developments in the Amazon. *Infect Genet Evol*. 2016;43:354–363.
13. Skoglund P, Mallick S, Bortolini MC, et al. Genetic evidence for two founding populations of the Americas. *Nature*. 2015;525:104–108.
14. Lin H, Zhang Q, Li X, et al. Identification of key candidate genes and pathways in hepatitis B virus-associated acute liver failure by bioinformatical analysis. *Medicine (Baltimore)*. 2018;97:e9687.
15. Purcell S, Neale B, Todd-Brown K, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am J Hum Genet*. 2007;81:559–575.
16. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc*. 1995;57:289–300.
17. Szklarczyk D, Franceschini A, Wyder S, et al. STRING v10: Protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res*. 2015;43:D447–D452.
18. Huang P, Chandra V, Rastinejad F. Retinoic acid actions through mammalian nuclear receptors. *Chemical Reviews*. 2014;114:233–254.

19. Li B, Wang Y, Shen F, et al. Identification of Retinoic Acid Receptor Agonists as Potent Hepatitis B Virus Inhibitors via a Drug Repurposing Screen. *Antimicrob Agents Chemother*. 2018;AAC.00465-18. doi:10.1128/AAC.00465-18

20. Tsukuda S, Watashi K, Iwamoto M, et al. Dysregulation of retinoic acid receptor diminishes hepatocyte permissiveness to hepatitis B virus infection through modulation of sodium taurocholate cotransporting polypeptide (NTCP) expression. *J Biol Chem*. 2015;290:5673–5684.

21. Cortes E, Lachowski D, Rice A, et al. RAR-β is downregulated in HCC & cirrhosis and its expression inhibits myosin-driven activation and durotaxis in hepatic stellate cells. *Hepatology*. 2018. doi:10.1002/hep.30193

22. Jung JK, Park SH, Jang KL. Hepatitis B virus X protein overcomes the growth-inhibitory potential of retinoic acid by downregulating retinoic acid receptor-b2 expression via DNA methylation. *J Gen Virol*. 2010;91:493–500.

23. Yang B, Guo M, Herman JG, et al. Aberrant promoter methylation profiles of tumor suppressor genes in hepatocellular carcinoma. *Am J Pathol*. 2003;163:1101–7.

24. Song M, Sun Y, Tian J, et al. Silencing Retinoid X Receptor Alpha Expression Enhances Early-stage Hepatitis B Virus Infection In Cell Cultures. *J Virol*. 2018;92:JVI.01771-17.

25. Zhang C, Chen X, Liu H, et al. Alpha fetoprotein mediates HBx induced carcinogenesis in the hepatocyte cytoplasm. *Int J Cancer*. 2015;137:1818–1829.

26. Yan H, Zhong G, Xu G, et al. Sodium taurocholate cotransporting polypeptide is a functional receptor for human hepatitis B and D virus. *Elife*. 2012;1:e00049.

27. Schieck A, Schulze A, Gähler C, et al. Hepatitis B virus hepatotropism is mediated by specific receptor recognition in the liver and not restricted to susceptible hosts.

Hepatology. 2013;58:43–53.

28. Akcay IM, Katrinli S, Ozdil K, et al. Host genetic factors affecting hepatitis B infection outcomes: Insights from genome-wide association studies. *World J Gastroenterol* 2018;24:3347-3360.

29. Hennig BJ, Fielding K, Broxholme J, et al. Host Genetic Factors and Vaccine-Induced Immunity to Hepatitis B Virus Infection. PLoS ONE 2008;3:e1898.

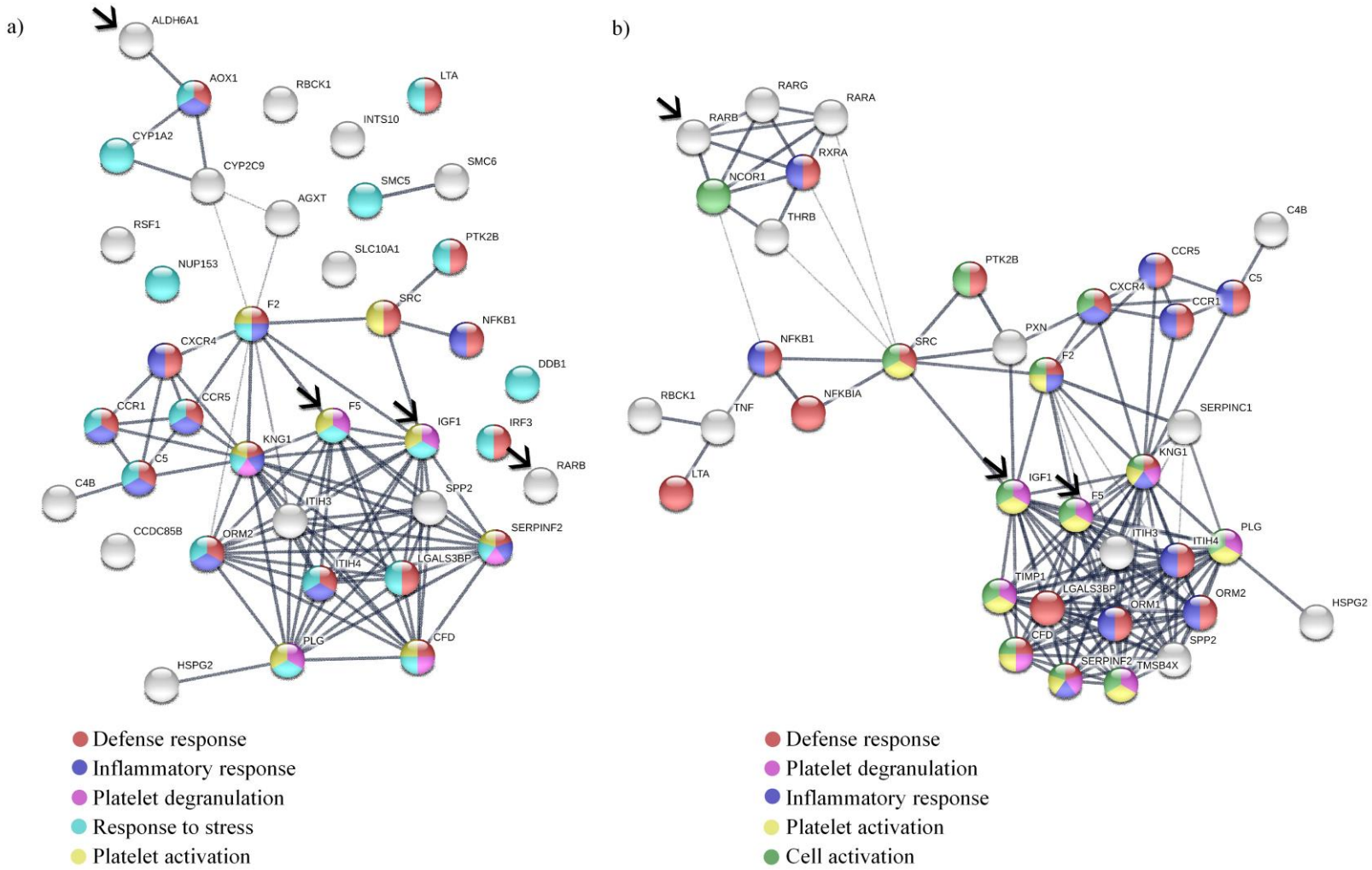
Table 1. Genetic associations for HBV+/HBV- status

Gene	Chr[1]	SNP	Alleles[2]	MAF[3]		*P*[4]	FDR-BH[5]
				HBV+	HBV-		
RARB	3	rs1286654	A/C	0.026	0.362	5.55E-05	0.016
		rs7616278	A/G	0.026	0.345	1.11E-04	0.016
		rs1299407	A/G	0.026	0.345	1.11E-04	0.016
		rs1286655	A/C	0.026	0.328	2.23E-04	0.016
		rs1286648	G/T	0.026	0.310	4.51E-04	0.016
		rs1153584	A/G	0.026	0.310	4.51E-04	0.018
		rs4681063	T/G	0.079	0.397	7.82E-04	0.018
		rs2568875	G/A	0.132	0.466	7.91E-04	0.018
		rs2056777	C/T	0.132	0.466	7.91E-04	0.018
		rs6768669	T/C	0.026	0.286	9.37E-04	0.031
		rs4607073*	G/T	0.026	0.276	1.88E-03	0.037
IGF1	12	rs10745940	G/A	0.184	0.552	3.33E-04	0.016
		rs978458*	C/T	0.184	0.552	3.33E-04	0.016
		rs972936*	C/T	0.184	0.552	3.33E-04	0.016
		rs11111274	A/G	0.184	0.552	3.33E-04	0.016
		rs12313279	G/A	0.184	0.536	6.33E-04	0.018
F5	1	rs9332531	C/A	0.605	0.241	5.35E-04	0.018
ALDH6A	14	rs2239556*	C/T	0.053	0.345	9.29E-04	0.021

1- Chromosome; 2- Major/minor allele; 3- Minor allele frequency; 4- Unadjusted *P*-value for Fisher's exact test; 5- Benjamini-Hochberg false discovery rate. *SNPs occurring in regulatory regions according to Ensembl (https://www.ensembl.org).

FIGURE LEGENDS

Fig. 1. Interaction network for proteins involved with HBV response. Nodes represent proteins and the edges represent their interactions according to a minimum score of medium confidence (0.40). The arrows indicate the proteins for which we found significant results in the case x control association test. Colors indicate the most enriched functional categories according to Gene Ontology for a) query proteins only; b) query proteins including other RAR receptors and RXRA with a maximum of five additional interactions.


a)
ALDH6A1
AOX1
RBCK1
LTA
INTS10
CYP1A2
CYP2C9
SMC6
SMC5
AGXT
RSF1
PTK2B
SLC10A1
NUP153
F2
SRC
NFKB1
CXCR4
DDB1
CCR1
CCR5
F5
IGF1
IRF3
KNG1
RARB
C5
C4B
SPP2
ITIH3
CCDC85B
ORM2
SERPINF2
ITIH4
LGALS3BP
PLG
CFD
HSPG2
Defense response
Inflammatory response
Platelet degranulation
Response to stress
Platelet activation
b)
RARG
RARA
RARB
RXRA
NCOR1
THRB
C4B
CCR5
PTK2B
CXCR4
C5
CCR1
PXN
NFKB1
SRC
F2
RBCK1
NFKBIA
TNF
SERPINC1
KNG1
LTA
IGF1
F5
ITIH3
ITIH4
PLG
TIMP1
LGALS3BP
ORM1
ORM2
HSPG2
CFD
SPP2
SERPINF2
TMSB4X
Defense response
Platelet degranulation
Inflammatory response
Platelet activation
Cell activation

SUPPLEMENTARY MATERIAL

Table S1. Further details about populations included in this study

Group	Population	N[1]	Linguistic affiliation	Region	Sampling year	HBV prevalence[2]
HBV+	Apalaí	4	Carib	Amazon	1983	12,5%
	Arara	4	Carib	Amazon	1985	4,3%
	Guarani	17	Tupi	Southern Brazil	1988	3,7%
	Karitiana	4	Tupi	Amazon	1986	10,0%
HBV-	Suruí	4	Tupi	Amazon	1990	0,0%
	Urubu-Kaapor	3	Tupi	Amazon	1983	0,0%
	Xavante	11	Macro-Ge	Brazilian Central Plateau	1990	0,0%
	Zoró	1	Tupi	Amazon	1990	0,0%

Notes: 1- according to Skoglund et al.[12]; 2- according to Godoy et al.[11]

