PATRÍCIA OLIVEIRA GIONDA

Análise de genomas completos do HBV genótipos F e H encontrados no Brasil e no México através do método de sequenciamento de nova geração

Dissertação apresentada à Faculdade de Medicina da Universidade de São Paulo para obtenção do título de Mestre em Ciências

Programa de Ciências em Gastroenterologia

Orientador: Dr. João Renato Rebello Pinho

São Paulo
2021

PATRÍCIA OLIVEIRA GIONDA

Análise de genomas completos do HBV genótipos F e H encontrados no Brasil e no México através do método de sequenciamento de nova geração

Dissertação apresentada à Faculdade de Medicina da Universidade de São Paulo para obtenção do título de Mestre em Ciências

Programa de Ciências em Gastroenterologia

Orientador: Dr. João Renato Rebello Pinho

São Paulo
2021

Dados Internacionais de Catalogação na Publicação (CIP)

Preparada pela Biblioteca da
Faculdade de Medicina da Universidade de São Paulo

©reprodução autorizada pelo autor

Gionda, Patrícia Oliveira

Análise de genomas completos do HBV genótipos F e H encontrados no Brasil e no México através do método de sequenciamento de nova geração / Patrícia Oliveira Gionda. -- São Paulo, 2021.

Dissertação (mestrado) -- Faculdade de Medicina da Universidade de São Paulo.

Programa de Ciências em Gastroenterologia.

Orientador: João Renato Rebello Pinho.

Descritores: 1.Vírus da hepatite B
2.Sequenciamento completo do genoma
3.Sequenciamento Illumina 4.Genótipos 5.Mutações
6.Brasil 7.México

USP/FM/DBD-310/21

Responsável: Erinalva da Conceição Batista, CRB-8 6755

Dedico este trabalho ao grande amor da minha vida, minha mãe Silvia.
Sem ela nada seria possível e eu seria ninguém.

AGRADECIMENTOS

Expresso aqui toda minha gratulação àqueles que, de um modo ou de outro, tornaram possível a realização da presente dissertação:

Agradeço primeiramente ao meu orientador Dr. João Renato por ter acreditado em mim e nas minhas capacidades e me permitir ser sua aluna, por todos os ensinamentos prestados, o grande compartilhamento de conhecimento e disponibilidade que sempre manteve.

Agradeço à Dra. Michele Gomes pela infinita paciência, pelas palavras certas que teve para que eu desenvolvesse este trabalho, por me guiar e fazer encontrar o meu caminho. Deixo uma palavra sincera de gratidão, porque sem essa paciência e sabedoria eu jamais seria esta pessoa tão realizada.

Agradeço à Dra. Fernanda Malta que se mostrou alguém tão fundamental para mim em diversos momentos, dentro do laboratório e fora dele, sempre disposta a me ajudar e me fazer crescer. Agradeço os muitos ensinamentos e as múltiplas sugestões que contribuíram para o desenvolvimento e aperfeiçoamento desta tese.

Agradeço à minha grande amiga pesquisadora e também mestranda Ana Paula Salles, a qual conheci graças a este trabalho, que se fez muito presente em todo o processo de elaboração desta tese e de grandes momentos de minha vida, minha cúmplice e parceira.

Agradeço ao colega e grande bioinformata desse projeto Pedro Sebe, por toda a paciência e conhecimento compartilhado ao longo do trajeto. Obrigada pela disposição em todos os momentos e por compartilhar toda sapiência da sua área tão complexa. Você é hoje, e com certeza no futuro será mais ainda, um grande profissional!

Agradeço ao Instituto de Medicina Tropical por ter proporcionado ambiente favorável à minhas pesquisas, bem como a estrutura necessária para que pudesse crescer e evoluir academicamente e pessoalmente, como também à

todas as pessoas que conheci na instituição nesse processo e que através do convívio se tornaram tão especiais para mim.

Agradeço à Faculdade de Medicina da USP pelo fornecimento de materiais que foram fundamentais para o desenvolvimento da pesquisa que possibilitou a realização deste trabalho.

Agradeço ao Hospital das Clínicas pelo provimento de todo conhecimento adquirido no local através dos profissionais da saúde, pelas palestras e aulas e pela elevada qualidade do ensino oferecido.

Agradeço a CAPES pela oportunidade e por me conceder os meios para chegar até aqui.

Essa dissertação está de acordo com:

Universidade de São Paulo. Faculdade de Medicina. Divisão de Biblioteca e Documentação. Guia de apresentação de dissertações, teses e monografias. Elaborado Anneliese Carneiro da Cunha, Maria Julia de A. L. Freddi, Maria Fazanelli Crestana, Marinalva de Souza Aragão, Suely Campos Cardoso. Valéria Vilhena. 3a ed. – São Paulo: Divisão de Biblioteca e Documentação - DBD/FMUSP, 2011.

Referências: adaptado de *Internacional Committee of Medical Journals Editors* (Vancouver).

Abreviaturas dos títulos dos periódicos de acordo com “*List of Journals Indexed in Index Medicus*” da *National Library of Medicine*.

LISTA DE ABREVIATURAS E SIGLAS

aa: Aminoácido

Anti-HBe: Anticorpo contra o antígeno e do vírus da hepatite B

Anti-HBs: Anticorpo contra o antígeno de superfície do vírus da hepatite B

cccDNA (circular covalent closed desoxiribonucleic acid): Ácido desoxirribonucleico circular covalentemente fechado

CHC: Carcinoma hepatocelular

Eco R1: Enzima de restrição EcoR1

DNA (deoxyribonucleic acid): Ácido desoxirribonucleico

HAV (Hepatitis A Virus): Vírus da hepatite A

HBcAg: Antígeno do core do vírus da hepatite B

HBeAg: Antígeno e do vírus da hepatite B

HBIG: Imunoglobulina anti-hepatite B

HBsAg: Antígeno de superfície do vírus da hepatite B

HBV (Hepatitis B Virus): Vírus da hepatite B

HBxAg: Antígeno x do vírus da hepatite B

HCV (Hepatitis C Virus): Vírus da hepatite C

HIV (Human Immunodeficiency Virus): Vírus da imunodeficiência humana

HLA (Human leukocyte antigen): Sistema antígeno leucocitário humano

MCMC (Markov Chain Monte Carlo): Monte Carlo via Cadeias de Markov

MHR (Major Hydrophilic region): Principal região hidrofílica

NGS (Next Generation Sequence): Sequenciamento de nova geração

OMS: Organização Mundial da Saúde

ORF (Open Reading Frame): Fase de leitura aberta

PCR (Polymerase Chain Reaction): Reação em cadeia da polimerase

PNI: Programa Nacional de Imunizações

RNA (Ribonucleic Acid): Ácido ribonucleico

RNAm: RNA mensageiro

RT (Reverse transcriptase): Transcriptase reversa

SBS (Sequencing by synthesis): Sequenciamento por síntese

SUS: Sistema Único de Saúde

Taq: *Thermus aquaticus*

LISTA DE FIGURAS

Figura 1 - Evolução da infecção pelo vírus da hepatite B.....	17
Figura 2 - Endemicidade global do HBV	19
Figura 3 - Representação esquemática do genoma do HBV.....	22
Figura 4 - Esquema das fases da replicação do HBV.	23
Figura 5 - Representação esquemática do princípio tecnológico da plataforma Illumina.....	29
Figura 6. Estratégia de amplificação do genoma completo do HBV em reação única utilizando os primers P1 e P2, protocolo A	33
Figura 7. Estratégia de amplificação do genoma completo do HBV utilizada no protocolo B	33
Figura 8 – Estratégia de amplificação do genoma completo do HBV utilizada no protocolo C	34
Figura 9 - Representação da qualidade de reads através pelo fastQC.	38
Figura 10 - Fluxograma de trabalho para análise de dados das sequências geradas	39
Figura 11 - Análise por eletroforese em gel de agarose de produtos de PCR gerados com os primers P1 e P2, protocolo A.	41
Figura 12 - Análise por eletroforese em gel de agarose de produtos de PCR amplificados	42
Figura 13 - Análise por eletroforese em gel de agarose de produtos de PCR do protocolo C	43
Figura 14: Representação esquemática do processo de trabalho feito para amplificação do genoma completo do HBV.....	44
Figura 15 - Representação do alinhamento do genoma da amostra CO20	48
Figura 16 - Árvore filogenética das sequências de genoma completo do HBV genótipo F caracterizadas usando sequenciamento de nova geração.....	50
Figura 17 - Árvore filogenética das sequências de genoma completo do HBV genótipo H caracterizadas usando sequenciamento de nova geração.	51

LISTA DE TABELAS

Tabela 1. Sequências dos primers utilizados nas reações de PCR.....	35
--	----

LISTA DE QUADROS

Quadro 1 - Resultados de qualidade e cobertura obtidos na análise de dados do sequenciamento dos genomas do HBV	46
Quadro 2 - Resultado da comparação dos contigs gerados com sequências referência do banco de dados NCBI (BLAST).....	47
Quadro 3 - Descrição das principais mutações relacionadas com impacto funcional nas amostras sequenciadas no presente estudo	55

SUMÁRIO

AGRADECIMENTOS.....	5
LISTA DE ABREVIATURAS E SIGLAS	8
LISTA DE FIGURAS	9
LISTA DE TABELAS.....	10
LISTA DE QUADROS.....	11
RESUMO.....	13
1. INTRODUÇÃO.....	15
1.1. História natural da hepatite B	15
1.2. Transmissão	17
1.3. Vacinação	19
1.4. Vírus da hepatite B.....	20
1.5. Diversidade genética viral	24
1.5.1. Genótipos e subgenótipos	24
1.6. Distribuição genotípica do HBV	25
1.7. Sequenciamento de nova geração	27
2. JUSTIFICATIVA.....	30
3. OBJETIVO.....	31
4. MATERIAIS & MÉTODOS	32
4.1. Amostras.....	32
4.2. Extração do DNA viral	32
4.3. Seleção de <i>primers</i> e estratégias para amplificação do genoma completo do HBV	32
4.4. Amplificação do genoma completo do HBV	35
4.5. Detecção do DNA amplificado.....	36
4.6. Preparo das bibliotecas para sequenciamento de nova geração e sequenciamento dos genomas do HBV amplificados.....	36
4.7. Análise das sequências de genoma completo do HBV geradas	37
4.8. Alinhamento de <i>contigs</i>	38
4.9. Análise Filogenética	39
4.10. Análise de Variantes.....	40
5. RESULTADOS	41
5.1. Amplificação de genomas completos do HBV	41
5.2. Sequenciamento	45
5.3. Análise de dados gerados	45
6. DISCUSSÃO	57
7. CONCLUSÕES.....	62
8. REFERÊNCIAS BIBLIOGRÁFICAS.....	63

RESUMO

Gionda PO. *Análise de genomas completos do HBV genótipos F e H encontrados no Brasil e no México através do método de sequenciamento de nova geração* [dissertação]. São Paulo: Faculdade de Medicina, Universidade de São Paulo; 2021.

INTRODUÇÃO: A hepatite B é uma das principais patologias hepáticas e uma das doenças infecciosas mais desafiadoras à saúde pública mundial, com altos índices de morbidade e mortalidade relacionadas à evolução da infecção pelo vírus da hepatite B (HBV). Dez diferentes genótipos (A-J) do HBV já foram caracterizados até o presente momento, sendo os genótipos F e H os que apresentam maior divergência e estão associados com casos de hepatite B nas Américas. **OBJETIVOS:** O objetivo deste trabalho foi sequenciar genomas completos de amostras dos genótipos F e H do HBV provenientes do Brasil e do México utilizando a metodologia de sequenciamento de nova geração; e analisar características genéticas relevantes para a compreensão da história natural das infecções associadas com esses genótipos. **MATERIAIS E MÉTODOS:** Foram incluídas 90 amostras de plasma com DNA do HBV detectável pertencentes aos genótipos F (n = 59) e H (n = 31), as quais foram submetidas à extração do DNA viral e posteriormente à amplificação do genoma completo do HBV. As amostras amplificadas foram submetidas à fragmentação do DNA, marcação e finalmente sequenciamento de última geração na plataforma MiSeq (Illumina). A análise dos dados foi desenvolvida por meio de ferramentas de bioinformática, nas quais foram avaliadas a qualidade das leituras obtidas e a cobertura de diferentes regiões do genoma. Os dados de sequenciamento foram analisados usando ferramentas de biotecnologia considerando qualidade e cobertura para alinhar os fragmentos e criar contigs mais longos. Sequências curadas e editadas foram então alinhadas com sequências de referência para determinar o subgenótipo viral e para detectar variantes de sequência em posições relevantes nos diferentes genes virais. Uma árvore filogenética foi construída usando métodos Bayesianos. Para a análise de variantes foi utilizada a ferramenta Freebayes. **RESULTADOS:** Foi possível amplificar o genoma completo do HBV de 31 amostras e dessas 18 foram sequenciadas com sucesso (HBV/F=16 e HBV/H=2).. A análise das sequências geradas mostrou que uma das amostras genótipo F apresentava co-infecção com os subgenótipos F1b e F3, enquanto as demais amostras eram todas subgenótipo F2a. Não foram observados eventos de recombinação nos genomas de HBV/F e H sequenciados. As principais variantes observadas nesse estudo foram aquelas com alterações nos genes Pré-S e S que levam à interrupção prematura da síntese ou não produção das proteínas de superfície (7/18); e mutações no códon de iniciação do gene pré-core que resultam na falta de síntese da proteína HBeAg (11/18). Essas variantes foram observadas em alguns casos compondo a população viral majoritária e em outros a minoritária. **CONCLUSÕES:** Neste trabalho, desenvolvemos uma metodologia de NGS aplicada para os genótipos F e H, característicos da América Latina em amostras obtidas no México e no Brasil. Estas amostras tiveram seus genomas completos sequenciados e encontrou-se 15 amostras do subgenótipo F2a, uma amostra apresentando co-infecção com os subgenótipos F1b e F3 e 2 amostras do genótipo H.