Table S2. HBV mutations in HBV+ populations

Populations	Mutations	
	RT domain	**HBsAg**
Apalaí	E11Q	G10AEGV
	H13Y	L12LMV
	R18RS	L13FL
	P20HLPR	Q16LPQR
	A21AV	F20FL
	L29FILV	T37INST
	D45DE	H60HQ
	L69FILV	K122R
	N123D	
	M129L	
	S137T	
	L151F	
	N246S	
	T259S	
	D271E	
	T313A	
Arara	D7T	A194V
	V27FV	S207N
	N122H	
	M129L	
	V163I	
	I253V	
	K333N	
Guarani	N122H	S207N
	M129L	
	V163I	
	I253V	
	T259S	
	V278I	

Note: There were no HBV strains isolated from Karitiana individuals in Godoy et al.[11]

CAPÍTULO V – DISCUSSÃO GERAL

A análise filogenética é uma ferramenta importante no estudo de patógenos porque permite caracterizar e estabelecer relações entre diferentes linhagens e hospedeiros, fornecendo informações que podem ser interpretadas no contexto histórico/evolutivo e também de patogenicidade (Hartfield et al. 2014). Este trabalho buscou contribuir na elucidação da história evolutiva dos genótipos americanos F e H através da comparação entre as duas principais topologias sugeridas para HBV. Nossos resultados mostraram a ocorrência de suporte decisivo em favor da topologia enraizada nos genótipos F e H sugerida por Inferência Bayesiana (BI). Adicionalmente, mostrou que a aceleração das taxas de substituição proposta para o ramo que leva aos genótipos F e H só é sustentada sob o critério da Máxima Verossimilhança (ML), desaparecendo quando BI é considerada. Esses resultados sugerem que, para o que se conhece de HBV em primatas até o momento, existe maior suporte estatístico para a interpretação de que o acúmulo de diferenças observado nas linhagens americanas F e H se deve a uma divergência antiga e não a uma aceleração provocada por seleção positiva. Essa afirmação é sustentada pela topologia de maior suporte que mostra a primeira divergência no clado formado pelo HBV de humanos e macacos do Velho Mundo separando linhagens americanas das demais. Além disso, essa interpretação filogenética concorda com a hipótese de dispersão do HBV junto aos primeiros colonizadores da América há ~16000 anos, o que justificaria também o acúmulo de diferenças tão importantes nessas linhagens, espelhando o ocorrido com as populações Nativas Americanas (Godoy et al. 2013; Raghavan et al. 2015).

Nossas análises também foram sugestivas de que a utilização de macacos do Novo Mundo para o enraizamento da filogenia de HBV não configura a melhor estratégia provavelmente devido à diferença excessiva dessas espécies em relação aos demais. No presente trabalho, o enraizamento nos genótipos F e H produziram análises de maior suporte. Por fim, mostramos também que os testes para seleção positiva se mantiveram iguais para ambas topologias, evidenciando ocorrência de seleção positiva para códons do gene P e X. O HBV apresenta uma organização genômica complexa com ocorrência de ORFs sobrepostas que configuram restrições seletivas adicionais (Torres et al. 2011). Nossas análises incluíram apenas partes não-sobrepostas do genoma de HBV, de modo que análises mais propícias para este tipo de genoma devem ser consideradas para confirmação desses resultados (Li et al. 2017).

Como previamente mencionado, as distribuições genotípicas do HBV refletem padrões históricos especialmente relacionados com os movimentos migratórios da população em questão. Devido à alta ocorrência do genótipo D do HBV na região sul do Brasil, trabalhos anteriores têm sugerido uma relação com o histórico importante de migrações europeias da região (Bertolini et al. 2012; Gusatti et al. 2015; Paoli et al. 2018). Nesse trabalho nós descrevemos a relação em nível filogenético e populacional entre isolados de HBV/D circulantes na Americana Latina e Europa.

Nossos resultados confirmaram as similaridades entre o HBV/D isolado da Itália com àqueles dos países com maior influência europeia recente (Argentina e das regiões Sul e Sudeste do Brasil), entretanto as análises a nível de subgenótipo evidenciaram que essa similaridade está relacionada especialmente com o subgenótipo D3 que é predominante nessas populações, enquanto para os genótipos D1 e D2 essas relações mudam, evidenciando múltiplas fontes do genótipo D na região. Adicionalmente, mostramos que a Itália apresenta uma expansão populacional muito mais antiga que todos os outros lugares amostrados, sugerindo o país como uma fonte deste genótipo na Europa. Como historicamente a região participou de rotas migratórias importantes, isso corrobora com o fato do genótipo D estar distribuído no mundo todo (Kostaki et al. 2018).

Por outro lado, mostramos também a dinâmica diferenciada existente na região do Caribe e Nordeste do Brasil, com predomínio do subgenótipo D4. Esse subgenótipo tem registros raros ao redor do mundo, ocorrendo em alguns países africanos (Candotti et al. 2006; Andernach et al. 2009; Baha et al. 2012; Kwange et al. 2013). Nossos resultados mostraram que esse subgenótipo apresentam certa estruturação formando grupamentos principais que envolvem o Nordeste brasileiro, Cuba, e o Haiti. Trabalhos anteriores tem sugerido possível origem africana para esta cepa, com dispersão durante o período do tráfico de escravos (Banerjee et al. 2014; Barros et al. 2014), o que corrobora com a alta prevalência observada na região caribenha. Em conjunto, nossos resultados salientam a influência de muitas fontes de genótipo D na América Latina, com especial papel da Itália na dispersão do subgenótipo D3.

Por fim, este trabalho comparou a ocorrência de variantes em genes relacionados com resposta ao HBV em duas amostras de Nativos Americanos, relacionando com a suscetibilidade à infecção. Populações Nativas Americanas apresentam hábitos e condições

sociais que propiciam as rotas de transmissão associadas ao HBV e frequentemente apresentam altas prevalências da infecção (Torres 1996; Braga 2004; Godoy et al. 2016). Nossos resultados evidenciaram uma importante associação do gene *RARB* com a suscetibilidade à infecção. RARB tem sido descrito em estudos funcionais como um importante regulador do gene que codifica o receptor específico ao qual HBV se liga para ser internalizado no hepatócito (Tsukuda et al. 2015; Li et al. 2018), além de estar relacionado com ocorrência de desfechos mais graves da doença (Yang et al. 2003; Jung et al. 2010; Zhang et al. 2015; Cortes et al. 2018). Adicionalmente, a presença e funcionalidade do receptor na superfície do hepatócito tem sido descrita como fator limitante para a ocorrência de infecção por HBV (Ni et al. 2014). Desta forma, nossos resultados são sugestivos de que o gene *RARB* é um fator genético importante na suscetibilidade do hospedeiro à infecção por HBV.

CAPÍTULO VI – CONCLUSÕES E PERSPECTIVAS

Este trabalho reuniu dados genéticos e evolutivos de ambos vírus e hospedeiro e buscou contribuir com o melhor conhecimento das linhagens de HBV circulantes na América Latina de três maneiras: através da análise e discussão da melhor topologia proposta para a intrigante história evolutiva dos genótipos americanos F e H de HBV; pelo estabelecimento da contribuição das cepas circulantes na Europa sobre o HBV latino americano; e pela determinação de fatores genéticos de suscetibilidade hospedeira à infecção.

Como perspectivas estão a descrição detalhada de toda a amostra viral isolada previamente de Nativos Americanos com relação ao perfil mutacional do HBV, além da melhor descrição do papel funcional das variantes genéticas de *RARB* indicadas aqui como fatores de risco ao HBV.

CAPÍTULO VII – REFERÊNCIAS BIBLIOGRÁFICAS

Alvarado-Mora M V, Botelho L, Gomes-Gouvêa MS, De Souza VF, Nascimento MC, Pannuti CS, Carrilho FJ and Pinho JR (2011) Detection of Hepatitis B virus subgenotype A1 in a Quilombo community from Maranhão, Brazil. Virol J 8:415. doi: 10.1186/1743-422X-8-415

Alvarado-Mora M V and Pinho JRR (2013) Distribution of HBV genotypes in Latin America. Antivir Ther 18:459–465. doi: 10.3851/IMP2599

Andernach IE, Nolte C, Pape JW and Muller CP (2009) Slave trade and hepatitis B virus genotypes and subgenotypes in Haiti and Africa. Emerg Infect Dis 15:1222–1228. doi: 10.3201/eid1508.081642

Araujo NM, Mello FCA, Yoshida CFT, Niel C and Gomes SA (2004) High proportion of subgroup A′ (genotype A) among Brazilian Isolates of Hepatitis B virus. Arch Virol 149:1383–1395. doi: 10.1007/s00705-003-0269-4

Baha W, Ennaji MM, Lazar F, Melloul M, El Fahime E, El Malki A and Bennani A (2012) HBV genotypes prevalence, precore and basal core mutants in Morocco. Infect Genet Evol 12:1157–1162. doi: 10.1016/j.meegid.2012.04.026

Banerjee P, Mondal RK, Nandi M, Ghosh S, Khatun M, Chakraborty N, Bhattacharya S, Choudhury AR, Banerjee S, Santra A et al. (2014) A rare HBV subgenotype D4 with unique genomic signatures identified in north-eastern india-an emerging clinical challenge? PLoS One. doi: 10.1371/journal.pone.0109425

Bar-Gal GK, Kim MJ, Klein A, Shin DH, Oh CS, Kim JW, Kim TH, Kim SB, Grant PR, Pappo O et al. (2012) Tracing hepatitis B virus to the 16th century in a Korean mummy. Hepatology 56:1671–1680. doi: 10.1002/hep.25852

Barros LMF, Gomes-Gouvêa MS, Kramvis A, Mendes-Corrêa MCJ, dos Santos A, Souza LAB, Santos MDC, Carrilho FJ, de Jesus Domicini A, Pinho JRR et al. (2014) High prevalence of hepatitis B virus subgenotypes A1 and D4 in Maranhão state, Northeast Brazil. Infect Genet Evol 24:68–75. doi: 10.1016/j.meegid.2014.03.007

Beck J and Nassal M (2007) Hepatitis B virus replication. World J Gastroenterol 13:48–64.

Bertolini DA, Gomes-Gouvêa MS, Carvalho-Mello IMVG de, Saraceni CP, Sitnik R, Grazziotin FG, Laurindo JP, Fagundes NJR, Carrilho FJ and Pinho JRR (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. doi: 10.1016/j.meegid.2012.04.009

Block TM, Guo H and Guo JT (2007) Molecular Virology of Hepatitis B Virus for Clinicians. Clin Liver Dis 11:685–706. doi: 10.1016/j.cld.2007.08.002

Braga WSM (2004) [Hepatitis B and D virus infection within Amerindians ethnic groups in the Brazilian Amazon: epidemiological aspects]. Rev Soc Bras Med Trop 37 Suppl 2:9–13.

Candotti D, Opare-Sem O, Rezvan H, Sarkodie F and Allain J-P (2006) Molecular and serological characterization of hepatitis B virus in deferred Ghanaian blood donors with and without elevated alanine aminotransferase. J Viral Hepat 13:715–724. doi: 10.1111/j.1365-2893.2006.00741.x

Chen D-S, Locarnini S and Wallace J (2015) From the big three to the big four. Lancet Infect Dis 15:626–7. doi: 10.1016/S1473-3099(15)00026-2

Chisari F V. and Ferrari C (1995) Hepatitis B Virus Immunopathogenesis. Annu Rev Immunol 13:29–60. doi: 10.1146/annurev.iy.13.040195.000333

Coppola N, Alessio L, Gualdieri L, Pisaturo M, Sagnelli C, Minichini C, Di Caprio G, Starace M, Onorato L, Signoriello G et al. (2017) Hepatitis B virus infection in undocumented immigrants and refugees in Southern Italy: Demographic, virological, and clinical features. Infect Dis Poverty. doi: 10.1186/s40249-016-0228-4

Cortes E, Lachowski D, Rice A, Chronopoulos A, Robinson B, Thorpe S, Lee DA, Possamai LA, Wang H, Pinato DJ et al. (2018) RAR-β is downregulated in HCC & cirrhosis and its expression inhibits myosin-driven activation and durotaxis in hepatic stellate cells. Hepatology. doi: 10.1002/hep.30193

Dandri M and Locarnini S (2012) New insight in the pathobiology of hepatitis B virus infection. Gut 61:i6–i17. doi: 10.1136/gutjnl-2012-302056

Devesa M and Pujol FH (2007) Hepatitis B virus genetic diversity in Latin America. Virus Res 127:177–184. doi: 10.1016/j.virusres.2007.01.004

Dill JA, Camus AC, Leary JH, Di Giallonardo F, Holmes EC and Ng TFF (2016) Distinct Viral Lineages from Fish and Amphibians Reveal the Complex Evolutionary History of Hepadnaviruses. J Virol 90:7920–7933. doi: 10.1128/JVI.00832-16

Ganem D and Prince AM (2004) Hepatitis B Virus Infection — Natural History and Clinical Consequences. N Engl J Med 350:1118–1129. doi: 10.1056/NEJMra031087

Geoghegan JL, Duchêne S and Holmes EC (2017) Comparative analysis estimates the relative frequencies of co-divergence and cross-species transmission within viral families. PLoS Pathog 13:e1006215. doi: 10.1371/journal.ppat.1006215

Gilbert C, Meik JM, Dashevsky D, Card DC, Castoe TA and Schaack S (2014) Endogenous hepadnaviruses, bornaviruses and circoviruses in snakes. Proceedings Biol Sci 281:20141122. doi: 10.1098/rspb.2014.1122

Godoy BA, Alvarado-Mora M V, Gomes-Gouv??a MS, Rebello Pinho JR and Fagundes NJR (2013) Origin of HBV and its arrival in the Americas - The importance of natural selection on time estimates. Antivir Ther 18:505–512. doi: 10.3851/IMP2600

Godoy BA, Gomes-Gouvêa MS, Zagonel-Oliveira M, Alvarado-Mora M V., Salzano FM, Pinho JRR and Fagundes NJR (2016) High prevalence of HBV/A1 subgenotype in native south Americans may be explained by recent economic developments in the Amazon. Infect Genet Evol 43:354–363. doi: 10.1016/j.meegid.2016.06.002

Gomes-Gouvea MS, Soares MCP, Bensabath G, de Carvalho-Mello IMVG, Brito EMF, Souza OSC, Queiroz ATL, Carrilho FJ and Pinho JRR (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 90:2638–2643. doi: 10.1099/vir.0.013615-0

Gómez-Moreno A and Garaigorta U (2017) Hepatitis B Virus and DNA Damage Response: Interactions and Consequences for the Infection. Viruses 9:304. doi: 10.3390/v9100304

Gusatti CS, Costi C, Halon ML, Grandi T, Medeiros AFR, Silva CMD, Gomes SA, Silva MSN, Niel C and Rossetti MLR (2015) Hepatitis B virus genotype D isolates circulating in Chapecó, Southern Brazil, originate from Italy. PLoS One 10:e0135816. doi: 10.1371/journal.pone.0135816

Hartfield M, Murall CL and Alizon S (2014) Clinical applications of pathogen phylogenies. Trends Mol Med 20:394–404. doi: 10.1016/j.molmed.2014.04.002

Huy TTT, Ngoc TT and Abe K (2008) New Complex Recombinant Genotype of Hepatitis B Virus Identified in Vietnam. J Virol 82:5657–5663. doi: 10.1128/JVI.02556-07

Jung JK, Park SH and Jang KL (2010) Hepatitis B virus X protein overcomes the growth-inhibitory potential of retinoic acid by downregulating retinoic acid receptor-b2 expression via DNA methylation. J Gen Virol 91:493–500. doi: 10.1099/vir.0.015149-0

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

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

Kurbanov F, Tanaka Y and Mizokami M (2010) Geographical and genetic diversity of the human hepatitis B virus. Hepatol Res 40:14–30. doi: 10.1111/j.1872-034X.2009.00601.x

Kwange SO, Budambula NLM, Kiptoo MK, Okoth F, Ochwoto M, Oduor M and Kimotho JH (2013) Hepatitis B virus subgenotype A1, occurrence of subgenotype D4, and S gene mutations among voluntary blood donors in Kenya. Virus Genes 47:448–455. doi: 10.1007/s11262-013-0976-1

Lampe E, Mello FCA, do Espírito-Santo MP, Oliveira CMC, Bertolini DA, Gonçales NSL, Moreira RC, Fernandes CAS, Nascimento HCL, Grotto RMT et al. (2017) Nationwide overview of the distribution of hepatitis B virus genotypes in Brazil: A 1000-sample multicentre study. J Gen Virol 98:1389–1398. doi: 10.1099/jgv.0.000789

Lanford RE, Chavez D, Brasky KM, Burns RB and Rico-Hesse R (1998) Isolation of a hepadnavirus from the woolly monkey, a New World primate. Proc Natl Acad Sci U S A 95:5757–61.

Lauber C, Seitz S, Mattei S, Suh A, Beck J, Herstein J, Börold J, Salzburger W, Kaderali L, Briggs JAG et al. (2017) Deciphering the Origin and Evolution of Hepatitis B Viruses by Means of a Family of Non-enveloped Fish Viruses. Cell Host Microbe 22:387–399.e6. doi: 10.1016/j.chom.2017.07.019

Leistner CM, Gruen-Bernhard S and Glebe D (2008) Role of glycosaminoglycans for binding and infection of hepatitis B virus. Cell Microbiol 10:122–133. doi: 10.1111/j.1462-5822.2007.01023.x

Li B, Wang Y, Shen F, Wu M, Li Y, Fang Z, Ye J, Wang L, Gao L, Yuan Z et al. (2018) Identification of Retinoic Acid Receptor Agonists as Potent Hepatitis B Virus Inhibitors via a Drug Repurposing Screen. Antimicrob Agents Chemother AAC.00465-18. doi: 10.1128/AAC.00465-18

Li S, Wang Z, Li Y and Ding G (2017) Adaptive evolution of proteins in hepatitis B virus during divergence of genotypes. Sci Rep 7:1990. doi: 10.1038/s41598-017-02012-8

Li Z, Hou X and Cao G (2015) Is mother-to-infant transmission the most important factor for persistent HBV infection? Emerg Microbes Infect 4:e30. doi: 10.1038/emi.2015.30

Lin HJ, Lai CL, Lau JY, Chung HT, Lauder IJ and Fong MW (1990) Evidence for intrafamilial transmission of hepatitis B virus from sequence analysis of mutant HBV DNAs in two Chinese families. Lancet (London, England) 336:208–12.

Locarnini S (2005) Molecular virology and the development of resistant mutants: Implications for therapy. Semin Liver Dis 25:9–19. doi: 10.1055/s-2005-915645

Locarnini S, Chen DS and Shibuya K (2016) No more excuses: Viral hepatitis can be eliminated. Lancet 387:1703–1704. doi: 10.1016/S0140-6736(16)30295-1

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

Lucifora J, Arzberger S, Durantel D, Belloni L, Strubin M, Levrero M, Zoulim F, Hantz O and Protzer U (2011) Hepatitis B virus X protein is essential to initiate and maintain virus replication after infection. J Hepatol 55:996–1003. doi: 10.1016/j.jhep.2011.02.015

Magnius LO and Norder H (1995) Subtypes, Genotypes and Molecular Epidemiology of the Hepatitis B Virus as Reflected by Sequence Variability of the S-Gene. Intervirology 38:24–34. doi: 10.1159/000150411

McCarthy AE, Weld LH, Barnett ED, So H, Coyle C, Greenaway C, Stauffer W, Leder K, Lopez-Velez R, Gautret P et al. (2013) Spectrum of illness in international migrants seen at geosentinel clinics in 1997-2009, part 2: Migrants resettled internationally and evaluated for specific health concerns. Clin Infect Dis 56:925–933. doi: 10.1093/cid/cis1016

Milich D and Liang TJ (2003) Exploring the Biological Basis of Hepatitis B e Antigen in Hepatitis B Virus Infection. Hepatology 38:1075–1086. doi: 10.1053/jhep.2003.50453

Montenegro RA and Stephens C (2006) Indigenous health in Latin America and the Caribbean. Lancet 367:1859–1869. doi: 10.1016/S0140-6736(06)68808-9

Moura IF, Lopes EP, Alvarado-Mora MV, Pinho JR and Carrilho FJ (2013) Phylogenetic analysis and subgenotypic distribution of the hepatitis B virus in Recife, Brazil. Infect Genet Evol 14:195–199. doi: 10.1016/j.meegid.2012.11.022

Mühlemann B, Jones TC, Damgaard P de B, Allentoft ME, Shevnina I, Logvin A, Usmanova E, Panyushkina IP, Boldgiv B, Bazartseren T et al. (2018) Ancient hepatitis B viruses from the Bronze Age to the Medieval period. Nature 557:418–423. doi: 10.1038/s41586-018-0097-z

Ni Y, Lempp FA, Mehrle S, Nkongolo S, Kaufman C, Fälth M, Stindt J, Königer C, Nassal M, Kubitz R et al. (2014) Hepatitis B and D viruses exploit sodium taurocholate co-transporting polypeptide for species-specific entry into hepatocytes. Gastroenterology 146:1070–1083.e6. doi: 10.1053/j.gastro.2013.12.024

Norder H, Couroucé A-M and Magnius LO (1994) Complete Genomes, Phylogenetic Relatedness, and Structural Proteins of Six Strains of the Hepatitis B Virus, Four of Which Represent Two New Genotypes. Virology 198:489–503. doi: 10.1006/viro.1994.1060

Ozasa A, Tanaka Y, Orito E, Sugiyama M, Kang J-H, Hige S, Kuramitsu T, Suzuki K, Tanaka E, Okada S et al. (2006) Influence of genotypes and precore mutations on fulminant or chronic outcome of acute hepatitis B virus infection. Hepatology 44:326–334. doi: 10.1002/hep.21249

PAHO (2016) Hepatitis B and C in the Spotlight.