Descritores: Vírus da hepatite B; Sequenciamento do genoma completo; Sequenciamento Illumina; Genótipos; Mutações; Brasil; México;

ABSTRACT

Gionda, PO. *Analysis of complete genomes of HBV F and H genotypes found in Brazil and Mexico using the next-generation sequencing method* [dissertation] São Paulo: "Faculdade de Medicina, Universidade de São Paulo"; 2021.

INTRODUCTION: Hepatitis B is one of the main liver diseases and one of the most challenging infectious diseases to public health worldwide, with high rates of morbidity and mortality related to the evolution of hepatitis B virus (HBV) infection. Ten different genotypes (A-J) of HBV have been characterized to date, with the F and H genotypes showing the greatest divergence and being associated with cases of hepatitis B in the Americas. **OBJECTIVES:** The aim of this work was to sequence complete genomes of samples of HBV F and H genotypes from Brazil and Mexico using the next-generation sequencing methodology; and analyze genetic characteristics relevant to understanding the natural history of infections associated with these genotypes. **MATERIALS AND METHODS:** 90 plasma samples with detectable HBV DNA belonging to the F (n = 59) and H (n = 31) genotypes were included. The amplified samples were submitted to DNA fragmentation, labeling and finally sequencing of the latest generation in the MiSeq platform (Illumina). Data analysis was carried out using bioinformatics tools, in which the quality of the readings obtained and the coverage of different regions of the genome were evaluated. Sequencing data was analyzed using biotechnology tools considering quality and coverage to align fragments and create longer contigs. Cured and edited sequences were then aligned with reference sequences to determine viral subgenotype and to detect sequence variants at relevant positions in the different viral genes. A phylogenetic tree was constructed using Bayesian methods. For the analysis of variants, the Freebayes tool was used. **RESULTS:** It was possible to amplify the complete HBV genome of 31 samples and of these 18 were successfully sequenced (HBV/F=16 and HBV/H=2). The analysis of the generated sequences showed that one of the F genotype samples presented co- infection with F1b and F3 subgenotypes, while the other samples were all F2a subgenotype. No recombination events were observed in the sequenced HBV/F and H genomes. The main variants observed in this study were those with alterations in the Pre-S and S genes that lead to premature interruption of the synthesis or non-production of surface proteins (7/18); and mutations in the pre-core gene start codon that result in lack of HBeAg protein synthesis (11/18). These variants were observed in some cases making up the majority viral population and in others the minority. **CONCLUSIONS:** In this work, we developed an NGS methodology applied to F and H genotypes, characteristic of Latin America, in samples obtained from Mexico and Brazil. These samples had their complete genomes sequenced and 15 samples of the F2a subgenotype were found, one sample showing co-infection with the F1b and F3 subgenotypes and 2 samples of the H genotype.

Descriptors: Hepatitis B virus; Whole genome sequencing; Illumina sequencing; Genotypes; Mutations; Brazil; Mexico;

1. INTRODUÇÃO

1.1. História natural da hepatite B

O termo hepatites virais é utilizado para designar as doenças infecciosas que acometem principalmente o tecido hepático e que são ocasionadas por um grupo de vírus hepatotrópicos. A história das hepatites virais remonta desde períodos anteriores a era cristã, pois existem relatos de surtos de icterícia na Babilônia, há mais de 2500 anos (1). No século XVIII em 1725 foi introduzido pela primeira vez o termo hepatite por Bianchi J. B. (2). São conhecidas epidemias de icterícia durante a guerra da Sucessão Austríaca em 1743, durante a invasão de Napoleão no Egito em 1798, na guerra Franco-Prussiana de 1870 e Secessão Americana (1861-1865) (1) (2). Durante a guerra da Secessão Americana, mais de 40.000 soldados dos exércitos da União foram atingidos (2).

Em 1947, MacCallum (3) denominou os agentes causadores da icterícia epidêmica de “vírus da hepatite A” (HAV) e “vírus da hepatite B” (HBV). O primeiro era referente à icterícia infecciosa, com período de incubação de 18 a 37 dias, e o segundo à icterícia sérica, com período de incubação de 50 a 180 dias. Tais termos foram aprovados pelo Comitê de Hepatites Virais da Organização Mundial de Saúde (OMS) em 1977, e continuam sendo utilizados até os dias de hoje (4).

Em 1963, Baruch Blumberg e cols. (5) descobriram o antígeno Austrália no soro de um aborígene australiano, que foi inicialmente associado com uma série de diferentes patologias. Sua associação com a hepatite B foi feita em 1968, por Prince (6) e Okochi e Murakami (7), recebendo posteriormente o nome de antígeno de superfície do vírus da hepatite B (HBsAg). Poucos anos depois, em 1970, Dane e cols. (8) visualizaram pela primeira vez a partícula viral do vírus da hepatite B, que foi então caracterizada e denominada de partícula de Dane. Desde então, progressos consideráveis têm sido feitos a respeito da epidemiologia, virologia, história natural e tratamento desse vírus hepatotrópico (9).

A hepatite B pode evoluir forma aguda ou crônica. Pessoas infectadas podem desenvolver vários sintomas ou não apresentarem nenhum. Em qualquer uma das situações, o indivíduo pode se recuperar da infecção pelo HBV, através do clareamento viral espontâneo ao menos do vírus circulante e desenvolver imunidade contra o vírus, ou desenvolver hepatite crônica que geralmente permanece sem cura durante toda a vida. Entretanto, mesmo nos indivíduos em que a infecção se torna aparentemente resolvida, o genoma viral pode permanecer latente nas células hepáticas (10).

Em pessoas que desenvolvem a hepatite B aguda, o período de incubação dura geralmente de 1 a 4 meses, podendo variar de 6 semanas a 6 meses. A infecção aguda costuma ser benigna na maioria das vezes. Dois terços dos indivíduos infectados apresentam formas assintomáticas e evoluem para cura, um terço tem manifestações clínicas, e desses, apenas 10% tornam-se portadores crônicos do vírus (11). Os sintomas persistem por várias semanas e incluem fraqueza, náuseas, falta de apetite, vômito, dores abdominais, icterícia, urina escurecida, fezes esbranquiçadas, erupção cutânea e dores nas articulações. Aproximadamente 1 a 2% das pessoas que desenvolvem hepatite B aguda vem a falecer de hepatite fulminante (10).

Na hepatite B crônica, os indivíduos infectados geralmente não apresentam sintomas por décadas, e mais de 40% das pessoas infectadas cronicamente com o vírus podem desenvolver complicações, relacionadas com a hepatite crônica, em especial, a cirrose e o carcinoma hepatocelular (CHC) (12).

A idade em que a pessoa entra em contato com o HBV é o principal fator para se determinar a evolução da infecção. Aproximadamente 10% das crianças abaixo de 5 anos que entram em contato com o vírus são sintomáticas. Já 90% das crianças infectadas durante o primeiro ano de vida e 50% das crianças infectadas entre 1 e 4 anos de idade, desenvolvem hepatite crônica. Por outro lado, quando o indivíduo já é um adulto ao ser infectado, 50% desenvolvem a infecção de forma sintomática e menos de 5% desenvolvem hepatite B crônica (10).

Na figura 1, vemos um esquema da evolução da infecção pelo HBV em diferentes situações.

Fonte: Ferreira M.S. e Borges A.S., 2007 (11)
Figura 1 - Evolução da infecção pelo vírus da hepatite B

A evolução da forma crônica da hepatite B pode estar relacionada a variações genéticas do HBV e a características genéticas do indivíduo hospedeiro (13). As variações genéticas do genoma viral podem ser adquiridas durante a evolução da infecção, levando a seleção de variantes que conseguem escapar da ação do sistema imune (14), anticorpos estimulados pela vacinação, ou mesmo da ação de drogas antivirais (15). Com relação a características genéticas do indivíduo infectado, diferentes alelos dos complexos de histocompatibilidade podem estar relacionados ao diferenças no desenvolvimento da doença (13).

1.2. Transmissão

O HBV é considerado bastante infectivo e sabe-se que uma só partícula viral é capaz de infectar o ser humano. O vírus circula no sangue e se replica nos hepatócitos em torno de 10^{11} (100.000.000.000) cópias por dia, muito mais que os vírus da hepatite C (HCV) e da imunodeficiência humana (HIV) (próximo de 10^9 cópias por dia) (16). O HBV é estável e resistente ao meio ambiente,

sobrevivendo até uma semana fora do corpo humano (17). No plasma, a vida média do HBV varia de 1 a 3 dias, enquanto nos hepatócitos a vida média varia de 10 a 100 dias (17).

Há três principais formas de transmissão do HBV: de mãe para o filho no nascimento (transmissão vertical), por contato com uma pessoa infectada (transmissão parenteral) ou através de contato sexual (transmissão sexual) (18).

A transmissão vertical é a principal via de transmissão atualmente no mundo e ocorre principalmente em países subdesenvolvidos onde existe alta endemicidade de HBV (19). O mecanismo preciso da transmissão vertical é desconhecido, já que o vírus não é capaz de cruzar a barreira placentária. A hipótese mais aceita é que a infecção ocorra com a exposição da mucosa do recém-nascido a sangue e fluídos corpóreos infectados na hora do parto (20).

Do mesmo modo que a transmissão vertical, a transmissão sexual se dá com o contato da mucosa com sangue ou fluidos corporais infectados. A transmissão sexual é responsável pela alta frequência da infecção em diferentes populações de diversas origens geográficas e é reconhecida como a via de propagação viral em grupos de risco, como profissionais do sexo e homossexuais (21).

A transmissão parenteral inclui transfusão de sangue e seus derivados, hemodiálise, acupuntura, manicure, o uso de drogas injetáveis, tatuagens e perfuração com material infectado em profissionais da área da saúde. O HBV não consegue cruzar a pele ou a mucosa íntegra, sendo necessário que essas barreiras naturais estejam rompidas, mesmo que de forma mínima ou insignificante (22).

A prevalência do HBV e os padrões de transmissão variam muito nas diferentes partes do mundo, dependendo da taxa de infecção crônica, da proporção de pessoas com o vírus replicando ativamente e da via predominante de transmissão em cada região (23). As regiões foram divididas de acordo com a prevalência do HBsAg na população geral, de: baixa (0,1 a 2%), intermediária (2 a 7%), alta (>8%). Aproximadamente 45% da população global vivem em áreas com alta prevalência de hepatite B, 43% vive em áreas com prevalência intermediária, e 12% em áreas com baixa prevalência (24).

As áreas de elevada endemicidade incluem o Sudeste Asiático, Bacia do Pacífico (excluindo Japão, Austrália e Nova Zelândia), África Subsaariana, Bacia Amazônica, partes do Oriente Médio e alguns países da Europa Oriental. Entre aquelas consideradas de baixa endemicidade podemos citar a América do Norte, países do Norte e da parte ocidental da Europa, Austrália e partes da América do Sul. As demais regiões do mundo enquadram-se no perfil de áreas de endemicidade intermediária (25).

Fonte: CDC, 2008 (26)

Figura 2 - Endemicidade global do HBV

Nas áreas com prevalência intermediária ou alta, a maioria dos indivíduos torna-se infectado no momento do nascimento ou durante os primeiros anos de vida pelo contato com portadores do vírus, o que ocorre com mais frequência entre membros da mesma família. Já nas áreas de baixa prevalência, a infecção é adquirida principalmente pela via sexual (23).

1.3. Vacinação

A vacina contra a hepatite B é a medida preventiva mais efetiva. As primeiras vacinas foram licenciadas em 1982 e eram derivadas de plasma de

pacientes com infecção crônica, com HBsAg inativados por métodos físico-químicos e em 1986 passou a ser obtida HBsAg recombinante obtido em células de leveduras. Na década de 90, a Organização Mundial de Saúde (OMS) recomendou que a vacina fosse incorporada nos programas de vacinação nacionais (27). Também a partir da década de 90, a vacina passou a ser oferecida pelo Sistema Único de Saúde (SUS) (28). Desde 1998, o Programa Nacional de Imunizações (PNI), do Ministério da Saúde, recomenda que crianças a partir do seu nascimento, sejam vacinadas contra a hepatite B. A primeira dose deve ser aplicada nas primeiras 12-24h de vida, o que resulta numa elevada eficácia na prevenção da transmissão vertical. O esquema da vacinação é feito da seguinte forma: 0, 1 e 6 meses, e deve ser aplicado independentemente da idade do indivíduo (29).

Em 2001, a faixa etária de vacinação contra hepatite B foi ampliada para 19 anos de idade, em 2011 para 24 anos, e em 2012 para 29 anos. Além disso, para reduzir a transmissão vertical do vírus, foi intensificada a oferta de triagem sorológica a todas as gestantes que fazem o pré-natal no SUS e todos os recém-nascidos de mães portadoras da doença recebem profilaxia – vacina e imunoglobulinas (28).

Diversos estudos mostram que as vacinas contra hepatite B têm boa imunogenicidade e são eficazes, com proteção em mais de 90% dos adultos jovens saudáveis e em mais de 95% em lactentes, crianças e adolescentes. A eficácia diminui gradativamente após os 40 anos de idade. A obesidade, o estresse, o tabagismo e o etilismo também são fatores associados a uma menor eficácia vacinal (29).

1.4. Vírus da hepatite B

O vírus da hepatite B é um vírus envelopado pertencente ao gênero *Orthohepadnavirus*, da família *Hepadnaviridae*, que possui tropismo por células hepáticas e contém um genoma circular de aproximadamente 3215 pb (pares de bases) (30). Todos os vírus da família *Hepadnaviridae* são capazes de causar infecções agudas com viremia e antigenemia (31).

As células infectadas pelo HBV produzem diferentes tipos de partículas relacionadas ao vírus, as quais podem ser identificadas por microscopia eletrônica e são classificadas em: (i) partículas completas infecciosas (partículas de Dane) com 42 a 47 nm de diâmetro sendo constituídas por um envelope externo e uma estrutura interna que compreende o antígeno central da hepatite B (HBcAg), o DNA viral e a proteína DNA polimerase (32); (ii) partículas incompletas esféricas de 17 a 25 nm; e (iii) partículas incompletas cilíndricas (tubulares) com cerca de 20 a 22 nm. As partículas incompletas são constituídas apenas pelo envelope viral e não contém DNA, portanto não são infecciosas. Essas partículas são secretadas em grande quantidade, excedendo a produção das partículas completas em 1.000 a 10.000 vezes (33).