Paoli J de, Wortmann AC, Klein MG, Pereira VRZB, Cirolini A, Godoy B, Fagundes NJR, Wolf JM, Lunge VR and Simon D (2018) HBV epidemiology and genetic diversity in an area of high prevalence of hepatitis B in southern Brazil. Brazilian J Infect Dis. doi: 10.1016/j.bjid.2018.06.006

Paraskevis D, Angelis K, Magiorkinis G, Kostaki E, Ho SYW and Hatzakis A (2015) Molecular Phylogenetics and Evolution Dating the origin of hepatitis B virus reveals

higher substitution rate and adaptation on the branch leading to F / H genotypes q. Mol Phylogenet Evol 93:44–54. doi: 10.1016/j.ympev.2015.07.010

Paraskevis D, Magiorkinis G, Magiorkinis E, Ho SYW, Belshaw R, Allain JP and Hatzakis A (2013) Dating the origin and dispersal of hepatitis B virus infection in humans and primates. Hepatology 57:908–916. doi: 10.1002/hep.26079

Petersen J, Thompson AJ and Levrero M (2016) Aiming for cure in HBV and HDV infection. J Hepatol 65:835–848. doi: 10.1016/j.jhep.2016.05.043

Rabe B, Vlachou A, Pante N, Helenius A and Kann M (2003) Nuclear import of hepatitis B virus capsids and release of the viral genome. Proc Natl Acad Sci 100:9849–9854. doi: 10.1073/pnas.1730940100

Raghavan M, Steinrucken M, Harris K, Schiffels S, Rasmussen S, DeGiorgio M, Albrechtsen A, Valdiosera C, Avila-Arcos MC, Malaspinas A-S et al. (2015) Genomic evidence for the Pleistocene and recent population history of Native Americans. Science (80-) 349:aab3884-aab3884. doi: 10.1126/science.aab3884

Rasche A, Souza BF de CD and Drexler JF (2016) Bat hepadnaviruses and the origins of primate hepatitis B viruses. Curr Opin Virol 16:86–94. doi: 10.1016/j.coviro.2016.01.015

Revill P, Testoni B, Locarnini S and Zoulim F (2016) Global strategies are required to cure and eliminate HBV infection. Nat Rev Gastroenterol Hepatol 13:239–248. doi: 10.1038/nrgastro.2016.7

Revill PA and Locarnini SA (2016) New perspectives on the hepatitis B virus life cycle in the human liver. J Clin Invest 126:833–836. doi: 10.1172/JCI86650

Richter C, Beest G Ter, Gisolf EH, Van Bentum P, Waegemaekers C, Swanink C and Roovers E (2014) Screening for chronic hepatitis B and C in migrants from Afghanistan, Iran, Iraq, the former Soviet Republics, and Vietnam in the Arnhem region, the Netherlands. Epidemiol Infect 142:2140–2146. doi: 10.1017/S0950268813003415

Roman S, Tanaka Y, Khan A, Kurbanov F, Kato H, Mizokami M and Panduro A (2010) Occult hepatitis B in the genotype H-infected Nahuas and Huichol native Mexican population. J Med Virol 82:1527–1536. doi: 10.1002/jmv.21846

Ross ZP, Klunk J, Fornaciari G, Giuffra V, Duchêne S, Duggan AT, Poinar D, Douglas MW, Eden JS, Holmes EC et al. (2018) The paradox of HBV evolution as revealed from a 16thcentury mummy. PLoS Pathog 14:e1006750. doi: 10.1371/journal.ppat.1006750

Schinzari V, Barnaba V and Piconese S (2015) Chronic hepatitis B virus and hepatitis C virus infections and cancer: synergy between viral and host factors. Clin Microbiol Infect 21:969–974. doi: 10.1016/J.CMI.2015.06.026

Schweitzer A, Horn J, Mikolajczyk RT, Krause G and Ott JJ (2015) Estimations of worldwide prevalence of chronic hepatitis B virus infection: A systematic review of data published between 1965 and 2013. Lancet 386:1546–1555. doi: 10.1016/S0140-6736(15)61412-X

Simmonds P (2001) Reconstructing the origins of human hepatitis viruses. Philos Trans R Soc B Biol Sci 356:1013–1026. doi: 10.1098/rstb.2001.0890

Souza BF de CD, König A, Rasche A, de Oliveira Carneiro I, Stephan N, Corman VM, Roppert PL, Goldmann N, Kepper R, Müller SF et al. (2018) A novel hepatitis B virus species discovered in capuchin monkeys sheds new light on the evolution of primate hepadnaviruses. J Hepatol 68:1114–1122. doi: 10.1016/j.jhep.2018.01.029

Steinhauer DA and Holland JJ (1986) Direct method for quantitation of extreme polymerase error frequencies at selected single base sites in viral RNA. J Virol 57:219–28.

Stevens CE, Beasley RP, Tsui J and Lee W-C (1975) Vertical Transmission of Hepatitis B Antigen in Taiwan. N Engl J Med 292:771–774. doi: 10.1056/NEJM197504102921503

Suh A, Brosius J, Schmitz J and Kriegs JO (2013) The genome of a Mesozoic paleovirus reveals the evolution of hepatitis B viruses. Nat Commun 4:1791. doi: 10.1038/ncomms2798

Suh A, Weber CC, Kehlmaier C, Braun EL, Green RE, Fritz U, Ray DA and Ellegren H (2014) Early Mesozoic Coexistence of Amniotes and Hepadnaviridae. PLoS Genet 10:e1004559. doi: 10.1371/journal.pgen.1004559

Summers J and Mason WS (1982) Replication of the genome of a hepatitis B--like virus by reverse transcription of an RNA intermediate. Cell 29:403–15.

Sunbul M (2014) Hepatitis B virus genotypes: global distribution and clinical importance. World J Gastroenterol 20:5427–34. doi: 10.3748/wjg.v20.i18.5427

Tatematsu K, Tanaka Y, Kurbanov F, Sugauchi F, Mano S, Maeshiro T, Nakayoshi T, Wakuta M, Miyakawa Y and Mizokami M (2009) A Genetic Variant of Hepatitis B Virus Divergent from Known Human and Ape Genotypes Isolated from a Japanese Patient and Provisionally Assigned to New Genotype J. J Virol 83:10538–10547. doi: 10.1128/JVI.00462-09

Torres C, Piñeiro y Leone FG, Pezzano SC, Mbayed VA and Campos RH (2011) New perspectives on the evolutionary history of hepatitis B virus genotype F. Mol Phylogenet Evol 59:114–22. doi: 10.1016/j.ympev.2011.01.010

Torres JR (1996) Hepatitis B and hepatitis delta virus infection in South America. Gut 38 Suppl 2:S48-55.

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

Tsukuda S, Watashi K, Iwamoto M, Suzuki R, Aizaki H, Okada M, Sugiyama M, Kojima S, Tanaka Y, Mizokami M et al. (2015) Dysregulation of retinoic acid receptor diminishes hepatocyte permissiveness to hepatitis B virus infection through modulation of sodium taurocholate cotransporting polypeptide (NTCP) expression. J Biol Chem 290:5673–5684. doi: 10.1074/jbc.M114.602540

Valaydon ZS and Locarnini SA (2017) The virological aspects of hepatitis B. Best Pract Res Clin Gastroenterol 31:257–264. doi: 10.1016/j.bpg.2017.04.013

van Hemert FJ, van de Klundert MAA, Lukashov V V., Kootstra NA, Berkhout B and Zaaijer HL (2011) Protein X of hepatitis B virus: Origin and structure similarity with the central domain of DNA glycosylase. PLoS One 6:e23392. doi: 10.1371/journal.pone.0023392

Wei Y, Neuveut C, Tiollais P and Buendia M-A (2010) Molecular biology of the hepatitis B virus and role of the X gene. Pathol Biol 58:267–272. doi: 10.1016/j.patbio.2010.03.005

WHO (2017) Global Hepatitis Report, 2017. World Heal Organ. doi: ISBN 978-92-4-156545-5

Yan H and Li W (2015) Sodium taurocholate cotransporting polypeptide acts as a receptor for hepatitis B and D virus. Digestive Diseases. pp 388–396

Yang B, Guo M, Herman JG and Clark DP (2003) Aberrant promoter methylation profiles of tumor suppressor genes in hepatocellular carcinoma. Am J Pathol 163:1101–1107. doi: 10.1016/S0002-9440(10)63469-4

Yuen M-F, Chen D-S, Dusheiko GM, Janssen HLA, Lau DTY, Locarnini SA, Peters MG and Lai C-L (2018) Hepatitis B virus infection. Nat Rev Dis Prim 4:18035. doi: 10.1038/nrdp.2018.35

Zanetti AR, Van Damme P and Shouval D (2008) The global impact of vaccination against hepatitis B: A historical overview. Vaccine 26:6266–6273. doi: 10.1016/j.vaccine.2008.09.056

Zhang C, Chen X, Liu H, Li H, Jiang W, Hou W, McNutt MA, Lu F and Li G (2015) Alpha fetoprotein mediates HBx induced carcinogenesis in the hepatocyte cytoplasm. Int J Cancer 137:1818–1829. doi: 10.1002/ijc.29548

Zhang Z, Filzmayer C, Ni Y, Sültmann H, Mutz P, Hiet M-S, Vondran FWR, Bartenschlager R and Urban S (2018) Hepatitis D virus replication is sensed by MDA5 and induces IFN-β/λ responses in hepatocytes. J Hepatol 69:25–35. doi: 10.1016/j.jhep.2018.02.021

Zhao XL, Yang JR, Lin SZ, Ma H, Guo F, Yang RF, Zhang HH, Han JC, Wei L and Pan X Ben (2016) Serum viral duplex-linear DNA proportion increases with the progression of liver disease in patients infected with HBV. Gut 65:502–511. doi: 10.1136/gutjnl-2014-308989

ANEXO

The Brazilian Journal of
INFECTIOUS DISEASES

www.elsevier.com/locate/bjid

Original article

HBV epidemiology and genetic diversity in an area of high prevalence of hepatitis B in southern Brazil

Juliana de Paoli[a], André Castagna Wortmann[b], Mirelli Gabardo Klein[a], Vagner Reinaldo Zingalli Bueno Pereira[a], Adriana Maria Cirolini[c], Bibiane Armiliato de Godoy[d], Nelson Jurandi Rosa Fagundes[d], Jonas Michel Wolf[a,*], Vagner Ricardo Lunge[a], Daniel Simon[a]

[a] Universidade Luterana do Brasil (ULBRA), Programa de Pós-Graduação em Biologia Celular e Molecular Aplicada à Saúde, Canoas, RS, Brazil
[b] Universidade Federal do Rio Grande do Sul (UFRGS), Programa de Pós-Graduação em Ciências em Gastroenterologia e Hepatologia, Porto Alegre, RS, Brazil
[c] Secretaria Municipal de Saúde de Bento Gonçalves, Serviço de Atendimento Especializado, Bento Gonçalves, RS, Brazil
[d] Universidade Federal do Rio Grande do Sul (UFRGS), Programa de Pós-Graduação em Genética e Biologia Molecular, Porto Alegre, RS, Brazil

ARTICLE INFO

Article history:
Received 26 April 2018
Accepted 22 June 2018
Available online 6 August 2018

Keywords:
HBV infection
Epidemiology
HBV genotypes
Risk factor
Transmission
Cross-sectional studies

ABSTRACT

Background: Hepatitis B virus (HBV) infection is a major public health problem in Brazil. HBV endemicity is usually moderate to low according to geographic regions, and high prevalence of this virus has been reported in people of some specific Brazilian counties, including those with a strong influence of Italian colonization in southern Brazil. Analysis of HBV diversity and identification of the main risk factors to HBV infection are necessary to understand hepatitis B epidemiology in these high prevalence regions in southern Brazil.
Objective: To investigate epidemiological characteristics and HBV genotypes and subgenotypes circulating in a specific city with high HBV prevalence.
Methods: A cross-sectional study was performed with 102 HBV chronically infected individuals, recruited in reference outpatient clinics for viral hepatitis in a city of high HBV prevalence (Bento Gonçalves) in Rio Grande do Sul state, Brazil between July and December 2010. Socio-demographic, clinical and behavior-related variables were collected in a structured questionnaire. HBV serological markers (HBsAg, anti-HBc), viral load, genotypes/subgenotypes and drug resistance were evaluated and comparatively analyzed among all patients.
Results: The HBV infected subjects had a mean age of 44.9 (±12.2) years, with 86 patients (84.3%) reporting to have a family history of HBV infection, 51 (50.0%) to share personal objects, and were predominantly of Italian descendants (61; 64.9%). There was a predominance of genotype D (49/54; 90.7%), but genotype A was also detected (5/54; 9.3%). Subgenotypes D1 (1; 4.7%), D2 (3; 14.3%), and D3 (17; 81.0%) were identified. LAM-resistant mutation (rtM204I) and ADV-resistant mutations (rtA181V) were detected in only one patient each.

* Corresponding author.
E-mail address: jonasmwolf@gmail.com (J.M. Wolf).
https://doi.org/10.1016/j.bjid.2018.06.006

Conclusions: These results demonstrate a pivotal role of intrafamilial transmission for HBV spreading in this population. Furthermore, there is a high prevalence of HBV genotype D in this region.

Introduction

Hepatitis B virus (HBV) infection is one of the most important human diseases with about two billion people infected worldwide, including 240–280 million with chronic hepatitis B.[1] HBV-infected individuals can develop cirrhosis, hepatocellular carcinoma and other hepatic injuries.[2] More than 600,000 people die each year due to clinical hepatic complications by HBV infection.[3]

Taxonomically, HBV belongs to the Hepadnaviridae family and presents a partial double strand DNA genome of approximately 3200 base pairs.[4] Ten HBV genotypes (A to J) have already been described, with nucleotide sequences divergences greater than 8% in the entire viral genome. In Brazil, the most frequent genotypes are A (58.7%), D (23.4%), and F (11.3%), with A more frequent in Southeast, North and Northeast regions, while D predominates in the South region.[5] Most genotypes present subgenotypes that may lead to different diagnostic and clinical profiles.[4]

HBV transmission occurs *via* contact with infected blood and body fluids (WHO 2016). Infection *via* contacts with medical/dental devices,[6,7] piercings use,[8] tattoo,[9] and illicit drugs use (principally injected) have been reported.[10] Further, this transmission may also occur in the family context[11,12] and sharing of personal objects with family members is strongly associated with HBV transmission.[13] In this sense, recent studies have reported important associations between HBV transmission in intrafamilial environmental.[13–15]

HBV vaccine has been used in Brazil in the last 20 years.[16] Immunization programs started in the infant population in the 1990s and gradually covered the rest of the population. Vaccination coverage for HBV reaches approximately 50% of the population and the effectiveness is estimated at 60%.[10,16] Despite elderly patients have higher rates of HBV infection (most of them were infected before the availability of the vaccine), contamination has been demonstrated in all age groups.[16]

HBV therapy is also largely used and aims to prevent progression to more severe clinical conditions as cirrhosis and hepatocellular carcinoma. The most used drugs in the last years were immunomodulators, such as interferon-alpha and pegylated interferon alpha, and nucleos(t)ide analogues, such as lamivudine (LAM), adefovir (ADV), entecavir (ETV), telbivudine (LdT) and tenofovir (TDF).[17] LAM was the main antiviral drug used in the HBV treatment, but this drug was associated with high rates of drug resistance-related mutations in the HBV polymerase.[18] Currently, the treatment of choice is TDF (disoproxil or alafenamide) since it has a high genetic barrier for drug resistance. The combination of therapies is not commonly recommended.[17]

Laboratory diagnosis is usually based on serology. Previous studies demonstrated markers for HBV infection in 12.7% (anti-HBc positive) and 0.7% (HBsAg positive) of the Brazilian population.[16] The highest HBV detection rate in the country (17.2 cases per 100,000 inhabitants) was observed in South Region, but even there the HBV prevalence was low. In this same geographic region, HBV frequency also increases with age: 1.6% anti-HBc positive in the range of 10–19 years old *versus* 11.3% among those aged 20–69 years.[10] In this whole region, some areas have been more studied because of the high HBV prevalence. It is noteworthy that these areas are generally of Italian descendants and with a high prevalence of genotype D.[19–21]

The present report is a cross-sectional study aimed to assess epidemiological risk factors for HBV infection, as well as to determine HBV genotypes/subgenotypes and resistance mutations in a sample from Bento Gonçalves, a city with a high prevalence of this disease in southern Brazil.

Methods

Subjects and data collect

Patients over 18 years old were recruited between July and December 2010 from two reference outpatient clinics for hepatitis in the city of Bento Gonçalves, Rio Grande do Sul state. The inclusion criterion was diagnosis of "chronic hepatitis B" (detectable HBsAg for more than six months in two tests performed before the beginning of the present study). Informed consent was obtained from all patients previous to the inclusion in the study. The study meets all applicable ethical standards for experimentation and research integrity, according to the Declaration of Helsinki.

Socio-demographic and potential risk factors for HBV infection were obtained through a standardized individual questionnaire that was administered by a trained interviewer in a private room. The Alcohol Use Disorders Identification Test (AUDIT), validated in Brazilian Portuguese, was used to screen for alcohol use disorders. A score ≥ 8 was considered alcohol abuse.[22] General clinical and laboratory data, including alanine aminotransferase (ALT) and aspartate aminotransferase (AST), creatinine, albumin, prothrombin time (PT), ascites, and liver biopsies were obtained from medical records, using data from the most recent tests at the time of interview. Liver biopsies were classified according to the

METAVIR scoring system (fibrosis stage on a five-point scale, with F0 = no fibrosis, F1 = portal fibrosis and periportal without septations, F2 = portal and periportal fibrosis with rare septations, F3 = portal fibrosis and periportal with many septations, F4 = cirrhosis; necroinflammatory activity estimating the degree of portal inflammatory lesions and hepatocellular necrosis on a four-point scale, A0 = no activity, A1 = mild, A2 = moderate, A3 = severe). The study was approved by the Ethics Committees of Hospital de Clínicas (Porto Alegre, RS, Brazil), Universidade Luterana do Brasil (Canoas, RS, Brazil) and Hospital Tacchini (Bento Gonçalves, RS, Brazil).

Sampling and laboratory tests

Blood samples (12 mL) were collected in the same moment for all the laboratory analysis. Serological tests were performed by commercially available assays: HBsAg using ETI-MAK-4 (DiaSorin, Saluggia, Italy) and anti-HBc using ETI-AB-COREK PLUS assay (DiaSorin, Sallugia, Italy).

Total DNA was extracted according to previously described methodology.[23] Real-time polymerase chain (PCR) reaction was used for HBV detection and quantification, as previously described.[24] Both negative and positive controls were included in each PCR run. HBV-DNA positive samples were further submitted to genotyping by amplification and sequencing of a 359 bp fragment of the reverse transcriptase (P gene). The primers pairs used for this PCR were: 5′-CAATGTGGWTAYCCTGCYTTAATGCC-3′ and 5′-GCACAGCCTAGCWGCCA TGG-3′. The PCR conditions consisted of 40 cycles of 94 °C for 15 s, 60 °C for 30 s, 72 °C for 120 s.

Twenty-one HBV positive samples presenting viral load higher than 2000 IU/mL were further submitted to sequencing of a larger fragment of 590 bp encompassing a region of the reverse transcriptase (P gene) and the overlapping S gene (surface antigen HBsAg gene). This region was selected for analysis of resistance to antivirals and was amplified by nested PCR. The primers pairs used were 5′-CASTCATCCWCAGGCMATGCAGTGGA-3′ and 5′-GGGTTGCGTCAGCAAACACTTGGC-3′ in the first round; and 5′-CATCCTGCTGCTATGCCTCATCTTC-3′ and 5′-ATDCKTTGACADACTTTCCARTCAAT-3′ in the second round of amplification. The first-round PCR consisted of 25 cycles of 94 °C for 15 s, 55 °C for 30 s, 72 °C for 120 s. The second-round PCR consisted of 35 cycles of 94 °C for 15 s, 60 °C for 30 s, 72 °C for 120 s.

Nucleotide sequences were obtained by Data Collection v1.0.1 (Applied Biosystems) and electropherograms were analyzed with Sequencing Analysis v.5.3.1. software (Applied Biosystems) and edited in SeqMan (DNAStar, Madison, WI, USA). Nucleotide and amino acids sequences were aligned by the MAFFT method[25] and phylogenies were assembled by Geneious 9.1.2 (Geneious, Inc) software. Neighbor-joining method[26] was used to reconstruct phylogenetic trees, which were performed by comparing sequences obtained in this study with Genbank data representative of other genotypes and subgenotypes (NCBI Genbank: https://www.ncbi.nlm.nih.gov/genbank/). Genetic distances were calculated using General Time Reversible (GTR + I + G) nucleotide substitution model, estimated using jModeltest v.2.1.4.[27] Bootstrap analysis with 1000 replicates were performed to test the reliability of the tree with values ≥ 60 indicated on the branches.

In the resistance analysis, rtM204I/V was defined as the signature of LAM-resistant mutations (LAM-R) and also encompassing resistance to LdT (LdT-R), rtA181V and rtN236T were defined as the signature of ADV-resistant mutations (ADV-R) and rtT184A/C/F/G/I/L/M/S, rtS202C/G/I and rtM250I/L/V were defined as the signature ETV-resistant mutations (ETV-R).[28]

Statistical analysis

Data were analyzed using the Statistical Package for Social Sciences (SPSS, version 18.0, Chicago, IL). The distributions of the quantitative variables were evaluated by the Kolmogorov–Smirnov test with Lilliefors correction. Variables with normal distribution were compared by Student's *t*-test and those with non-parametric distribution were compared by Mann–Whitney test. Categorical variables were compared by Chi-Square or Fisher's Exact Test, according to the recommendation. All statistical tests in this study were two-sided and p-values < 0.05 were considered as statistically significant.

Results

The epidemiologic profile of HBV patients

One hundred and two HBV chronically infected patients were included in the study. The mean age was 44.9 ± 12.2 years and there were 55 males (53.9%) and 47 females (46.1%). No significant differences between sexes were observed in most socio-demographic variables, excepting higher level of schooling and less than five siblings (Table 1). Most patients were married (n = 79, 77.5%), lived in rural area of the municipality (n = 68, 66.7%) during childhood, self-defined white ethnic group (n = 76, 74.5%), and were predominantly of Italian descendants (n = 61, 64.9%) (Table 1). There was a great heterogeneity related to occupation of the individuals, including the economic segments of industry/commerce (n = 82, 80.4%), farming (n = 8, 7.8%), and housewife/housemaid (n = 6, 5.9%).

Additionally, 90.2% (n = 92) of patients reported normal delivery, 50.0% (n = 51) shared personal objects, 49.0% (n = 50) previous use of glass syringe, 21.6% (n = 22) blood transfusion, 15.7% (n = 16) sexually transmitted infection (STI), 7.8% (n = 8) tattoo, and 4.9% (n = 5) body piercing. HBV patients reported different means of infection, such as exposure to sharp objects (n = 20, 19.6%), sexual intercourse (n = 16, 15.7%), vertical transmission (n = 16, 15.7%), blood transfusion (n = 11, 10.8%), and syringe sharing (n = 8, 7.8%) (Table 2). The family history of HBV infection was also analyzed: 86 (84.3%) patients referred having HBV-positive family members, with 42 (41.2%) patients reporting a history of HBV infection in siblings, 28 (27.5%) in the mother, and 11 (10.8%) in the father.

The frequencies of alcohol and illicit drugs use were also evaluated (Table 2). Men presented higher scores in the AUDIT test (3.4 ± 4.3) than women (1.40 ± 4.1), with a consequent higher frequency of alcohol use disorder (p < 0.01). Use of illicit drugs was also reported (6.9% for smoked drugs, 1.9% for

Table 1 – Socio-demographic characteristics of HBV patients stratified by sex (Bento Gonçalves, RS, Brazil, 2010).

Variables	Total ($n=102$) n (%)	Men ($n=55$) n (%)	Women ($n=47$) n (%)	p-Value[a]
Age (mean ± SD)	44.9 ± 12.2	46.3 ± 12.8	43.3 ± 11.3	0.22
Level of schooling				0.01
Elementary school or less	55 (53.9)	38 (69.1)	17 (36.2)	
High school	47 (46.1)	17 (30.9)	30 (63.8)	
Marital status				0.30
Married	79 (77.5)	44 (80.0)	35 (74.5)	
Not married	23 (22.5)	11 (20.0)	12 (25.5)	
Place of residence in childhood				0.48
Rural area	68 (66.7)	35 (63.6)	33 (70.2)	
Urban	34 (33.3)	20 (36.4)	14 (29.8)	
Number of siblings[b]				0.01
>5	46 (45.1)	31 (56.3)	15 (31.9)	
≤5	56 (54.9)	24 (43.7)	32 (68.1)	
Ethnic group[b]				0.16
Not white	15 (14.7)	10 (18.2)	4 (8.5)	
White	76 (74.5)	39 (70.9)	37 (78.7)	
Occupation				–
Farmer	8 (7.8)	8 (14.5)	0 (0.0)	
Industry/commerce	82 (80.5)	44 (80.0)	38 (80.9)	
Housewife/housemaid	6 (5.9)	0 (0.0)	6 (12.8)	
Retired	1 (0.9)	0 (0.0)	1 (2.1)	
Other	5 (4.9)	3 (5.5)	2 (4.3)	
Ancestry				
Number of grandparents Italians[b]				
Four	61 (64.9)	32 (58.2)	29 (61.7)	0.63
Tree	8 (8.5)	5 (9.1)	3 (4.5)	
Two	10 (10.6)	2 (3.6)	8 (17.0)	
One	2 (2.1)	1 (1.8)	1 (2.1)	
None	13 (13.8)	10 (18.2)	3 (6.4)	

[a] Pearson's chi-square or Fisher's exact test (qualitative variables) and Student's t-test for independent samples (quantitative variables).
[b] Totals do not coincide due to lack of data from certain participants in the study.

sniffed, and 0.9% for injected), but they did not present any significant difference between sexes (all with p-values > 0.05) (Table 2).