O genoma viral apresenta uma estrutura peculiar, sendo constituído por uma molécula de DNA circular, com uma fita parcialmente dupla apresentando uma região de fita simples com tamanho variável. A fita longa, de polaridade negativa, possui um tamanho fixo, com cerca de 3200 nucleotídeos. Já a fita curta, de polaridade positiva, pode variar entre 1700 e 2800 nucleotídeos nas diferentes moléculas (34).

Na fita maior estão presentes as quatro regiões abertas de leitura (ORFs, do inglês *Open Reading Frame*): S, C, P e X (35). A região S codifica proteínas de superfície do envelope do vírus (HBsAg) é constituída de 226 aminoácidos (36). A cadeia de leitura pre-S/S pode iniciar-se em 3 diferentes códon de iniciação, codificando as glicoproteínas de superfície grande (L), média (M) e pequena (S) que contém até 3 regiões diferentes: pre-S1 (envolvida no reconhecimento do HBV pelos receptores do hepatócito), pre-S2 e S (37). Os aminoácidos 21 a 47 da região pre-S1 codificam o sítio de reconhecimento dos receptores para os hepatócitos, portanto fundamentais para entrada do vírus nas células (36). A região pre-S é a mais variável do genoma sendo que somente 49,6% dos nucleotídeos são conservados em um estudo que envolveu diversas sequências de diferentes genótipos virais (38). Em contraste, a parte do gene S que codifica a proteína de superfície pequena (nucleotídeos 155-832) parece ser a mais conservada do genoma viral, com 81,9% dos nucleotídeos conservados (38). Essa porção do genoma pode ser utilizada para genotipagem com um alto

grau de confiabilidade, pois tem regiões conservadas em meio a regiões pouco conservadas, o que a faz também útil para análises filogenéticas (38).

Fonte: Kidd-Ljunggren K. e cols., 2002 (39)

Figura 3 - Representação esquemática do genoma do HBV

A fase de leitura aberta C contém a região C e pre-C. Essas regiões são transcritas em dois RNA mensageiros (RNAm): pré-genômico e pré-core. O RNAm pré-genômico é traduzido na proteína do nucleocapsídeo (HBcAg), que está presente no core de vírions em células hepáticas infectadas, contudo não é detectado no soro. Além disso, o HBcAg está integrado ao genoma viral e à DNA polimerase, sendo essencial para a função e maturação do vírion (40). O RNAm pré-core é traduzido na proteína pré-core, que depois de processada é secretada como o antígeno e (HBeAg), presente no soro de pacientes com replicação do HBV. Em algumas situações podem ocorrer mutações na região pré-core ou na região promotora e o paciente pode apresentar replicação do HBV, mesmo tendo sorologia negativa para o HBeAg.

O gene P é o mais longo do HBV, estendendo-se por aproximadamente 80% do genoma viral e se sobrepondo a três outros genes (S, C e X) (41). A polimerase do HBV é uma proteína multifuncional que contém 4 domínios; uma região iniciadora, uma região espaçadora de função desconhecida, uma região catalítica que funciona como DNA polimerase RNA dependente e uma região

carboxi-terminal que possui atividade ribonuclease H (42). Assim, a polimerase viral funciona como uma transcriptase reversa (para a síntese da cadeia negativa do DNA a partir RNA genômico), e como uma DNA polimerase endógena (37).

O gene X codifica o HBxAg, que parece estar envolvido nos processos de carcinogênese, através de transativação de promotores celulares e virais (32) (43).

O ciclo de vida do HBV é caracterizado pela síntese do DNA de fita dupla parcial, através da transcrição reversa do RNA intermediário (43). A replicação começa com a ligação do vírion com o hepatócito. Dentro do núcleo do hepatócito, a síntese da fita positiva de DNA do HBV é completada e o genoma viral é convertido para a forma de DNA circular (44). O DNA circular serve como modelo para a transcrição de diversas espécies de RNAs genômicos e subgenômicos, sendo o componente estável do ciclo de replicação que é relativamente resistente à ação de antivirais e ao clareamento (45). Estudos revelaram que a meia-vida plasmática do HBV é de 24 horas, com um *turnover* diário de 50% (46).

Fonte: Okamoto H. e cols., 1988 (47), modificada

Figura 4 - Esquema das fases da replicação do HBV. Vírions infectam a célula, liberando o DNA no citoplasma. O DNA é transportado para o núcleo e reparado para formar o cccDNA [1]. A transcrição de cccDNA pela RNA polimerase [2] produz, entre outros transcritos, o RNA pré-genômico. O RNA pré-genômico é encapsidado, juntamente com a proteína P, e transcrito reversamente dentro do nucleocapsídeo [3]. Novos ciclos levam à amplificação do cccDNA intracelular e a liberação de novos vírions

1.5. Diversidade genética viral

Em 1971, Le Bouvier (48) fez o primeiro relato de diversidade de subtipos sorológicos no HBV. A divergência das sequências do genoma completo dentro de um mesmo subtipo é de aproximadamente 8% semelhante à encontrada em diferentes subtipos. Assim, a subgenotipagem não reflete uma verdadeira variação genotípica nas sequências codificadoras da proteína S. Por essa razão, o método de classificação de genótipos do HBV baseados na sequência de nucleotídeos do HBsAg passou a ser adotado (47). A classificação genética das cepas de HBV usando a sequência de nucleotídeos do genoma completo foi realizada primeiramente por Okamoto e cols. (47) em 1988. A análise permitiu a identificação de quatro grupos genômicos: A, B, C e D (49).

Posteriormente Norder e cols. (49), mediante o sequenciamento apenas do gene S, identificaram dois novos genótipos: E e F, sendo o genótipo F o mais divergente de todos e mais frequente na América do Sul (35).

Stuyver e cols. (50) descreveram o genótipo G em soro de pacientes franceses e americanos com infecção crônica pelo HBV e Arauz-Ruiz e cols. (51) relataram um oitavo genótipo do HBV, o genótipo H, isolado em cepas virais oriundas da Nicarágua, México e Califórnia.

A genotipagem baseada no sequenciamento apenas do gene S tem se mostrado equivalente a realizadas com o genoma completo (52). Os estudos baseados no sequenciamento parcial do genoma do HBV também têm sido usados não só para esclarecer questões de epidemiologia ou mecanismos básicos da evolução do vírus, mas também no diagnóstico molecular correlacionando a evolução molecular do vírus com a evolução clínica da doença (53) (54) (50).

1.5.1. Genótipos e subgenótipos

Como visto acima, foram identificados inicialmente oito genótipos (A-H) do HBV (60). Os genótipos I e J foram descritos depois e sua denominação como novos genótipos ainda é fruto de continuas discussões entre os pesquisadores da área. Tal indefinição se dá pelo fato desses supostos novos genótipos serem resultantes de recombinações entre genomas de diferentes genótipos do HBV

isolados em humanos (genótipo I) (61) ou entre genótipos isolados em humanos e em primatas não humanos (genótipo J) (62).

Alguns genótipos são ainda classificados como subgenótipos (63). São caracterizados por > 8% de nucleotídeos diferentes para genótipo, e 4%-8% de nucleotídeos diferentes para subgenótipos (64). Mais de 30 subgenótipos foram determinados até agora (65), identificados entre os genótipos A-D e F (66).

1.6. Distribuição genotípica do HBV

Mais de 2 bilhões de pessoas já tiveram contato com o HBV em algum momento de suas vidas, e dessas, aproximadamente 350 milhões desenvolveram hepatite B crônica e são portadores do vírus. A incidência da infecção por HBV e sua forma de transmissão varia largamente dependendo da região geográfica, grupos populacionais (67) e via de exposição (68). Por exemplo, os genótipos B e C são prevalentes em áreas altamente endêmicas, onde a transmissão perinatal ou vertical desempenha um papel importante na disseminação viral, enquanto que os genótipos remanescentes são frequentemente encontrados em áreas onde a transmissão horizontal é o principal modo de transmissão (69).

O genótipo A predomina em algumas regiões da África, regiões norte e central da Europa e na América do Norte. Os genótipos B e C são comuns no sudeste asiático, China e Japão. O genótipo D tem uma ampla distribuição geográfica, mas é frequente em países localizados na bacia do Mediterrâneo e Oriente Médio. O genótipo E é considerado nativo da África e predomina na região ocidental do continente (70) (71). O genótipo F é característico da América do Sul, sendo o único genótipo encontrado entre populações indígenas isoladas (72) (73) e amplamente distribuído nos países com forte contribuição desse grupo étnico como Venezuela (72), Colômbia (74) e Chile (75). O genótipo G foi relatado inicialmente na Europa (Alemanha e França) e posteriormente nas Américas (Canadá, EUA, México e Colômbia) (76). O genótipo H é encontrado na América Central, sobretudo no México (77). O genótipo I, uma nova recombinação intergenotípica entre os genótipos A, C e G, foi isolado no Vietnã e em Laos (78). O genótipo J, o mais recente, foi identificado nas Ilhas Ryukyu

no Japão e parece ser também um recombinante entre vírus humanos e de primatas não humanos (79).

A variação subgenotípica também encontra diversidade de distribuição geográfica mundial. O genótipo A é dividido em diferentes subgenótipos, de A1 até A7. O subgenótipo A1 é encontrado principalmente em algumas regiões dos continentes africano (África Oriental e do Sul), asiático (Índia, Bangladesh) e americano (Brasil, Haiti, Martinica) enquanto A2 predomina na Europa (80). Posteriormente, mais cinco subgenótipos foram propostos para a classificação de cepas isoladas em diferentes regiões da África (A3, A4, A5 e A7) e na Bélgica, entre pacientes de origem africana (A6) (62).

O genótipo B foi classificado em 9 subgenótipos diferentes, que se distribuem por diferentes populações. Os subgenótipos são classificados em: B1 no Japão; B2 no Japão e China; B3 na Indonésia; B4 no Vietnã; B5 nas Filipinas; B6 em populações nativas do Ártico incluindo Alasca, norte do Canadá e Groenlândia; e B7-B9 na Indonésia.

O genótipo C é classificado em 16 subgenótipos diferentes com distribuição nos países asiáticos: C1 no Sudeste da Ásia e Sul da China; C2 no Sudeste asiático, China e Japão; C3 na Oceania e Indonésia; C4 na Austrália; C5 nas Filipinas, Vietnã e Indonésia; C6 na Indonésia e Filipinas; C7 nas Filipinas; e C8-C16 na Indonésia (62).

O genótipo D é dividido em 9 subgenótipos diferentes e estes têm distribuição mundial não seletiva e dividem-se em: D1 na Bacia do Mediterrâneo, Oriente médio, Índia e China; D2 no Japão e Norte da Europa; D3 na Índia, África do Sul e sul da Europa; D4 na Oceania e Haiti; D5 na Índia e Japão; D6 na Indonésia; D7 na Tunísia e D8 na Nigéria (62) (81).

As cepas do genótipo F isoladas em diferentes regiões geográficas são classificadas em 4 subgenótipos: F1, cujo clado é subdividido em F1a e encontrado na América Central e F1b na Argentina; F2, que também tem uma subdivisão em F2a (encontrado na Venezuela e Argentina) e F2b (na Venezuela); F3 na Venezuela e Colômbia; e F4 na Argentina (82).

No Brasil, os genótipos A, D e F circulam em todas as regiões, sendo que o genótipo A é o mais prevalente em todas elas com exceção da região Sul, onde o genótipo D predomina (83) (84). Estudos mais recentes apontam que o

subgenótipo A1 de origem africana é o mais frequente em todas as regiões do país (73). Enquanto que os subgenótipos D2 e D3 são os mais frequentes no Sul e Sudeste do país, o subgenótipo D4 foi identificado no Maranhão (73) (83) (85). O genótipo F é encontrado com maior frequência no norte do país, especialmente entre populações indígenas, sendo o subgenótipo F2a o mais frequente (73).

Os genótipos B e C também circulam no Brasil, porém praticamente restritos a indivíduos com origem oriental (84) (86). O genótipo E é muito raro no país, sendo só encontrado entre imigrantes africanos (87) ou naqueles que passam algum período neste continente (88). O genótipo G também já foi descrito em alguns casos no Rio de Janeiro (89), São Paulo (88) (90) (91) e Acre (92). O genótipo H até o momento só foi encontrado em um caso no sul do Brasil (83).

Alguns estudos evidenciaram a relação entre o genótipo e a gravidade da doença. Na Ásia, os genótipos B e C são mais frequentes, e o genótipo C estaria associado a complicações mais graves que o genótipo B (93). Os genótipos C e D apresentam pior resposta ao interferon e associam-se com doença hepática mais avançada que os genótipos A e B (94). Os portadores do genótipo B também apresentaram melhor resposta ao tratamento combinado com interferon e Lamivudina em comparação ao genótipo C (95), bem como resposta viral sustentada após tratamento. O genótipo F, altamente predominante em populações indígenas, foi implicado em uma progressão frequente e rápida para CHC (96), mesmo sendo menos propenso a produzir infecção crônica do que, por exemplo, o genótipo D (97). No entanto o genótipo H, a outra variante americana do HBV, raramente é associado ao CHC. Este genótipo é predominante na população mexicana, onde está associado a baixa carga viral e a hepatite oculta (98).

1.7. Sequenciamento de nova geração

São processos de sequenciamento de DNA que utilizam metodologias com o objetivo de acelerar e baixar o custo do processo de sequenciamento.

Apesar de se diferenciarem consideravelmente entre si, todos os sequenciadores de nova geração se baseiam no processamento paralelo massivo de fragmentos de DNA (99). Cada tecnologia de sequenciamento possui uma estratégia diferente, mas em geral podemos identificar etapas comuns entre todas elas: preparo da amostra, amplificação da biblioteca e sequenciamento.

Baseia-se no conceito de “sequenciamento por síntese” (SBS, do inglês *sequencing by synthesis*), semelhante ao sequenciamento Sanger. O protocolo de sequenciamento requer que as sequências a serem determinadas sejam convertidas numa biblioteca de sequenciamento especial, permitindo a amplificação e imobilização das sequências para serem submetidas ao sequenciamento. Para este propósito, dois adaptadores diferentes são adicionados às terminações 5' e 3' de todas as moléculas (100).