HBV detection, quantitation and genotyping

HBV-DNA was detected in 54 patients (52.9%) who presented a viral load mean of 3.0 ± 1.7 $\log_{10}$ IU/mL. Thirty-three patients had low viral load (<2000 IU/mL), including six in active treatment against HBV (four using LAM, two TDF plus LAM). Among the remaining 21 patients with a viral load > 2000 IU/mL, six were also on antiviral treatment against HBV (three using -interferon, three LAM). Thirteen of these 21 patients underwent liver biopsy and the results demonstrated that seven (53%) did present no fibrosis, five had mild fibrosis, and only one presented cirrhosis. Necroinflammatory activity (indicating hepatocellular necrosis) was mild in seven, moderate in four, and severe in two patients (Table 3).

A partial region of the HBV polymerase gene (in a region coding for the reverse transcriptase) of 359 bp was sequenced in all the 54 patients. The phylogenetic analysis demonstrated that 49 (90.7%) patients were infected with genotype D and five (9.4%) with genotype A (Fig. 1). An additional HBV genotyping analysis was carried out sequencing a larger fragment of the 590 bp encompassing *P/S* gene in the 21 HBV-infected patients with a viral load > 2000 IU/mL (all infected with genotype D). The phylogenetic analysis demonstrated that subgenotype D3 was the most frequent ($n=17$, 81.0%), but subgenotypes D2 ($n=3$, 14.3%) and D1 ($n=1$, 4.7%) were also detected (Fig. 2). Of these patients, 13 (61.9%) were male. In addition, Italian ancestry was predominant, of which 13 (61.9%) had four grandparents of Italian origin (Table S1).

Mutation patterns

Amino acid substitutions in the *P/S* gene were evaluated in the 21 patients with a viral load > 2000 IU/mL (Fig. 3). A total of 26 amino acid positions presented modifications in comparison to a reference sequence (NCBI GenBank accession: X69798). The majority of the patients presented the following modifications: Y135S in 17 (80.9%) patients and I266V in 14 (66.7%) patients. Interestingly, three of the four patients without any of these modifications presented four other specific polymorphisms: L122F, H126R, P130Q and E263D. LAM-resistant mutation rtM204I was observed in only one patient (BG-96), who was on LAM treatment. This patient also presented the compensatory mutation rtL180M. The rtA181V, defined as the signature of ADV-resistant mutations, was also observed in one patient (BG-43). This patient had already been treated with lamivudine and adefovir.

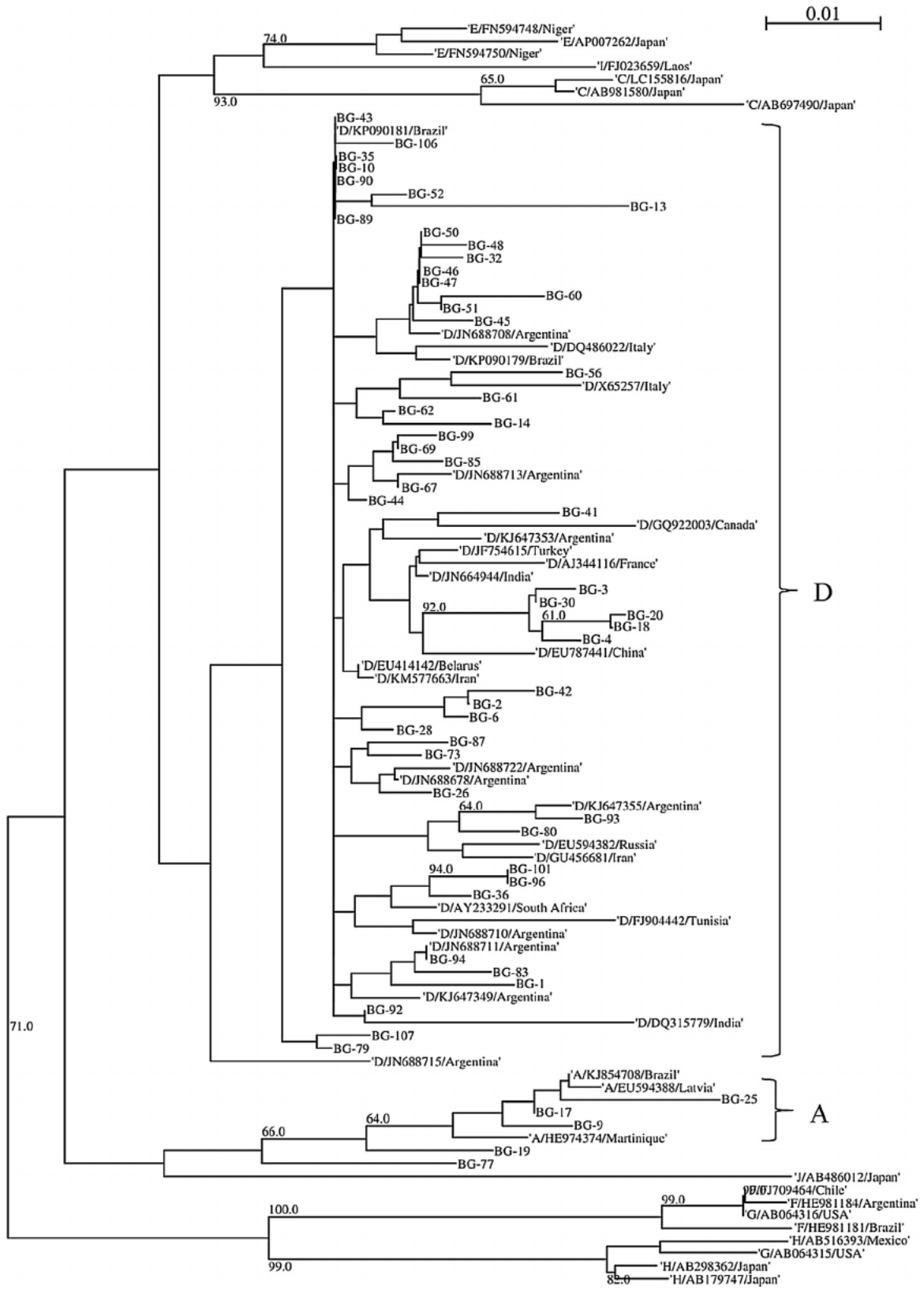


Fig. 1 – Phylogenetic tree of gene fragment (359 bp) of 54 patients with hepatitis B infection in Bento Gonçalves city, Rio Grande do Sul State, 2010. The sequences evaluated in this study are represented by the acronym BG (Bento Gonçalves) followed by the identification number of the sample. Sequences obtained on GenBank are demonstrated by the genotype information followed by the sequence accession number. The numbers at each node correspond to bootstrap values (greater than 60%) obtained with 1000 replicates. The scale bar indicates the genetic distances.

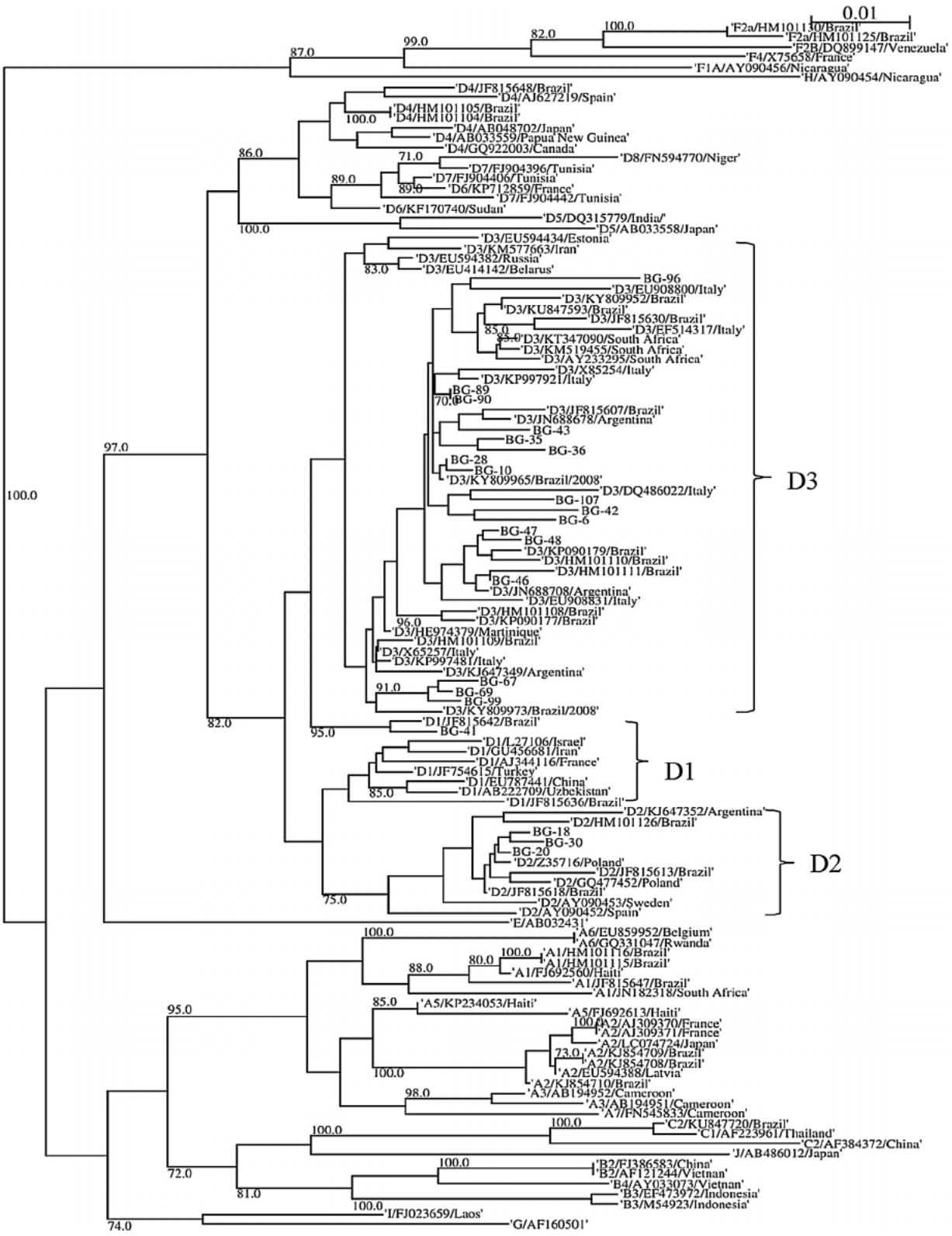


Fig. 2 – Phylogenetic tree of / gene fragment (590 bp) of 21 patients with active hepatitis B infection in Bento Gonçalves city, Rio Grande do Sul State, 2010. The sequences evaluated in this study are represented by the acronym BG (Bento Gonçalves) followed by the identification number of the sample. Sequences obtained on GenBank are demonstrated by the subgenotype information followed by the sequence accession number. The numbers at each node correspond to bootstrap values (greater than 60%) obtained with 1000 replicates. The scale bar indicates the genetic distances.

Table 2 – Risk factors for HBV infection stratified by sex (Bento Gonçalves, RS, Brazil, 2010).

Variables	Total ($n=102$) n (%)	Men ($n=55$) n (%)	Women ($n=47$) n (%)	p-Value[a]
History of siblings infected (HBV)	42 (41.2)	20 (36.4)	22 (46.8)	0.28
History of mother infected (HBV)	28 (27.5)	15 (27.3)	13 (27.7)	0.70
History of father infected (HBV)	11 (10.8)	7 (12.7)	4 (8.5)	0.52
Type of delivery[b]				0.06
Normal	92 (90.2)	49 (89.1)	43 (91.5)	
Cesarean	4 (3.9)	0 (0)	4 (8.5)	
Previous STI	16 (15.7)	11 (20.0)	5 (10.6)	0.18
Sharing of personal objects	51 (50.0)	18 (32.7)	33 (70.2)	<0.01
Blood Transfusion history	22 (21.6)	15 (27.3)	7 (14.9)	0.13
Tattoo	8 (7.8)	2 (3.6)	6 (12.8)	0.14
Body piercing	5 (4.9)	0 (0)	5 (10.6)	0.02
Previous use of glass syringe	50 (49.0)	28 (50.9)	22 (46.8)	0.58
Possible forms of infection				0.88
Sexual intercourse	16 (15.7)	8 (14.5)	8 (17.0)	
Blood transfusion	11 (10.8)	7 (12.7)	4 (8.5)	
Sharp objects	20 (19.6)	9 (16.4)	11 (23.4)	
Syringe sharing	8 (7.8)	4 (7.3)	4 (8.5)	
Birth (vertical transmission)	16 (15.7)	9 (16.4)	7 (14.9)	
Other forms	30 (28.0)	13 (23.6)	17 (36.2)	
AUDIT Score (Mean ± SD)	2.5 ± 4.3	3.4 ± 4.3	1.40 ± 4.1	<0.01
Alcohol use disorder				<0.01
Yes	14 (13.7)	12 (21.8)	2 (4.3)	
Smoked drugs use				0.29
Yes	7 (6.9)	5 (9.1)	2 (4.3)	
Sniffed drugs use				0.30
Yes	2 (1.9)	2 (3.6)	0 (0)	
Injected drugs use				0.54
Yes	1 (0.9)	1 (1.8)	0 (0)	

STI, sexually transmitted disease; AUDIT, Alcohol Use Disorders Identification Test.

[a] Pearson's chi-square or Fisher's exact test (qualitative variables) and Student's t-test for independent samples (quantitative variables).

[b] Totals do not coincide due to lack of data from certain participants in the study.

Table 3 – Clinical characteristics of HBV patients stratified by sex (Bento Gonçalves, RS, Brazil, 2010).

Variables	Total n (%)	Men ($n=55$) n (%)	Women ($n=47$) n (%)	p-Value[a]
PCR				0.45
Positive	54 (52.9)	31 (56.4)	23 (48.9)	
Negative	48 (47.1)	24 (43.6)	24 (51.1)	
Diagnostic time in years (mean ± SD)	6.5 ± 5.8	6.8 ± 6.1	6.0 ± 5.5	0.56
HBV genotypes				0.28
A	5 (9.3)	4 (12.9)	1 (4.3)	
D	49 (90.7)	27 (87.1)	22 (95.7)	
Viral load (mean log10 IU/mL ± SD)	3.0 ± 1.7	3.0 ± 2.0	3.0 ± 2.0	0.97
HBV treatment	25 (24.5)	16 (29.1)	9 (19.1)	0.28

PCR, polymerase chain reaction.

[a] Pearson's chi-square or Fisher's exact test (qualitative variables) and Student's t-test for independent samples (quantitative variables).

Discussion

HBV infection is a serious health problem in Brazil. Recent data demonstrated a prevalence rate higher than the average in some cities and regions in South Brazil.[19,21] Bento Gonçalves is one of these cities, with an incidence of 41.5 cases per 100,000 inhabitants.[29] The whole city is geographically located in the mountains of the Rio Grande do Sul state, a region colonized by European immigrants, mainly Italians. Therefore, socio-demographic characteristics and risk factors were evaluated to identify the frequency of epidemiological characteristics associated with HBV infection in this population.

In this study, the socio-demographic characteristics of HBV infected patients were having elementary school or less (53.9%), be married (77.5%), living in rural area (66.7%), having ≥ 5 siblings (45.1%), of white skin color (74.5%), and of Italian ancestry (four grandparents) (64.9%). Further, considering risk factors, the patients evaluated had 41.2% of siblings, 27.5% of mothers, and 10.8% of fathers HBV infected. STI, blood history of transfusion, tattoo, body piercing, and previous use of glass syringe were identified in 15.7%, 21.6%, 7.8%,

	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2
	0	1	2	2	2	2	3	3	4	4	4	6	8	8	0	0	1	1	2	2	3	3	4	6	6	6
	5	8	2	3	6	9	0	5	1	2	5	4	0	1	4	8	5	9	1	9	3	8	5	3	6	7
Ref seq.	S	N	L	N	H	M	P	Y	Y	V	L	L	L	A	M	V	Q	S	F	L	I	N	Y	E	I	Q
BG-6	.	.	.	.	.	.	.	S	.	.	.	.	.	.	.	.	.	.	.	.	.	.	R	.	V	.
BG-10	.	.	.	.	.	.	.	S	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	V	.
BG-18	.	.	F	.	R	.	Q	.	.	.	.	.	M	.	.	.	.	.	.	.	.	.	.	D	.	.
BG-20	.	.	F	.	R	.	Q	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	D	.	.
BG-28	.	.	.	.	.	.	.	S	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	V	.
BG-30	.	.	F	.	R	.	Q	.	.	.	.	.	.	.	.	.	.	.	.	.	.	H	.	D	.	.
BG-35	.	.	.	.	.	.	.	S	H	.	M	.	.	.	.	A	H	A	Y	V	.	.	.	.	V	.
BG-36	.	D	.	.	.	.	.	S	.	.	.	.	.	.	.	.	.	A	Y	.	.	.	.	.	V	H
BG-41	T	.	.	D	.	.	Q	S	.	.	.	M	.	.	.	.	.	.	.	.	V	.	.	D	.	.
BG-42	.	.	I	.	.	.	.	S	.	.	.	.	.	.	.	.	.	.	.	.	.	D	.	.	V	H
BG-43	.	T	I	.	.	.	.	S	.	.	.	.	.	V	.	.	.	.	Y	.	.	.	.	.	V	.
BG-46	.	.	.	.	.	.	.	S	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	V	.
BG-47	.	.	.	.	.	.	.	S	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	V	.
BG-48	.	.	.	.	.	.	.	S	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	V	.
BG-67	.	.	.	.	.	.	.	S	.	E	.	.	.	.	.	.	.	.	.	.	V	.	.	.	.	.
BG-69	.	.	.	.	.	.	.	S	.	.	.	.	.	.	.	.	.	.	.	.	V	.	.	.	.	.
BG-89	.	.	.	.	.	.	.	S	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	V	.
BG-90	.	.	.	.	.	.	.	S	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	V	.
BG-96	.	.	.	.	.	L	.	N	.	.	.	.	M	.	I	.	.	.	.	.	.	.	.	.	V	.
BG-99	.	.	.	.	.	.	.	S	.	.	.	.	.	.	.	.	.	.	.	.	V	.	.	.	.	.
BG-107	.	.	.	.	.	.	.	S	.	.	.	.	.	.	.	.	S	.	.	.	.	.	.	.	V	.

Fig. 3 – Amino acids modifications in comparison to the reference sequence X69798.

4.9%, and 49.0% of the sample, respectively. Smoked drugs use was the most frequently reported (6.9%). Additionally, sharp objects use (19.6%) and vertical transmission (15.7%) were the most reported means of HBV infection. Importantly, there was a high frequency of HBV-infected family members among the study sample. In a previous study, conducted in a city geographically and culturally close to Bento Gonçalves (Caxias do Sul), important associations were detected between intrafamilial transmission of HBV,[30] a fact that justifies our findings. Evidence also corroborates this type of transmission in Northern Brazil.[13] Currently, phylogenetic approaches reinforce the context of intrafamilial transmission in different studies.[14,15,31]

Over 50% of patients had at least one relative infected with HBV demonstrating the importance of HBV transmission in the family environment, mainly in childhood as previously observed in southern Brazil.[30] In addition, there is evidence of a high frequency of HBV infection in siblings (75%) born to HBsAg-positive mothers (p <0.01) as observed in other study.[13]

Alcohol consumption was also investigated in the present study. We observed that 13.7% of patients had alcohol use disorder by AUDIT scores. Considering the additional damage that alcohol ingestion may produce in patients with chronic hepatitis B, these findings highlight the importance of identifying the pattern of alcohol ingestion in patients with chronic liver diseases.[32] In this study, frequencies for illicit drugs use of 6.9% for smoked drugs, 1.9% for sniffed drugs, and 0.9% for injected drugs were observed. Alcohol consumption and drugs use (especially injected) had already been identified as relevant risk factors for HBV infection.[10] In Brazil, association between heavy alcohol consumption and HBV infection was previously seen in one region (Southeast), but not in others (Central-West, North, Northeast, and South).[10] Further, in our country, illicit drugs use has also not been associated with HBV infection in a nationwide study.[33] However, sniffed and inhaled drugs was shown to be associated with HBV in South Brazil in a more recent investigation.[10] Finally, in the Southern region of the country (Caxias do Sul City), HBV infection was not associated with neither illicit drugs or alcohol consumption.[30]

In this study, 15.7% of the sample had a history of STI. This classical risk factor (STI) was not shown to be associated to HBV infection in Southern Brazil in a recent study.[30] On the other hand, other studies found an association between the history of STI and infection for HBV in a multicentric population-based study in Central-West, Northeast and Fed-

eral Districts of Brazil,[33] and in population-based survey in North, Southwest, and South of Brazil.[10]

The frequency of body piercing and tattoo were 10.6% and 7.8%, respectively. Although these esthetic features have been associated with HBV infection,[8,9,33] studies in Southern Brazil have not detected any association.[10,30] In this study, 49.0% and 21.6% of HBV-infected people had previous use of glass syringe and blood transfusion, respectively. These factors were shown to be associated with HBV infection in a previous study in Southern Brazil.[10,33] Further, in this study, the frequency of share of personal objects in the family environment was of 50.0%. This variable was associated with HBV in other previous studies.[13,30,34] Generally, sharing razor, cutlery, face towels, and toothbrush with one infected individual is strongly associated with HBV transmission.[10,13,30]

Additionally, we evaluated the frequency of HBV genotypes among chronically infected patients in Bento Gonçalves, Southern Brazil. Previous studies reported the major prevalence of genotypes A, D, and F among Brazilians.[5,19,21] Our results revealed marked predominance of genotype D, with a minority of patients infected with genotype A. Moreover, no other genotype was detected in the present study. The high prevalence of genotype D is in agreement with studies conducted in Southern Brazil.[19,21] However, this high prevalence of genotype D in Southern Brazil differs from the findings of the whole country (A is the most frequent genotype).[5] This regional epidemiological difference must be taken into account because of the continental extent of the country.