Primeiramente, realiza-se a fragmentação de forma aleatória e, esses fragmentos são, posteriormente, ligados a adaptadores localizados nas duas extremidades. A adesão das moléculas de DNA fita simples ocorre por afinidade ao suporte sólido, local onde se encontram os oligonucleotídeos complementares a esses adaptadores. Na fase de anelamento, o adaptador da extremidade da molécula que se encontra livre, se une ao seu oligonucleotídeo complementar no suporte e forma uma estrutura em ponte, que logo dá início a PCR, fazendo uso da extremidade 3' livre do oligonucleotídeo como primer. Já durante a etapa de desnaturação, essa ponte é desfeita através do uso de altas temperaturas. Posteriormente, a etapa de anelamento é repetida, dando origem a novas estruturas em ponte, iniciando um novo ciclo de amplificação. É importante destacar que nesta plataforma podem ser gerados fragmentos tanto *paired-end*, onde o sequenciamento é feito para fragmentos de DNA com o tamanho entre 200-500bp, quanto para *single-read*, onde apenas uma extremidade do fragmento de 50-70bp é sequenciada, para a preparação de bibliotecas de DNA (101).

No momento em que ocorrer uma quantidade suficiente desses ciclos, são obtidos *clusters* de moléculas idênticas ligadas ao suporte e, através da incorporação de nucleotídeos terminadores marcados e da excitação a laser, é gerado sinal, que será captado por um aparelho que realizará a leitura e a

interpretação dos possíveis nucleotídeos componentes da cadeia. O procedimento de leitura é feito de forma sequencial, o que permite a montagem da sequência completa de cada um dos clusters gerados (101).

Fonte: Illumina INC. 2010 (102)

Figura 5 - Representação esquemática do princípio tecnológico da plataforma Illumina. Com a incorporação de nucleotídeos marcados, é gerado sinal, que é captado pelo dispositivo de leitura e interpretado como um dos quatro nucleotídeos componentes da cadeia. A leitura é feita de forma sequencial, o que permite a montagem da sequência completa de cada cluster

2. JUSTIFICATIVA

Apesar de terem se passado mais de cinco décadas da identificação do agente etiológico da hepatite B e de todos os estudos subsequentes contribuírem para o conhecimento desta infecção, a hepatite B persiste como um problema de saúde pública mundial, sendo ainda associada a muitos casos de cirrose, hepatocarcinoma e a alta mortalidade (103) (104). Conhecer a sequência de DNA dos vírus é um grande interesse da clínica humana, pois tem impacto no diagnóstico, no entendimento da patogênese da doença, no desenvolvimento de medidas preventivas, e na compreensão de processos de dispersão dos vírus nas populações humanas.

NGS é uma metodologia que pode ser caracterizada como um sequenciamento automatizado, paralelo e de alto rendimento, permitindo a obtenção de sequências longas. Sequências longas são muito importantes para que possamos fazer a caracterização de várias propriedades virais em um só ensaio. A possibilidade de termos uma metodologia para obter a sequência do genoma completo do HBV permite que diferentes propriedades virais sejam estudadas. É possível um teste que possibilite caracterizar ao mesmo tempo o genótipo viral, a presença de possíveis recombinantes, a presença de mutações relevantes nos diferentes genes virais que vão impactar na imunogenicidade e resistência a vacinas (análise dos genes do envelope viral); capacidades de replicação, “fitness” e resistência aos medicamentos antivirais (análise do gene da DNA polimerase viral); padrões imunológicos diferentes durante a evolução (análise das regiões basal do core e pré-core) e hepatocarcinogenicidade (análise na região do gene X).

Neste trabalho, desenvolvemos a técnica de sequenciamento completo do genoma do HBV focando inicialmente na caracterização dos genomas nos genótipos característicos de populações ameríndias.

3. OBJETIVO

O objetivo deste trabalho foi sequenciar o genoma completo de amostras de genótipo F e H do HBV provenientes do Brasil e do México utilizando-se a metodologia de sequenciamento de nova geração, e estudar características relevantes para a compreensão das doenças associadas com esse vírus a partir dos dados obtidos pelas análises das sequências geradas:

- 1) Caracterização dos subgenótipos virais;
- 2) Pesquisa de presença de possíveis recombinantes e de infecções mistas com diferentes genótipos/subgenótipos;
- 3) Detecção de presença de mutações relevantes nos diferentes genes virais que vão implicar em modificações que tem impacto na evolução da doença: a) alteração no padrão de imunogenicidade e escape vacinal (pela análise dos genes do envelope viral); b) capacidades de replicação, “fitness” e resistência aos medicamentos antivirais (análise do gene da DNA polimerase viral); c) padrões imunológicos diferentes durante a evolução (análise das regiões basal do core e pré-core); d) hepatocarcinogenicidade (análise na região do gene X).

4. MATERIAIS & MÉTODOS

4.1. Amostras

Para a realização deste estudo foram incluídas 90 amostras de plasma obtidas de pacientes infectados pelo HBV genótipo F (n=59) ou genótipo H (n=31). As amostras de genótipo F foram obtidas de pacientes provenientes de diferentes estados brasileiros: Maranhão (n=6), Minas Gerais (n=4), São Paulo (n=36), Pará (n=12) e Rio Grande do Sul (n=1). As amostras pertencentes ao genótipo H foram obtidas de pacientes procedentes de Guadalajara, México e foram cedidas pelo Dr. Arturo Panduro Cerda, do Serviço de Biologia Molecular do Antigo Hospital Civil de Guadalajara – México.

Todas as 59 amostras do genótipo F eram HBsAg positivas. Sua carga viral variou entre 2.65 log e 8.95 log, sendo a grande maioria das amostras quantificada com carga viral abaixo de log 5. Das 31 amostras do genótipo H utilizadas no trabalho, 15 eram HBsAg positivo e das demais essa informação não estava disponível, porém todas com HBV DNA detectável.

4.2. Extração do DNA viral

O DNA viral de cada amostra foi extraído a partir de 200 µl de plasma usando o QiaAmp DNA mini Kit (Qiagen, Hilden, Alemanha), de acordo com as instruções do fabricante.

4.3. Seleção de *primers* e estratégias para amplificação do genoma completo do HBV

O genoma completo do vírus da hepatite B foi amplificado por meio da reação em cadeia da polimerase (PCR) utilizando-se três diferentes protocolos de amplificação (A, B e C), nos quais foram empregadas diferentes combinações de primers que anelam em posições estratégicas do genoma viral de forma a amplificar fragmentos que englobem toda a sequência nucleotídica do mesmo.

No protocolo A foram utilizados os primers P1 e P2 descritos por Günther e cols. (105), os quais anelam em posições do genoma que permite a amplificação dos 3215 pb do genoma do HBV em uma única reação (Figura 6).

Fonte: Gunther S. e cols., 1995 (105)

Figura 6. Estratégia de amplificação do genoma completo do HBV em reação única utilizando os primers P1 e P2, protocolo A

O protocolo B consistiu na realização de três reações de PCR usando diferentes combinações de primers para amplificação de três fragmentos sobrepostos: P1.2 e P194 foram utilizados para gerar o primeiro fragmento de 1611 pb; PS3076F e P1196R para gerar o segundo fragmento de 1335 pb; e 5'LAM5 e P2.2 para gerar o terceiro fragmento de 1210 pb (Figura 7) (106).

Fonte: Gomes M.M.S., 2005 (106), modificado.

Figura 7. Estratégia de amplificação do genoma completo do HBV utilizada no protocolo B

No protocolo C o genoma completo do HBV foi amplificado utilizando-se primers que geram 2 fragmentos sobrepostos e foi empregada a metodologia de *nested* PCR, que consiste em melhorar a sensibilidade da reação para amplificação do DNA de interesse, possibilitando dessa forma maior sucesso na amplificação de amostras com baixos níveis de carga viral, como descrito abaixo.

A reação *nested* PCR é uma modificação da PCR convencional projetada para melhorar a sensibilidade e a especificidade da reação. Envolve o uso de dois conjuntos de primers e duas reações de PCR sucessivas. Os amplicons resultantes da primeira reação de PCR são usados como modelo para o segundo passo da amplificação. A sensibilidade e especificidade da amplificação do DNA podem ser significativamente melhoradas com esta técnica.

Os primers utilizados nesta estratégia foram escolhidos a partir do estudo realizado por Wang e cols. (107) cuja finalidade era conseguir amplificar e sequenciar com alta acurácia e sensibilidade amostras de pacientes infectados pelo HBV com baixa carga viral (Figura 8). Para amplificação do primeiro fragmento foram utilizados os primers P1 e AR1 para realização da primeira PCR, e P1 e AR2 na reação *nested* PCR, gerando um fragmento de 2092 pb. Para amplificação do segundo fragmento, os primers P2 e AF1 foram utilizados na primeira reação, e P2 e AF2 na reação *nested* PCR, resultando em um fragmento de 1320 pb.

Todos os primers escolhidos nos protocolos realizados e sua respectiva sequência e posição no genoma estão listados na Tabela 1.

Fonte: Wang X. e cols., 2017 (107), modificado

Figura 8 – Estratégia de amplificação do genoma completo do HBV utilizada no protocolo C. As setas azuis indicam a primeira rodada de amplificação e as setas de cor rosa indicam a segunda rodada

Tabela 1. Sequências dos primers utilizados nas reações de PCR para amplificação do genoma completo do HBV

Primer	Sequência 5'-3'	Posição	Referência
P1	CCGGAAAGCTTGAGCTCTTC TTTTTCACCTCTGCCTAATCA	1821-1841	Günther <i>et al.</i> , 1995
P2	CCGGAAAGCTTGAGCTCTTC AAAAAGTTGCATGGTGCTGG	1825-1806	Günther <i>et al.</i> , 1995
P1.2	TTTTTCACCTCTGCCTAATCA	1821-1841	Günther <i>et al.</i> , 1995 (com modificações)
P2.2	AAAAAGTTGCATGGTGCTGG	1825-1806	Günther <i>et al.</i> , 1995 (com modificações)
P194	AAACCCCGCCTGTAACACGA	211-192	Gomes, 2015
PS3076F	TGGGGTGGAGCCCTCAG	3079-3095	Stuyver <i>et al.</i> , 2000
P1193R	GCGTCAGCAAACACTTGGA	1193-1174	Gomes, 2015
5'LAM5	TGCRYTGTATTCCCATCCCATC	593-615	Da Silva <i>et al.</i> , 2000
AR1	ACAGTGGGGGAAAGC	759-745	Wang <i>et al.</i> , 2017
AR2	AGAAACGGRCTGAGGC	702-687	Wang <i>et al.</i> , 2017
AF1	GTCTGCGGCGTTTTATC	419-435	Wang <i>et al.</i> , 2017
AF2	TGCCCGTTTGTCCTCTA	503-519	Wang <i>et al.</i> , 2017

4.4. Amplificação do genoma completo do HBV

Para as reações de amplificação, foi utilizada em todos os protocolos a enzima Q5® *Hot Start High-Fidelity DNA Polymerase* (New England Biolabs, Massachusetts, EUA). Essa enzima foi escolhida para o presente trabalho por apresentar uma baixa taxa de inserção de erros durante o processo de amplificação (aproximadamente 280 vezes menor que a *Taq* DNA polimerase). Foi utilizado 0,5 µl da enzima em um mix contendo: 10 µl do tampão de reação que acompanha a enzima (5X Q5 *Reaction Buffer*); 26,5 µl de H₂O; 1 µl de dNTPs a 10 mM e 1 µl dos primers específicos a cada protocolo na concentração de 30 pmol/ µl. Posteriormente, as amostras foram submetidas a repetidos ciclos

de variação de temperatura no termociclador Mastercycler proS (Eppendorf AG, Hamburg, Germany) para amplificação de cada alvo.

Para as reações utilizando o protocolo A, foi realizada uma etapa de desnaturação inicial a 98 °C por 30 segundos seguido por 40 ciclos a 98 °C por 10 segundos, 64 °C por 30 segundos, 72 °C por 2 minutos. Ao término dos ciclos de amplificação foi acrescentado um passo de extensão a 72 °C por 2 minutos para finalização de alguma polimerização incompleta.

As reações nas quais foi empregada a combinação de primers do protocolo B foram submetidas aos seguintes ciclos de variação de temperatura: 98 °C por 5 min seguido por 34 ciclos de 98 °C por 10 segundos, 61 °C por 30 segundos, 72 °C por 50 segundos; e um ciclo final a 72 °C por 5 minutos.

As condições de ciclagem aplicadas às reações montadas com os *primers* do protocolo C foram as seguintes: 98 °C por 2 minutos seguido por 30 ciclos de 98 °C por 10 segundos, 58 °C (para a primeira rodada de amplificação) e 53,7 °C (para a reação de *nested* PCR) por 30 segundos; 72 °C por 1 minuto e 30 segundos; e um ciclo final a 72 °C por 5 minutos.

4.5. Detecção do DNA amplificado

O produto amplificado foi identificado por eletroforese em gel de agarose. Para isso, foram misturados 4 µl do produto final da PCR com 1 µl do tampão 10X Bluejuice™ Gel Loading Buffer (Invitrogen™ Life Technologies, Carlsbad, CA, EUA) e a mistura foi aplicada em gel de agarose 1% preparado com tampão TAE 1X e 7 µl de SYBR Safe. Foi também aplicado no gel um marcador de tamanho molecular de 10.000 bp (High DNA Mass Ladder) (Invitrogen™ Life Technologies, Carlsbad, CA, EUA), para avaliação do tamanho do fragmento obtido. O gel foi submetido à eletroforese de 110 V por 40 minutos, e ao final o produto amplificado foi visualizado em um transiluminador de luz Ultra Violeta (U.V.).

4.6. Preparo das bibliotecas para sequenciamento de nova geração e sequenciamento dos genomas do HBV amplificados

Inicialmente foi preparada a biblioteca de DNA a partir dos *amplicons* gerados nas reações de PCR, para tal foi utilizado o *kit Nextera XT DNA Sample*

Preparation, seguindo os procedimentos estabelecidos pelo fabricante. Após a quantificação dos produtos de PCR utilizando o método fluorimétrico Qubit (Qubit 4 Fluorometer; Thermo Fisher Scientific, Waltham, Massachusetts, EUA) a concentração de todas as amostras de PCR foi ajustada para 0,8 ng/ μ l, e em seguida foi realizada a etapa de fragmentação e tagmentação do DNA.

As amostras fragmentadas foram submetidas a uma reação de PCR para a inserção de adaptadores, cujo objetivo é a diferenciação de cada amostra. Ao final da etapa de amplificação foi realizada a purificação do DNA utilizando *beads* magnéticas AMPure XP (Beckman Coulter; Life Sciences Division Headquarters; Indianapolis, USA). Essas *beads* são capazes de selecionar por tamanho o material a ser purificado, então fragmentos que excedem 1000 bp são descartados, e fragmentos muito curtos também são descartados. Posteriormente, com as amostras devidamente purificadas e quantificadas, foi realizada a etapa de mistura das mesmas compondo uma única solução a 2 nM de concentração final. A etapa de denaturação foi realizada utilizando uma solução de NaOH a 0,2 N (v/v). As bibliotecas prontas foram pipetadas em uma lâmina específica para o sequenciador (*flow cell*) e então o processo de sequenciamento foi iniciado.

A plataforma MiSeq (Illumina Inc., San Diego, CA, U.S.A.) foi utilizada para a realização da etapa de sequenciamento do genoma completo do HBV. Essa plataforma tem capacidade para gerar 15 milhões de *reads*.

4.7. Análise das sequências de genoma completo do HBV geradas

O arquivo FASTQ usado para montagem de um genoma ou região de um genoma contém muitas sequências curtas chamadas *reads*. Os *reads* são o resultado da reação de sequenciamento e posteriormente são utilizados para leitura das bases que estão na sequência.

Após o sequenciamento das amostras, os dados brutos recebidos passaram por uma análise de qualidade utilizando-se a ferramenta FastQC (Babraham Bioinformatics) e After QC. Desses, foram obtidas estatísticas básicas sobre a biblioteca tais como: tipo de arquivo, formato, total de sequências, sequências filtradas, comprimento das sequências e porcentagem

de conteúdo de guanina + citosina (GC) de cada base em cada sequência gerada. Também foram verificados parâmetros de score de qualidade por sequência, qualidade da sequência por base, qualidade da curva, descontinuidade da curva, conteúdo GC por sequência, conteúdo de N por base e distribuição do comprimento das sequências.

Figura 9 - Representação da qualidade de reads através pelo fastQC. O gráfico mostra um valor de qualidade (eixo Y) para todas as bases em cada uma das posições do read (eixo X). A linha azul representa a qualidade média. O gráfico é dividido em três: verde (boa qualidade), laranja (qualidade razoável) e vermelho (baixa qualidade).

Para a limpeza de ruídos das sequências de cada biblioteca, foi feito um *trimming*, seguido da remoção de *reads* curtos e com sobreposição ruim, remoção de *reads* com baixa qualidade de sequência, com poly-X ou excesso de N.

4.8. Alinhamento de *contigs*

Contig é o nome dado à junção de várias *reads* sobrepostas que formam uma sequência. Os *contigs* foram gerados com o programa SPAdes, usando diversos tamanhos de k-mer e selecionando os de resultado com maiores

contigs. Primeiramente, os *contigs* foram comparados às sequências de um banco local curado somente para HBV, usando a ferramenta de alinhamento local BLAST (Basic Local Alignment Search Tool). Isso foi realizado com o objetivo de identificar a similaridade com as sequências do banco, permitindo identificar genótipo e subgenótipo. Para a classificação de cada *contig*, foram utilizados os campos: query name, que é o nome da sequência de consulta; subject name, que é nome da sequência cujo alinhamento foi encontrado; o campo percent identities é porcentagem de bases idênticas; aligned length é o tamanho do alinhamento; e o e-value, que é a probabilidade do alinhamento ser uma coincidência. Para montar a sequência final, utilizamos múltiplos contigs de cada amostra, conforme o alinhamento local do BLAST.

Figura 10 - Fluxograma de trabalho para análise de dados das sequências geradas

4.9. Análise Filogenética

As sequências geradas foram alinhadas com sequências de referência de cada genótipo e subgenótipo do HBV, incluindo várias sequências dos genótipos F e H isoladas na América Latina.

Para o alinhamento foi utilizado o programa Clustal_W integrado ao programa BioEdit (108). Posteriormente, o alinhamento foi submetido à análise filogenética utilizando-se o programa BEAST v.1.8.3 (109) com 10.000.000 etapas da Cadeia de Markov Monte Carlo (MCMC) amostragem de cada 1.000

passos, e rejeitando os primeiros passos como 1.000.000 *burn-in*. O programa FigTree v1.4.2 foi utilizado para visualização e edição das árvores filogenéticas.

4.10. Análise de Variantes

Na chamada de variantes foi utilizada a ferramenta Freebayes (109) e a anotação funcional com o programa SnpEff (110), que é capaz de analisar e anotar milhares de variantes e prever seus possíveis efeitos genéticos. O filtro de qualidade foi definido considerando uma correção de Bonferroni (110) partindo-se de um cut-off de $Q=30$, ou seja, uma taxa de falsos positivos de $1/1000$, onde foram avaliados ~ 3215 sítios com um cut-off corrigido de $1.8e-8$, que corresponde a $Q=77$.

Foram colocadas como preferência no filtro de análise as principais variantes de alto impacto encontradas na literatura do HBV sobre os genes pré-core e promotor basal do core, preS, S, X e gene P. Todas as variantes foram levantadas em comparação com a sequência KJ843191, que trata-se de um genótipo F2a, com genoma completo depositado no GenBank.

5. RESULTADOS

5.1. Amplificação de genomas completos do HBV

Todas as 90 amostras (HBV/F=59; HBV/H=31) incluídas no projeto foram processadas utilizando algum dos protocolos de amplificação do genoma completo do HBV descritos.

O primeiro protocolo a ser testado para amplificação foi o protocolo A, que utilizou os primers P1 e P2 e resultou num produto de 3,2 Kb, correspondente ao genoma completo do HBV (Figura 11). Inicialmente todas as 59 amostras pertencentes ao genótipo F foram testadas com esse protocolo e 18 apresentaram bandas do tamanho esperado visíveis no gel de agarose. Para o genótipo H, 22 amostras foram testadas e nenhuma apresentou banda visível no gel de agarose.

Figura 11 - Análise por eletroforese em gel de agarose de produtos de PCR gerados com os primers P1 e P2, protocolo A.

Após o resultado desse primeiro protocolo, para evitar desperdício dos reagentes e de amostra, os próximos testes foram divididos em etapas. A princípio, das 41 amostras nas quais não foi possível visualizar banda no gel de agarose pelo protocolo A, 35 amostras genótipo F foram separadas e submetidas à amplificação utilizando-se os primers do protocolo B (P1.2/ P194; PS3076F/ P1196R e 5'LAM5/P2.2). Por meio dessa estratégia foi possível amplificar o genoma completo do HBV a partir de três fragmentos sobrepostos, como pode

ser observado na Figura 12. Das amostras testadas neste protocolo, 5 exibiram bandas do tamanho esperado visíveis no gel de agarose.

Figura 12 - Análise por eletroforese em gel de agarose de produtos de PCR amplificados com os primers P1.2/ P194; PS3076F/ P1196R e 5'LAM5/ P2.2; protocolo B.

Para a amplificação das 36 amostras genótipo F que não apresentaram bandas no gel nos protocolos anteriores, foi realizado o protocolo de *nested* PCR para amplificação de dois fragmentos sobrepostos (protocolo C). As amostras foram separadas e testadas, e foi possível amplificar o genoma completo de 5 amostras (Figura 13). Para o genótipo H, as 22 amostras já testadas e não amplificadas com o protocolo A e as 9 amostras que não haviam sido processadas ainda, foram testadas utilizando o protocolo C e 3 tiveram o genoma completo do HBV amplificado.

Figura 13 - Análise por eletroforese em gel de agarose de produtos de PCR do protocolo C. Poços 1 a 6 representam fragmentos de 2092 pb. Poços 7, 9 e 10 representam fragmentos de 1320 pb.

O fluxograma de trabalho utilizado para amplificação do genoma completo está representado na figura abaixo (Figura 14).

Figura 14: Representação esquemática do processo de trabalho feito para amplificação do genoma completo do HBV

5.2. Sequenciamento

No total foi possível amplificar o genoma completo do HBV de 31 amostras das 90 testadas, sendo 28 pertencentes ao genótipo F e 3 ao genótipo H. Todas foram submetidas ao sequenciamento de nova geração.

5.3. Análise de dados gerados

Uma boa montagem de genoma requer a combinação adequada entre cobertura de sequenciamento e *reads* com comprimento e qualidade adequados (110). Os arquivos gerados foram avaliados quantos aos critérios de qualidade e cobertura, os quais estão representados no Quadro 1.

A maioria das amostras apresentou uma cobertura mediana acima de 800, isso significa que o número de *reads* que cobriram cada base atingiu uma qualidade considerada ótima no sequenciamento. Da mesma forma, a maioria das amostras apresentou uma cobertura geral acima de 85%, o que significa que os *reads* se concentraram, em sua maioria, em toda a extensão do genoma.

Das 31 amostras submetidas ao sequenciamento, 18 foram sequenciadas com sucesso e 13 não apresentaram bons níveis de qualidade e cobertura, o que pode ter sido em decorrência do baixo número de *reads* gerados na reação de sequenciamento, impossibilitando, portanto, a montagem do *contig* do genoma completo do HBV.

Quadro 1 - Resultados de qualidade e cobertura das amostras incluídas no projeto obtidos na análise de dados do sequenciamento

Amostra	Cobertura Geral	Cobertura Média	Cobertura Mediana
15457	100%	1997,1	1895
AAS	100%	500,5	385
ANT	100%	3870,6	3601
BH144	100%	1550,2	854
BH57	100%	1474,3	1461
CO20	100%	2276,6	2066
HBNTA62	100%	1111,3	1113
PA17	100%	1338,3	1296
VOWS	100%	2372,7	2367
PA26	98%	215,4	228
TN102	100%	8487,9	2966
1634	100%	5072,2	5086
468	100%	2017,1	2025
8069	87%	4877,5	5715
RP175	100%	29875,5	29001
RP176	100%	332,6	278
BH206	100%	10495,7	11084
TRA49	100%	32546,9	33332

A montagem dos genomas foi feita e sua identificação baseou-se nos resultados do alinhamento local que compararam as sequências produzidas com o banco de sequência de nucleotídeos não redundantes do (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>). No quadro 2 é possível observar o percentual de similaridade dos genomas montados com as sequências referência no GenBank. O genótipo identificado nas análises foi aquele mais similar com o da sequência de referência.

Quadro 2 - Resultado da comparação dos contigs gerados com sequências referência do banco de dados NCBI (BLAST)

Amostra	Referência (BLAST)	% de identidade	Genótipo e Subgenótipo
15457	KC494403	99,35%	F2a
AASF	KC494394	99,94%	F2a
ANTF	KC494403	98,78%	F2a
BH144	KX264497	98,57%	F2a
BH57	KC494403	99,72%	F2a
CO20	FJ709457	99,80%	F1b
CO20	MH051986	99,22%	F3
HBNTA62	KC494402	99,56%	F2a
PA17	KC494403	99,69%	F2a
VOWS	KC494403	99,94%	F2a
PA26	X69798	98,64%	F2a
TN102	KC494403	99,32%	F2a
1634	HM117851	99,32%	H
468	HM066946	98,54%	H
8069	X69798	99,71%	F2a
RP175	KC494403	98,72%	F2a
RP176	X69798	99,07%	F2a
BH206	KC494402	99,44%	F2a
TRA49	KC494402	99,53%	F2a

As sequências do genoma completo do HBV geradas aplicando-se a metodologia de NGS apresentaram >97% de semelhança com sequências referências depositadas no banco de dados consultado, reafirmando os genótipos identificados previamente no sequenciamento parcial do genoma e identificando os subgenótipos, uma vez que não eram todos conhecidos anteriormente. Não foram identificados eventos de recombinações nas sequências HBV/F e HBV/H caracterizadas.

Todas as amostras pertencentes ao genótipo F foram classificadas como subgenótipo F2a, com exceção da amostra CO20 na qual foram identificados reads pertencentes a dois subgenótipos diferentes, onde o maior read apresentou 99% de semelhança a sequência referência do subgenótipo F3 e o outro read gerado de tamanho menor apresentou 99,80% de semelhança a uma sequência referência completa do subgenótipo F1b. Esses dois *reads* alinharam-se à mesma porção do genoma do HBV em suas respectivas sequências referência: o primeiro *contig* alinhou-se à todo o genoma da referência subgenótipo F3, de 2 a 3215, com 99,80% de identidade; o segundo *contig* apresentou uma identidade de 99,80% com as regiões dos nucleotídeos 2013 a 3019 e 2 a 286 de uma sequência referência do genótipo F1b, ou seja, 40% do genoma completo, o que evidência a coinfeção com dois subgenótipos do genótipo F, sendo representado na figura abaixo (Figura 15):

Figura 15 - Representação do alinhamento do genoma da amostra CO20, onde o menor contig alinhou-se à referência do genoma referência de uma amostra F1b (a esquerda), e o maior contig alinhou-se à referência F3.

As duas amostras pertencentes ao genótipo H agruparam-se, como esperado, com as demais sequências referências isoladas também no México.

Tais resultados foram também confirmados pela realização da análise filogenética das sequências geradas (Figura 16 e Figura 17). A análise foi realizada pelo método Bayesiano a partir de 185 sequências do genoma completo do HBV de diferentes genótipos (167 do GenBank e 18 deste estudo).

O resultado da filogenia das sequências F2a caracterizadas nesse estudo aponta para a formação de um agrupamento geográfico específico com sequências F2a circulantes no Brasil, porém não se observa a formação de clados de acordo com a região de procedência da amostra dentro do país, sequências isoladas de casos do norte ou sudeste do país apresentam proximidade no agrupamento.