Phylogenetic studies have been able to trace geographic routes and speed of dissemination of HBV genotypes and subgenotypes.[35] Socio-demographic characteristics influence these processes, with the population profiles defining the migration and speed of HBV dissemination.[19] These variables are useful for the characterization of risk factors for hepatitis B and definition of the clinical evolution of the disease and responses to antivirals.[15,31,36] In the present study, the subgenotype D3 (81.0%) was the most frequent, followed by D2 (14.3%), and only one (4.7%) case of D1. The same scenario of D3 subgenotype was observed in cities with wide Italian immigration in southern[19,21] and southeastern Brazil.[15] In addition, the predominance of D3 and D2 subgenotypes was reported in the Missiones region of Argentina, which has a high frequency of Eurodescendant inhabitants.[36]

The results obtained also allow us to observe that D subgenotype samples were divided into seven branches in the phylogenetic tree, of which they have a high genetic identity, probably having the same evolutionary ancestry of South Europe. In this sense, studies have reported that D3 subgenotype is largely prevalent in Italy.[37,38] One could than hypothesize that this subgenotype probably emerged from this region to the Southern of Brazil, especially in the historical period of immigration.

HBV genotyping and sequencing studies provide a better understanding of the risk factors for infection, prognosis and treatment for hepatitis B.[15,31,35,36] Phylogenetic analyses conducted in different regions of the world have demonstrated that migratory processes influence the distribution of HBV genotypes and subgenotypes.[35] In addition, phylogenetic inferences allow the identification of intrafamilial transmission of HBV, due to the similarities of the viral sequences found in relatives sharing the same domestic environments, as observed in different studies.[14,15,31] However, future studies that evaluate HBV phylodynamics and phylogeographic will be useful for the understanding of viral migration and transmission routes in Southern Brazil.

In the present study, the most frequently detected amino acid substitutions in the RT enzyme were Y135S (80.9%) and I266V (66.7%). However, there is no evidence in the literature that Y135S and I266V mutations are related to antiviral resistance. Further, LAM-resistant mutation rtM204I was observed in only one patient, but this patient also presented the compensatory mutation rtL180M. The rtA181V, defined as the signature of ADV-resistant mutations, was also observed in another patient. In this sense, is important to identify antiviral-resistant mutations when using nucleos(t)ide analogues therapy.[39] Some HBV mutants have the capacity of exhibiting resistance to antivirals, enhanced virulence, facilitated cell attachment or alteration of epitopes which are important in host immune response.[40]

Particularly, in our study, ethnicity and familial issues seem to have influence in the distribution of genotype D. Approximately 60% of patients reported that the four grandparents were of Italian ancestry, while 18% reported having at least two grandparents of Italian ancestry. These findings support the hypothesis that genotype D may have been disseminated in this region by Italian immigrants who arrived in Rio Grande do Sul in the 19th century,[19,21] since there are studies that show high prevalence of this genotype in Southern Europe including Mediterranean countries, especially in Italy.[41,42] Different studies suggested that the high prevalence of genotype D in the Southeast of Brazil is associated with Mediterranean immigrants.[19,21] Novel studies are necessary to observe if this landscape is indeed a specific epidemiological characteristic of the other counties in the South region of the country.

Conclusions

In this study, there was a predominance of HBV genotype D (mainly subgenotype D3) among chronically infected patients in the city of Bento Gonçalves, suggesting an association between Italian immigration and the spread of this infection. Moreover, our results suggest an important role of intrafamilial transmission of HBV infection in the study sample.

Authorship

J. Paoli, A.C. Wortmann, D. Simon, N.J.R. Fagundes, and V.R. Lunge designed the study and wrote the protocol. J. Paoli, M.G. Klein, A. Cirolini, and B. Godoy managed to recruit participants and performed the lab work. J.M. Wolf, D. Simon, V.R. Lunge, and V.R.Z.B. Pereira performed the data statistical analyses. J.M. Wolf, D. Simon, V.R. Lunge wrote the first draft of the manuscript and contributed to literature review and discussion of results. All authors contributed to and have approved the final manuscript.

Funding

This research was funded by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, research grant 559598/2009-2).

Conflicts of interest

The authors declare no conflicts of interest.

Acknowledgments

The authors would like to thank the patients for their collaboration to make this study possible.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.bjid.2018.06.006.

REFERENCES

1. WHO, World Health Organization, Available from: http://www.who.int/mediacentre/factsheets/fs204/en/, 2016 [accessed 27.03.18].
2. Trepo C, Chan HL, Lok A. Hepatitis B virus infection. Lancet. 2014;384:2053–63.
3. Niederau C. Chronic hepatitis B in 2014: great therapeutic progress, large diagnostic deficit. World J Gastroenterol. 2014;20:11595–617.
4. Tong S, Revill P. Overview of viral replication and genetic variability. J Hepatol. 2016;64:S4–16.
5. Lampe E, Mello FCA, do Espírito-Santo MP, et al. Nationwide overview of the distribution of hepatitis B virus genotypes in Brazil: a 1000-sample multicenter study. J Gen Virol. 2017;98:1389–98.
6. Diercke M, Monazahian M, Petermann H, et al. Hepatitis B outbreak in a nursing home associated with reusable lancet devices for blood glucose monitoring, Northern Germany 2010. J Med Virol. 2015;87:583–8.
7. Bârlean L, Săveanu I, Balcoş C. Dental patients' attitudes towards infection control. Rev Med Chir Soc Med Nat Iasi. 2014;18:524–7.
8. Yang S, Wang D, Zhang Y, et al. Transmission of hepatitis B and C virus infection through body piercing: a systematic review and meta-analysis. Medicine. 2015;94:e1893.
9. Jafari S, Buxton JA, Afshar K, Copes R, Baharlou S. Tattooing and risk of hepatitis B: a systematic review and meta-analysis. Can J Public Health. 2012;103:207–12.
10. Ximenes RA, Figueiredo GM, Cardoso MR, et al. Population-based multicentric survey of hepatitis B infection and risk factors in the North, South, and Southeast Regions of Brazil, 10–20 years after the beginning of vaccination. Am J Trop Med Hyg. 2015;93:1341–8.
11. Shepard CW, Simard EP, Finelli L, et al. Hepatitis B virus infection: epidemiology and vaccination. Epidemiol Rev. 2006;28:112–25.
12. WHO, World Health Organization. http://apps.who.int/iris/bitstream/10665/154590/1/9789241549059_eng.pdf, 2015 [accessed 27.03.18].
13. Lobato C, Tavares-Neto J, Rios-Leite M, et al. Intrafamilial prevalence of hepatitis B virus in Western Brazilian Amazon region: epidemiologic and biomolecular study. J Gastroenterol Hepatol. 2006;21:863–8.
14. Ragheb M, Elkady A, Tanaka Y, et al. Multiple intra-familial transmission patterns of hepatitis B virus genotype D in north-eastern Egypt. J Med Virol. 2012;84:587–95.
15. Chachá SG, Gomes-Gouvêa MS, Malta FM, et al. Distribution of HBV subgenotypes in Ribeirão Preto, Southeastern Brazil: a region with history of intense Italian immigration. Braz J Infect Dis. 2017;21:424–32.
16. Souto FJ. Distribution of hepatitis B infection in Brazil: the epidemiological situation at the beginning of the 21st century. Rev Soc Bras Med Trop. 2016;49:11–23.
17. European Association for the Study of the Liver. European Association for the Study of the Liver. EASL 2017 Clinical Practice Guidelines on the management of hepatitis B virus infection. J Hepatol. 2017, pii:S0168-8278(17)30185-X. http://www.easl.eu/medias/cpg/management-of-hepatitis-B-virus-infection/English-report.pdf [accessed 27.03.18].
18. Lim YS. Management of antiviral resistance in chronic hepatitis B. Gut Liver. 2017;15:189–95.
19. Bertolini DA, Gomes-Gouvêa MS, Carvalho-Mello IM, et al. Hepatitis B virus genotypes from European origin explains the high endemicity found in some areas from southern Brazil. Infect Genet Evol. 2012;12:1295–304.
20. Menegol D, Spilki FR. Seroprevalence of hepatitis B and C markers at the population level in the municipality of Caxias do Sul, southern Brazil. Braz J Microbiol. 2014;44:1237–40.
21. Gusatti CS, Costi C, Halon ML, et al. Hepatitis B virus genotype D isolates circulating in Chapecó, Southern Brazil, Originate from Italy. PLOS ONE. 2015;10:e0135816.
22. Mendoza-Sassi RA, Béria JU. Prevalence of alcohol use disorders and associated factors: a population-based study using AUDIT in southern Brazil. Addiction. 2003;98:799–804.
23. Boom R, Sol CJ, Salimans MM, et al. Rapid and simple method for purification of nucleic acids. J Clin Microbiol. 1990;28:495–503.
24. Welzel TM, Miley WJ, Parks TL, et al. Real-time PCR assay for detection and quantification of hepatitis B virus genotypes A to G. J Clin Microbiol. 2006;44:3325–33.
25. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30:772–80.
26. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol. 1987;4:406–25.
27. Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing. Nat Methods. 2012;9:772.
28. Reijnders JG, Pas SD, Schutten M, de Man RA, Janssen HL. Entecavir shows limited efficacy in HBeAg-positive hepatitis B patients with a partial virologic response to adefovir therapy. J Hepatol. 2009;50:674–83.
29. Gonçalves B, Available from: http://www.bentogoncalves.rs.gov.br/downloads/Saude/Plano-Municipal-de-Saude-2014-2017.pdf [accessed 27.03.18].
30. Pereira VRZB, Wolf JM, Luz C, et al. Risk factors for hepatitis B transmission in South Brazil. Mem Inst Oswaldo Cruz. 2017;112:544–50.
31. Ciccozzi M, Chaouch H, Lo Presti A, et al. Evolutionary dynamics of HBV-D7 subgenotype in Tunisia. J Med Virol. 2017;89:469–75.
32. Gitto S, Golfieri L, Caputo F, Grandi S, Andreone P. Multidisciplinary view of alcohol use disorder: from a psychiatric illness to a major liver disease. Biomolecules. 2016;6:11.

33. Pereira LM, Martelli CM, Merchán-Hamann E, et al. Population-based multicentric survey of hepatitis B infection and risk factor differences among three regions in Brazil. Am J Trop Med Hyg. 2009;81:240–7.
34. Clemente CM, Carrilho FJ, Pinho JR, et al. A phylogenetic study of hepatitis B virus in chronically infected Brazilian patients of Western and Asian descent. J Gastroenterol. 2009;44:568–76.
35. Pourkarim MR, Amini-Bavil-Olyaee S, Kurbanov F, Van Ranst M, Tacke F. Molecular identification of hepatitis B virus genotypes/subgenotypes: revised classification hurdles and updated resolutions. World J Gastroenterol. 2014;20:7152–68.
36. Mojsiejczuk LN, Torres C, Sevic I, et al. Molecular epidemiology of hepatitis B virus in Misiones, Argentina. Infect Genet Evol. 2016;44:34–42.
37. Zehender G, Ebranati E, Gabanelli E, et al. Enigmatic origin of hepatitis B virus: an ancient travelling companion or a recent encounter? World J Gastroenterol. 2014;20:7622–34.
38. Sagnelli C, Ciccozzi M, Pisaturo M, et al. The impact of viral molecular diversity on the clinical presentation and outcome of acute hepatitis B in Italy. New Microbiol. 2015;38:137–47.
39. Shaw T, Bartholomeusz A, Locarnini S. HBV drug resistance: mechanisms, detection and interpretation. J Hepatol. 2006;44:593–606.
40. Zhang Q, Cao G. Genotypes, mutations, and viral load of hepatitis B virus and the risk of hepatocellular carcinoma: HBV properties and hepatocarcinogenesis. Hepat Mon. 2011;11:86–91.
41. Scotto G, Martinelli D, Di Tullio R, Fazio V. Epidemiological and clinical features of hepatitis B virus genotypes among immigrants in southern Italy. Hepat Res Treat. 2010;2010:878356.
42. Ozaras R, Balkan I, Yemisen M, Tabak F. Epidemiology of HBV subgenotypes D. Clin Res Hepatol Gastroenterol. 2015;39:28–37.