Em relação às sequências F1b e F3 isoladas da amostra CO20 se observa que ambas agrupam próximas a sequências dos respectivos subgenótipos isoladas na Venezuela (Figura 16), o que sugere uma origem comum dessas cepas identificadas no mesmo paciente.

Figura 16 - Árvore filogenética das sequências de genoma completo do HBV genótipo F caracterizadas usando sequenciamento de nova geração. Os valores de probabilidade posterior são apresentados acima dos ramos principais.

Figura 17 - Árvore filogenética das sequências de genoma completo do HBV genótipo H caracterizadas usando sequenciamento de nova geração. Os valores de probabilidade posterior são apresentados acima dos ramos principais.

Na anotação de mutações foram avaliadas mutações com impacto clínico associado à evolução da doença, escape vacinal, resistência os análogos nucle(s)ídeos. Foram avaliadas as variantes nas regiões pre-core, core, pre S, S, X e polimerase.

a) Genes Pré-S

As mutações de maior impacto observadas nessas regiões foram aquelas associadas com alteração da sua estrutura, ausência ou interrupção prematura da síntese da proteína (Quadro 3), correlacionando-se, particularmente deleções (111) (112), com uma forma mais progressiva da doença hepática (113) (114).

Códon de parada prematuro no aminoácido 85 da região pré-S1 (ps1L85STOP) foi observado na amostra BH57 (genótipo F2a), sendo que 40% da população viral apresentava essa alteração. Variantes com mutação no códon de iniciação do gene pré-S2 foi observada na amostra TN102 (genótipo F2a) com frequência de 98%. Variantes com deleção no gene pré-S2 foram observadas na amostra TRA62 (genótipo F2a) com frequência de 55% e compondo 94,6% da população viral da amostra 1634, pertencente ao genótipo H.

b) Gene S

As mutações analisadas na região S foram aquelas relacionadas com alterações na síntese da proteína e mutações dentro da principal região hidrofílica do HBsAg, localizada entre os aminoácidos 100 a 169, que podem levar a redução na capacidade de ligação dos anticorpos anti-HBs à proteína (Quadro 3).

Cepas com códons de parada prematuro no gene S (sC69STOP, sW196STOP ou sL216STOP) foram identificadas em 5 amostras, sendo essas F=3 e H=2, compondo de 12,5 a 70% da população viral. Em 9 amostras, genótipo F=7 e genótipo H=2, foram identificadas cepas do HBV com mutações dentro da principal região hidrofílica do HBsAg, sendo que apenas duas das

mutações observadas (sQ129H e sS140T) já foram associadas com escape dos anticorpos anti-HBs. Cepas do HBV com as mutações sQ129H ou sS140T foram identificadas em 6 amostras: 8096, BH144, BH57, PA17, RP176 e amostra 468, sendo que em 3 dessas cepas representavam a população viral majoritária (29,5% e 93,7%).

c) Gene Polimerase

A análise de mutações foi focada nos diferentes domínios (A-E) da região com função de transcriptase reversa e interdomínios B-C e C-D, onde a presença de mutações pode ter o impacto clínico da resistência ao tratamento com análogos nucleot(s)ídeos, principalmente a Lamivudina, podendo ocorrer por causa do tratamento com esse mesmo antirretroviral ou naturalmente devido à incorporação incorreta de nucleotídeos aleatórios em pacientes com a forma crônica da doença (115) (116). Mutações em 16 diferentes aminoácidos codificados por essa região genômica foram identificadas em 9 amostras (Genótipo F=7 e Genótipo H=2), sendo que nenhuma dessas mutações conferem resistência comprovada aos AN, apenas duas já foram associadas com potencial impacto na resposta ao tratamento com Adefovir (rtS85A e rtS219A). Dentre essas apenas a rtS219A foi observada na população viral majoritária (frequência de 88,8 a 94%), na amostra ANT123.

d) Região do promotor basal do core, genes pré-core, core e X

Nessa região foram investigadas mutações nas posições 1653, 1753, 1762 (xL130M), 1764 (xV131I), 1766, 1768, 1814, 1815, 1816, 1858, 1862 e 1888, sendo essas segundo a literatura, envolvidas com o desenvolvimento da doença de forma mais agressiva, incluindo a reativação da forma crônica e até curso fulminante da doença (117) (118), além da progressão para o HCC (119) (120).

Nas amostras analisadas foram observadas variantes com mutações na região do promotor basal do core: A1762T (4 amostras: 15457, ANT123, TN102

e 1634) com frequência de 78,3 a 99,5% e G1764A (1 amostra: PA26) constituindo 98,1 % da população viral. Em nenhuma amostra essas mutações foram observadas concomitantemente. As substituições de nucleotídeos nas posições 1762 e 1764 citadas anteriormente levam a mutações na proteína X nas posições 130 (xL130M) e 131 (xV131I), respectivamente. Outra mutação no HBx associada com impacto clínico foi observada no aminoácido 127 (xI127L) no genoma do HBV F2a isolado da amostra RP17.

Variantes com mutações (substituições ou deleções) no códon de iniciação do gene pré-core foram encontradas em 11 amostra (Genótipo F=8 e Genótipo H=2), em todos os casos representando a população viral dominante (87,5 a 100%).

Quadro 3 – Descrição das principais mutações relacionadas com impacto funcional nas amostras sequenciadas no presente estudo

Amostra	Subgenótipo	Gene	Mutação	Frequência na população viral
15457	F2a	S	C69STOP*	58,00%
		RT	S78T*	58,00%
		BCP	A1762T	99,30%
		PC	M1?	98,00%
		X	L130M	99,30%
8069	F2a	S	S140T	15,10%
ANTF123	F2a	RT	N76D	19,40%
		RT	S219A	94,40%
		RT	H234Q	11,80%
		BCP	A1762T	97,10%
		X	L130M	97,10%
BH144	F2a	S	S140T	93,73%
		PC	M1?	98,00%
BH57	F2a	PS1	L85STOP	40,20%
		S	C69STOP*	30,90%
		S	L109R	26,77%
		S	S140T	11,09%
		RT	S78T*	30,90%
		RT	S85A	3,60%
CO20	F2a	S	K160R*	15,45%
		RT	I169L*	15,50%
		RT	S238A	86,10%
		PC	M1?	92,10%
PA17	F2a	S	S140T	64,12%
		PC	M1?	100,00%
PA26	F2a	S	L122R	100,00%
		BCP	G1764A	98,10%
		X	V131I	98,10%
RP175	F2a	RT	V214E	97,40%
RP176	F2a	S	W196STOP*	12,50%
		S	L216STOP	12,50%
		S	L110I	28,84%
		S	L127P	7,08%
		S	S140T	3,45%
		S	L158F*	11,92%
		S	A168V	38,19%

		RT	L91I	15,60%
		RT	I163V*	29,50%
		RT	D205N*	12,50%
		RT	L207V	42,90%
		RT	V214E	32,80%
		RT	Q215H	8,50%
		RT	V231P	24,20%
		PC	M1?	100,00%
		X	I127T	68,80%
TN102	F2a	PS2	M1I	97,90%
		BCP	A1762T	99,50%
		PC	M1?	93,30%
		X	L130M	99,50%
AAS	F2a	PC	M1?	100,00%
TRA62	F2a	PS2	del nt 37 - 56	54,90%
		RT	V253I	100,00%
		PC	M1?	94,00%
VOWS	F2a	PC	M1?	87,50%
468	H	S	L216STOP	14,40%
		S	Q129H	29,48%
		RT	I163V	35,40%
		RT	I253V	71,10%
		PC	M1?	100,00%
		X	H94Y	76,60%
1634	H	PS2	del nt 47 - 55	94,60%
		S	C69STOP*	70,10%
		S	K160R	16,22%
		S	S167L	57,55%
		RT	S78T*	70,10%
		RT	A219A	88,80%
		RT	I233V	13,80%
		RT	I253V	99,40%
		BCP	A1762T	78,30%
		PC	M1?	100,00%
		X	L130M	78,30%

* Alterações de aminoácidos nas proteínas S e Polimerase provocadas pela substituição do mesmo nucleotídeo

6. DISCUSSÃO

A infecção pelo HBV é considerada um sério problema de saúde pública em todo o mundo, especialmente em países menos desenvolvidos. Estima-se que 70% das novas infecções crônicas ocorram em países de baixa renda (121). Estima-se também que 257 milhões de pessoas em todo o mundo sejam portadores crônicos do HBV, e que 887.220 mortes são em decorrência da infecção, a maioria das quais relacionadas a complicações como cirrose e carcinoma hepatocelular. Informações sobre a epidemiologia do HBV em países da América do Sul geralmente são fragmentadas, mas é estimado que entre 7 e 12 milhões de latino-americanos sejam portadores crônicos dessa doença (122) (123). A taxa de pacientes HBsAg positivos varia, sendo os valores mais altos detectados em jovens adultos de 20 a 40 anos, infecções possivelmente relacionadas com a transmissão horizontal (124) (125).

Devido ao seu ciclo de vida único, no qual uma transcriptase reversa propensa a erros é empregada para a replicação do genoma, o HBV apresenta uma grande variabilidade genética, resultando em 10 genótipos (A – J), quase 40 subgenótipos, diferentes recombinantes, clados e quasi-espécies (126). Dentre os 10 genótipos conhecidos até então, os genótipos F e H predominam nas populações indígenas da América Latina (127) (128), sendo o genótipo F considerado como o mais antigo da América (129). Os quatro subgenótipos do genótipo F (F1-F4) predominam na América Central, são mais frequentes em ameríndios e estão presentes em todos os países da América do Sul [174-176]. O genótipo H é encontrado quase que predominantemente em mestiços no México (130), mas também foram registrados casos pontuais em outros países: Argentina (131), Brasil (83), Turquia (132) e Japão (133). Entre os dez genótipos conhecidos do HBV, esses dois são os mais destoantes dos outros oito, chegando a >14% de divergência na sequência completa, tendo assim uma relação filogenética próxima entre si sugerindo que provavelmente se originaram no continente americano antes da colonização europeia (134). Apesar disso, não há evidências conclusivas sobre os ancestrais desses genótipos e sua relação filogenética com as outras cepas humanas e não humanas (135) (136).

O presente estudo analisou 90 amostras que apresentaram previamente resultado positivo para o HBV no teste molecular PCR e sequenciadas pela metodologia Sanger sendo classificadas como genótipos F (n=59) e H (n=31).

Durante as últimas décadas, os métodos de diagnóstico molecular experimentaram um rápido desenvolvimento. PCR é uma técnica que revolucionou a ciência, pois permitiu rápido desenvolvimento do estudo de ácidos nucleicos, proporcionando grande avanço em diversas áreas como na genética humana e de microrganismos e no diagnóstico de doenças infecciosas.

A amplificação do genoma completo do HBV vêm sendo um desafio para pesquisadores desde 1995 quando Günther e cols. (105) propuseram em seu estudo uma maneira pioneira de amplificação do genoma completo do vírus utilizando um par de primers. Nesse estudo, apesar de audaciosa a ideia de usar somente um par de primers para amplificar 3200 pares de bases de uma só vez foi aplicada e foi utilizada também a técnica de clonagem em plasmídeos do produto da primeira PCR para amplificar o segmento de DNA do vírus neles inserido. Os próximos estudos realizados nas seguintes décadas sobre o assunto continuam tendo certa dificuldade de reprodução, e em sua maioria os estudos caracterizam um número pequeno de amostras, como podemos ver, por exemplo, pelos estudos de Sasha e cols (137), no qual foram utilizado 6 primers para amplificação do genoma completo do HBV de somente uma amostra; de Spitz e cols (138) que utilizou o mesmo método de Günther de clonagem em plasmídeos para amplificação de também apenas uma amostra; de Fonseca e cols (139) que usou 10 primers para amplificação de três amostras; e até mesmo o estudo que caracterizou o genótipo G, de Stuyver e cols (50), que diferente dos outros estudos caracterizou o genoma completo das amostras fazendo amplificação com um número maior de primers, utilizando 21 pares em uma reação nested e semi-nested PCR.

No presente trabalho, ainda na investida de otimizar a reação de amplificação do genoma completo do HBV, além de dispor-se de um número menor de pares de primers, também foi empregada a enzima Q5 high-fidelity, comprovada no estudo de Wang e cols (140) como sendo a melhor enzima para amplificação do HBV, pois além de ter uma taxa de erro menor que 10 vezes em

relação às demais enzimas comumente utilizadas, também reduz o tempo necessário para a realização da amplificação de um fragmento de DNA longo, chegando a uma hora de diferença das demais.

Outra parte importante para o sucesso da execução do teste de PCR está relacionada à carga viral presente na amostra investigada. Essa relação é diretamente proporcional, quanto maior a carga viral na amostra, maior sua amplificação e mais fácil a montagem, no caso, de um genoma completo confiável. O genótipo H do HBV é em diversos estudos relacionado com baixa carga viral, como demonstrado por Sozzi e cols (141) no estudo onde comprova-se através de testes com cultura celular que esse genótipo tem taxas mais baixas de replicação e de expressão de HBsAg e HBeAg em comparação com outro genótipo.

Ainda assim, a utilização de um número grande de amostras em um estudo de caracterização de genoma completo do HBV não é algo corriqueiro, uma vez que sua reprodutibilidade é baixa. Não existe uma abundância de estudos de biologia molecular e bioinformática comparativa das diversas cepas de HBV, e os existentes são na grande maioria estudos caracterizando os genótipos B e C, que juntos correspondem a mais de 55% dos genomas completos publicados (dados de 2019). A caracterização do genoma completo de um vírus pode ajudar no diagnóstico clínico, seja ao revelar o agente envolvido na infecção ou na identificação de variações genéticas clinicamente importantes. A exemplo disso temos estudos que demonstram que, mutações do promotor precore/core e mutações, incluindo deleções no pré-S sendo relatadas em associação à doença hepática progressiva e risco de desenvolvimento do CHC (142).

Durante a replicação do HBV ocorre um alto grau de variação na sequência nucleotídica devido à falta de atividade corretiva da enzima viral durante a etapa de transcrição reversa, assim é necessário ter uma investigação abrangente sobre as sequências de aminoácidos das proteínas do HBV, a fim de explorar a variabilidade e a conservação dos domínios que determinam sua função, bem como as variações entre as diferentes cepas. Em 2019, Cao e cols. (143) demonstraram em seu estudo que, devido ao genoma do HBV ser

compacto e sobreposto, quando há mutação em uma ORF, o mesmo provavelmente também acontece em outra, o que pode acabar afetando a função ou a transcrição das proteínas virais. Apesar de o sequenciamento e a genotipagem de isolados de HBV não serem feitos rotineiramente, os diferentes genótipos do vírus podem levar a diferentes consequências clínicas durante a evolução do curso natural da infecção e progressão da doença. O genótipo F, por exemplo, é em diversos artigos colocado como sendo responsável pela progressão mais rápida da doença para cirrose hepática e CHC, com taxas de mortalidade mais altas do que aquelas observadas nas infecções com o genótipo A ou D, mesmo sendo detentor de menos de 2% das infecções crônicas mundialmente (144) (96). Já o genótipo H está associado a baixa incidência de CHC (98), mas maior incidência de hepatite oculta, devido a caracterização de níveis mais baixos de DNA e expressão de HBsAg reduzida ou negativa (145).

Em relação à vacinação há vários relatos de falha devido principalmente a mutações relacionadas ao gene S, uma vez que as vacinas geram resposta a epítomos localizados na proteína HBsAg. Em 2011, Tackle e cols (146) relataram o caso de um paciente que desenvolveu hepatite B aguda após infecção pelo subgenótipo F1b, mesmo tendo sido submetido à vacinação completa. Não foram identificadas mutações no gene S as mutações classicamente associadas com escape dos anticorpos anti-HBs, como a sG145R, porém os autores destacam as diferenças marcantes na sequência do gene S do genótipo F em relação aos demais genótipos. O estudo realizado por Limeres e cols. (147) fortalece a evidência da influência da variabilidade do gene S do genótipo F, que podemos estender ao genótipo H, na redução da capacidade de ligação aos anticorpos anti-HBs. No referido estudo os autores avaliaram a influência da variabilidade genética do gene S do genótipo F na sensibilidade de dois kits comerciais amplamente utilizados para a pesquisa do HBsAg, e observaram que as taxas de detecção desse antígeno produzido pelos subgenótipos F1b e F4 diferiram significativamente daquelas obtidas com antígenos produzidos pelo HBV genótipo A.

Assim como mutações em diversas regiões do genoma do HBV foram correlacionadas à resistência antiviral, escape da vacina e carcinoma

hepatocelular, eventos de coinfeção e recombinação viral também podem desencadear uma maior virulência e piora no quadro clínico dos pacientes. A recombinação é fácil de ser detectada no processo de montagem de árvores filogenéticas, no entanto, a coinfeção de dois genótipos diferentes pode ser fruto principalmente de artefatos durante a amplificação por PCR ou pelo processamento errado dos dados da sequência, levando à conclusão errônea de que há uma coinfeção. Em 2000, Morozov e cols. (148) alegaram que para obter sequências confiáveis do evento de coinfeção viral, é aconselhável que a sequência de nucleotídeos do genoma completo seja feita através de um único fragmento de PCR. No presente estudo também houve a caracterização de uma coinfeção viral entre os genótipos F1b e F3, feita através da amplificação dos 3215pb do genoma do HBV pela utilização dos primers P1 e P2 descritos por Gunther.

A distribuição de mutações no genoma do HBV com implicações clínicas representa um desafio para diagnóstico e as diferentes estratégias de tratamento, podendo até impactar diretamente no sucesso a longo prazo dos programas de vacinação da população (149). O conhecimento da prevalência de mutações clinicamente relevantes entre os diferentes genótipos e subgenótipos facilita a melhoria dos procedimentos diagnósticos, programas de imunização, protocolos terapêuticos e prognóstico da doença (150). Essas mutações podem ser analisadas por região do genoma e comparando a sequência de nucleotídeos do gene específico e seu impacto na sequência de aminoácidos.

Sendo assim, o NGS tem se mostrado cada vez mais como uma ferramenta poderosa no diagnóstico médico. Através dessa técnica é possível sequenciar fragmentos do genoma do HBV sendo possível identificar além de genótipos mistos e quasispécies do vírus, variações e polimorfismos importantes na avaliação de resistência à medicamentos.

7. CONCLUSÕES

- O método aplicado permitiu a caracterização de genomas completos do HBV genótipos F e H e possibilitou a identificação de infecções mistas com diferentes subgenótipos;
- Os genomas do genótipo F caracterizados pertenciam predominantemente ao subgenótipo F2a, e um caso de coinfecção com subgenótipos F3 e F1b também foi identificado;
- A ocorrência de eventos de recombinação não foi observada entre as amostras HBV genótipos F e H sequenciadas;
- Foram identificados diferentes tipos de mutações (deleções, inserções e substituições) nos genes virais analisados, assim como a frequência das variantes com essas alterações na população viral de cada amostra;
- As principais alterações genéticas observadas nas cepas HBV/F e H caracterizadas nesse estudo foram:
 - alterações nos gene Pré-S e S, que promovem interrupção prematura da síntese ou não produção das proteínas de superfície do HBV (HBsAg, LHBs e MHBs). Nesses genes foram identificadas 7 diferentes mutações isoladas ou combinadas em 38,8% (7/18) das amostras analisadas. Essas variantes foram observadas em alguns casos compondo a população viral majoritária e em outro minoritária;
 - mutações no códon de iniciação do gene pré-core, que promovem a falta de síntese da proteína HBeAg, foram identificadas em 61,1% (11/18) das amostras. Em todos os casos as cepas com essas alterações constituíram mais de 87 – 100% da população viral.
- Variantes com mutações associadas com escape dos anticorpos anti-HBs e com resistência aos análogos núcleo(s)ídeos foram pouco frequentes entre as amostras HBV/F e H analisadas,. assim como as mutações no promotor basal do core e gene X.

8. REFERÊNCIAS BIBLIOGRÁFICAS

1. **Freitas, J.** Hepatitis víricas: perspectiva histórica. [Online] 2003. [Citado em: 16 de 05 de 2017.] <http://www.aidsportugal.com/hepatitte>.
2. **Reuben, A.** *Landmarks in hepatology: the thin red line*. s.l. : Hepatol, 2002. Vol. 36.
3. **MacCallum, F. O.** *Homologous serum jaundice*. s.l. : Lancet, 1947. Vol. 2.
4. **ICTV.** Hepatitis B virus. *ICTVdB - The Universal Virus Database*. [Online] 2006. <http://www.ncbi.nlm.nih.gov/ICTVdb/ICTVdB/>.
5. **Blumberg, B. S., Alter, H. J. e Vinich, S.** A “New” Antigen in Leukemia Sera. s.l. : Jama, 1965. Vol. 191.
6. **Prince, A. M.** *An antigen detected in the blood during the incubation period of serum hepatitis*. s.l. : Proc Natl Acad Sci USA, 1968. Vol. 60.
7. **Okochi, K. e Murakami, S.** *Observations on Australia antigen in Japanese*. s.l. : Vox Sang, 1968. Vol. 15.
8. **Dane, D. S., Cameron, C. H. e Briggs, M.** *Virus-like particles in serum of patients with Australia-antigen-associated hepatitis*. s.l. : Lancet, 1970. Vol. 1.
9. **Purcell, R. H.** *The Discovery of the Hepatitis Viruses*. s.l. : Gastroenterology, 1993. Vol. 104.
10. **World Health Organization.** *Introduction of hepatitis B vaccine into childhood immunization services*. s.l. : WHO/V&B/01.31, 2001.
11. **Ferreira, M. S. e Borges, A. S.** *Avanços no tratamento da hepatite pelo vírus B*. s.l. : Revista da Sociedade Brasileira de Medicina Tropical, 2007. Vol. 40.
12. **Purow, D. B. e Jacobson, I. M.** *Slowing the progression of chronic hepatitis B. Early antiviral therapy can help minimize complications*. s.l. : Postgrad Med, 2003. Vol. 114.

13. **Wang, F. S.** *Current status and prospects of studies on human genetic alleles associated with hepatitis B virus infection.* s.l. : World J Gastroenterol, 2003. Vol. 9.
14. **Karawasa, T., et al.** *Association between frequency of amino acid changes in core region of hepatitis B virus (HBV) and the presence of precore mutation in Japanese HBV carriers.* s.l. : J Gastroenterol, 1997. Vol. 35.
15. **Torresi, J.** *The virological and clinical significance of mutations in the overlapping envelope and polymerase genes of hepatitis B virus.* s.l. : J Clin Virol, 2002. Vol. 25.
16. **Locarnini, S.** *Hepatitis B viral resistance: mechanisms and diagnosis.* s.l. : Journal of Hepatology, 2003. Vol. 39.
17. —. *Molecular virology of hepatitis B virus.* s.l. : Seminars in Liver Disease, 2004. Vol. 24.
18. **Mahoney, F. J.** *Update on Diagnosis, Management, and Prevention of Hepatitis B Virus Infection.* s.l. : Clin Microbiol Rev, 1999. Vol. 12.
19. **Hollinger, F. B. e Liang, T. J.** *Hepatitis B virus.* s.l. : Fields Virology, 2001. Vol. 4.
20. **Coordenação de Vigilância em Saúde COVISA.** *Programa Municipal de Hepatites Virais.* s.l. : Secretaria Municipal de Saúde de São Paulo -SMS/SP, 2016.
21. **Aguilera-Guirao, A., Romero, S. Y. e Regueiro, B. J.** *Epidemiology and clinical manifestations of viral hepatitis.* s.l. : Enferm Infec Microbiol Clin, 2006. Vol. 24.
22. **Melo, F. C. A. e Isolani, A. P.** *Hepatite B e C: Do Risco de Contaminação por Materiais de Manicure/Pedicure à Prevenção.* s.l. : R. Saúde e Biol., 2011. Vol. 6.
23. **Te, H. S. e Jensen, D. M.** *Epidemiology of hepatitis B and C viruses: a global overview.* s.l. : Clin Liver Dis, 2010. Vol. 14.

24. **Centers for Disease Control and Prevention.** *Epidemiology and Prevention of Vaccine-Preventable Diseases*. s.l. : Public Health Foundation, 2011. Vol. 12.
25. **Merrill, R. M. e Hunter, B. D.** *Seroprevalence of markers for hepatitis B viral infection*. s.l. : Int J Infect Dis, 2011. Vol. 15.
26. **CDC.** *Travelers' health; yellow book*. s.l. : CDC - Department of Health and Human Services, 2008.
27. **Liaw, Y. F. e Chu, C. M.** *Hepatitis B Virus Infection*. s.l. : Lancet, 2009. Vol. 373.
28. **Ministério da Saúde.** *Saúde amplia faixa etária para vacinação gratuita contra hepatite B a partir de 2011*. 2010.
29. **Secretaria de Saúde do Estado de São Paulo.** *Vacina contra hepatite B*. s.l. : Rev Saúde Pública, 2006.
30. **Seitz, S., et al.** *Cryoelectron microscopy of hepatitis B virions reveals variability in envelope capsid interactions*. s.l. : EMBO, 2007. Vol. 26.
31. **Glebe, D. e König, A.** *Molecular Virology of Hepatitis B Virus and Targets for Antiviral Intervention*. s.l. : Intervirology, 2014. 57.
32. **Hussain, B. K. e Lok, A. S. F.** *Hepatitis B virology: acute and chronic Infection - wild-type HBV and HBV variants*. s.l. : Management of Chronic Viral Hepatitis, 2001.
33. **Glebe, D. e Bremer, C. M.** *The Molecular Virology of Hepatitis B Virus*. s.l. : Thieme Medical Publishers, 2013. Vol. 333.
34. **Gerlich, W. H.** *Medical Virology of Hepatitis B: how it began and where we are now*. s.l. : Virology Journal , 2013. Vol. 10.
35. **Norder, H., Couroucé, A. M. e Maginus, L. O.** *Molecular basis of hepatitis B virus serotype variations within the four major subtypes*. s.l. : J Gen Virol, 1992. Vol. 73.
36. **Yokosuka, O. e Arai, M.** *Molecular biology of hepatitis B virus: effect of nucleotide substitutions on the clinical features of chronic hepatitis B*. s.l. : Med Mol Morphol, 2006. Vol. 39.

37. **Ganem, D. e Schneider, R. J.** *Hepadnaviridae and their replication*. s.l. : Fields virology, 2001. Vol. 4.
38. **Norder, H., Courouce, A. M. e Magnius, L. O.** *Complete genomes, phylogenetic relatedness, and structural proteins of six strains of the hepatitis B virus, four of which represent two new genotypes*. s.l. : Virology, 1994. Vol. 198.
39. **Kidd-Ljunggren, K., Mlyakawa, Y. e Kldd, A. H.** *Genetic variability m hepatitis B viruses*. s.l. : J Gen Virol, 2002. Vol. 83.
40. **Soriano, M. J. A.** *Virus de la Hepatitis B: estructura genómica y marcadores clínicos*. 2003.
41. **Sheldon, J., et al.** *Mutations affecting the replication capacity of the hepatitis B virus*. s.l. : J Viral Hep, 2006. Vol. 13.
42. **Ghany, M. e Liang, T. J.** *Drug Targets and Molecular Mechanisms of Drug Resistance in Chronic Hepatitis B*. s.l. : Gastroenterol, 2007. Vol. 132.
43. **Carrilho, F. J. e Ono-nita, S. K.,** O ABC das Hepatites: manual clínico para o manuseio e prevenção da Hepatite B. [A. do livro] E. S. A. Araujo. *Virologia do HBV*. s.l. : Bristol-Myers Squibb, 2008.
44. **Lin, C. L., et al.** *Application of hepatitis B virus genotyping and phylogenetic analysis in intrafamilial transmission of hepatitis B virus*. s.l. : Clin Infect Dis, 2005. Vol. 41.
45. **Lok, A. S. F. e Macmahon, B. J.** *Chronic Hepatitis B*. s.l. : Baltimore, 2007. Vol. 45.
46. **Nowak, T. H.** *HBV kinetics*. s.l. : Proc. Natl. Acad. Sci., 1996. Vol. 93.
47. **Okamoto, H., et al.** *Typing hepatitis B virus by homology in nucleotide sequence: comparison of surface antigen subtypes*. s.l. : J Gen Virol, 1988. Vol. 69.
48. **Le Bouvier, G. L.** *The heterogeneity of Australia antigen*. s.l. : J Infect Dis, 1971. Vol. 123.

49. **Norder, H., et al.** *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.* s.l. : J Gen Virol, 1992. Vol. 73.
50. **Stuyver, L., et al.** *A new genotype of hepatitis B virus: complete genome and phylogenetic relatedness.* s.l. : J Gen Virol, 2000.
51. **Arauz-Ruiz, P., et al.** *Genotype H: a new Amerindian genotype of hepatitis B virus revealed in Central America.* s.l. : J Gen Virol, 2002.
52. **Chu, C. J. e Lok, A. S.** *Clinical significance of hepatitis B virus genotypes.* s.l. : Hepatology, 2002.
53. **Lauder, I. J., et al.** *The variability of the hepatitis B virus genome: statistical analysis and biological implications.* 1993, Vol. 10.
54. **Karasawa, T., et al.** *Genetic heterogeneity in the precore region of hepatitis B virus in hepatitis B e antigen-negative chronic hepatitis B patients: spontaneous seroconversion and interferon-induced seroconversion.* s.l. : J Med Virol, 1995. Vol. 4.
55. **Grandjacques, C., et al.** *Rapid detection of genotypes and mutations in the pre-core promoter and the pre-core region of hepatitis B virus genome: correlation with viral persistence and disease severity.* s.l. : J Hepatol, 2000. Vol. 3.
56. **Kikuchi, L., et al.** *Genoepidemiology and its relationship to clinical features in patients infected chronically with hepatitis B virus.* s.l. : Hepatology Research, 2000. Vol. 14.
57. **Weinberger, K. M., et al.** *High genetic variability of the group-specific a-determinant of hepatitis B virus surface antigen (HBsAg) and the corresponding fragment of the viral polymerase in chronic virus carriers lacking detectable HBsAg in serum.* s.l. : J Gen Virol, 2000.
58. **Bowyer, S. M. e Sim, J. G. M.** *Relationships within and between genotypes of hepatitis B virus at points across the genome: footprints of recombination in certain isolates.* s.l. : J Gen Virol, 2000.

59. **Dumpis, U., et al.** *Transmission of hepatitis B virus infection in Gambian families revealed by phylogenetic analysis.* s.l. : J Hepatol, 2001. Vol. 35.
60. **Huang, C. C., et al.** *One single nucleotide difference alters the differential expression of spliced RNAs between HBV genotypes A and D.* s.l. : Virus Res, 2013. Vol. 174.
61. **Kurbanov, F., et al.** *When should "I" consider a new hepatitis B virus genotype?* s.l. : J Virol, 2008. Vol. 82.
62. **Shi, W., et al.** *Hepatitis B virus subgenotyping: history, effects of recombination, misclassifications, and corrections.* s.l. : Infect Genet Evol, 2013. Vol. 16.
63. **Cao, G. W.** *Clinical relevance and public health significance of hepatitis B virus genomic variations.* s.l. : World J Gastroenterol, 2009. Vol. 15.
64. **Kramvis, A., et al.** *Relationship of serological subtype, basic core promoter and precore mutations to genotypes/subgenotypes of hepatitis B virus.* s.l. : J Med Virol, 2008. Vol. 80.
65. **Guettouche, T. e Hnatyszyn, H. J.** *Chronic hepatitis B and viral genotype: the clinical significance of determining HBV genotypes.* s.l. : Antivir. Ther., 2005. Vol. 10.
66. **Olinger, C. M., et al.** *Possible new hepatitis B virus genotype, Southeast Asia.* s.l. : Emerg Infect Dis, 2008. Vol. 14.
67. **WHO.** *Hepatitis B. World Health Organization.* [Online] 2007. [Citado em: 21 de 02 de 2017.] <http://www.who.int/mediacentre/factsheets/fs204/en/>.
68. **Schaefer, S.** *Hepatitis B virus taxonomy and hepatitis B virus genotypes.* s.l. : World J Gastroenterol, 2007. Vol. 13.
69. **Allain, J. P.** *Epidemiology of Hepatitis B virus and genotype.* s.l. : J Clin Virol, 2006. Vol. 36.
70. **Andernach, I. E., Hubschen, J. M. e Muller, C. P.** *Hepatitis B virus: the genotype E puzzle.* s.l. : Rev Med Virol, 2009. Vol. 19.

71. **Kramvis, A. e Kew, M. C.** *Epidemiology of hepatitis B virus in Africa, its genotypes and clinical associations of genotypes.* s.l. : Hepatol Res, 2007. Vol.37.
72. **Devesa, M., et al.** *Subgenotype diversity of hepatitis B virus American genotype F in Amerindians from Venezuela and the general population of Colombia.* s.l. : J Med Virol, 2008. Vol. 80.
73. **Mello, F. C., et al.** *Hepatitis B virus genotypes circulating in Brazil: molecular characterization of genotype F isolates.* s.l. : BMC Microbiol, 2007. Vol. 23.
74. **Alvarado Mora, M. V., et al.** *Molecular characterization of the Hepatitis B virus genotypes in Colombia: a Bayesian inference on the genotype F.* s.l. : Infect Genet Evol, 2011. Vol. 11.
75. **Di Lello, F. A., et al.** *Diversity of hepatitis B and C viruses in Chile.* s.l. : J Med Virol, 2009. Vol. 81.
76. **Kurbanov, F., Tanaka, Y. e Mizokami, M.** *Geographical and genetic diversity of the human hepatitis B virus.* s.l. : Hepatol Res, 2010. Vol. 40.
77. **Panduro, A., et al.** *Distribution of HBV genotypes F and H in Mexico and Central America.* s.l. : Antivir Ther, 2013. Vol. 18.
78. **Phung, T. B., et al.** *Genotype X/C recombinant (putative genotype I) of hepatitis B virus is rare in Hanoi, Vietnam—genotypes B4 and C1 predominate.* s.l. : J Med Virol, 2010. Vol. 82.
79. **Tatematsu, K., Tanaka, Y. e Kurbanov, F.** *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.* s.l. : J Virol, 2009. Vol. 83.
80. **Kramvis, A., Kew, M. e Francois, G.** *Hepatitis B virus genotypes.* s.l. : Vaccine, 2005. Vol. 31.
81. **Yousif, M. e Kramvis, A.** *Genotype D of hepatitis B virus and its subgenotypes: An update.* s.l. : Hepatol Res, 2013. Vol. 43.
82. **Schaefer, S.** *Hepatitis B virus taxonomy and hepatitis B virus genotype.* s.l. : W J Gastroenterol, 2007. Vol. 13.

83. **Bertolini, D. A., et al.** *Hepatitis B virus genotypes from European origin explains the high endemicity found in some areas from southern Brazil.* s.l. : Infect Genet Evol, 2012. Vol. 12.
84. **Sitnik, R., et al.** *Hepatitis B virus genotypes and precore and core mutants in Brazilian patients.* s.l. : J Clin Microbiol, 2004. Vol. 42.
85. **Barros, L. M., et al.** *High prevalence of hepatitis B virus subgenotypes A1 and D4 in Maranhão state, Northeast Brazil.* s.l. : Infect Genet Evol, 2014. Vol.24.
86. **Ribeiro, N. R., et al.** *Distribution of hepatitis B virus genotypes among patients with chronic infection.* s.l. : Liver Int, 2006. Vol. 26.
87. **Sitnik, R., et al.** *Hepatitis B virus genotype E detected in Brazil in an African patient who is a frequent traveler.* s.l. : Braz J Med Biol Res, 2007. Vol. 40.
88. **Gomes-Gouvêa, M. S., et al.** *HBV carrying drug-resistance mutations in chronically infected treatment-naïve patients.* s.l. : Antivir Ther, 2015. Vol. 20.
89. **Bottecchia, M., et al.** *Hepatitis B virus genotypes and resistance mutations in patients under long term lamivudine therapy: characterization of genotype G in Brazil.* s.l. : BMC Microbiol, 2008. Vol. 8.
90. **Mendes-Correa, M. C., et al.** *High frequency of lamivudine resistance mutations in Brazilian patients co-infected with HIV and hepatitis B.* s.l. : J Med Virol, 2010. Vol. 82.
91. **Silva, A. C., et al.** *Hepatitis B genotype G and high frequency of lamivudine-resistance mutations among human immunodeficiency virus/hepatitis B virus co-infected patients in Brazil.* s.l. : Mem Inst Oswaldo Cruz, 2010. Vol. 105.
92. **Lobato, C., et al.** *Intrafamilial prevalence of hepatitis B virus in Western Brazilian Amazon region: epidemiologic and biomolecular study.* s.l. : J Gastroenterol Hepatol, 2006. Vol. 21.
93. **Huy, T. T. e Abe, K.** *Molecular epidemiology of hepatitis B and C virus infections in Asia.* s.l. : Ped Int, 2004. Vol. 46.

94. **Zhang, X., et al.** *Analysis of hepatitis B virus genotypes and pre-coe region variability during interferon treatment of HBe antigen negative chronic hepatitis.* s.l. : B J Med Virol, 1996. Vol. 48.
95. **Wai, C. T., et al.** *HBV genotype B is associated with better response to interferon therapy in HBeAg (+) chronic hepatitis than genotype C.* s.l. : C Hepatol, 2002. Vol. 36.
96. **Livingston, S. E., et al.** *Hepatitis B virus genotypes in Alaska native people with hepatocellular carcinoma: Preponderance of genotype F.* 2007 : J Infect Dis, Vol. 195.
97. **Menzo, S., et al.** *Long-term follow-up of acute hepatitis b: new insights in its natural history and implications for antiviral treatment.* *Genes.* 2018, Vol. 293.
98. **Roman, S. e Panduro, A.** *HBV endemicity in Mexico is associated with HBV genotypes H and G.* *World J Gastroenterol*, 2013. Vol. 19.
99. **Varuzza, L.** *Introdução à análise de dados de sequenciadores de nova geração.* 2013.
100. **Kircher, M. e Kelso, J.** *High-throughput DNA sequencing- concepts and limitations.* s.l. : Bioessays, 2010. Vol. 32.
101. **Pinto, L. S. e Kremer, F. S.** *Plataformas de Sequenciamento de Nova Geração & pré-processamento de dados.* 2016.
102. **Illumina, Inc.** *Illumina Sequencing Technology.* 2010.
103. **Maddrey, W. C.** *Hepatitis B: an important public health issue.* s.l. : J Med Virol, 2000. Vol. 61.
104. **Lavanchy, D.** *Hepatitis B virus epidemiology, disease burden, treatment, and current and emerging prevention and control measures.* s.l. : J Viral Hepat., 2004. Vol. 11.
105. **Gunther, S., et al.** *A Novel Method for Efficient Amplification of Whole Hepatitis B Virus Genomes Permits Rapid Functional Analysis and Reveals Deletion Mutants in Immunosuppressed Patients.* s.l. : Journal of Virology, 1995. Vol. 69.

106. **Gomes, M. M. S.** . *Caracterização de genomas completos do vírus da hepatite B de diferentes genótipos isolados no Brasil*. São Paulo : Programa de Pós-Graduação em Ciências da Coordenadoria de Controle de Doenças da Secretaria de Estado da Saúde de São Paulo, 2005. Vol. Dissertação (mestrado).
107. **Wang, X., et al.** *Integrating nested PCR with high-throughput sequencing to characterize mutations of HBV genome in low viral load samples*. s.l. : Medicine, 2017.
108. **Hall, T. A.** *BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT*. s.l. : Nucleic Acids Symp. Ser., 1999. Vol. 41.
109. **Drummond, A. J., et al.** *Bayesian phylogenetics with BEAUti and the BEAST 1.7*. s.l. : Molecular Biology And Evolution, 2012. Vol. 29.
110. **Ekblom, R. e Wolf, J. B. W.** *A field guide to whole-genome sequencing, assembly and annotation*. n. 9, 2014, Evolutionary Application, Vol. v. 7, p. 1026.
111. **Hung, J. H., et al.** *Endoplasmic reticulum stress stimulates the expression of cyclooxygenase-2 through activation of NF- κ B and pp38 mitogen-activated protein kinase*. Journal of Biol Chemistry, 2004. Vol. 279.
112. **Wang, H. C., et al.** *Hepatitis B virus pre-S mutants, endoplasmic reticulum stress and hepatocarcinogenesis*. s.l. : Cancer Sci, 2006, Vol. 97.
113. **Hsieh, Y., et al.** *Pre-S mutant surface antigens in chronic hepatitis B virus infection induce oxidative stress and DNA damage*. s.l. : Carcinogenesis, 2004. Vol. 10.
114. **Tratwein, C., et al.** *Hepatitis B virus mutations in the pre-S genome before and after liver transplantation*. s.l. : Hepatology, 1996. Vol. 3.
115. **Tsubota, A.** *How do naturally occurring YMDD-motif mutants influence the clinical course of lamivudine-naïve patients with chronic hepatitis B virus infection?* s.l. : J Gastroenterol Hepatol, 2006, Vol. 12.
116. **Tipples, G. A., et al.** *Mutation in HBV RNA-dependent DNA polymerase confers resistance to lamivudine in vivo*. s.l. : Hepatology, 1996, Vol. 3.

117. **Ozasa, A., et al.** *Influence of genotypes and precore mutations on fulminant or chronic outcome of acute hepatitis B virus infection.* s.l. : Hepatology, 2006. Vol. 44.
118. **Omata, M., et al.** *Mutations in the precore region of hepatitis B virus DNA in patients with fulminant and severe hepatitis.* s.l. : N Engl J Med, 1991. Vol. 324.
119. **Takahashi, K., et al.** *Hepatitis B virus genomic sequence in the circulation of hepatocellular carcinoma patients: comparative analysis of 40 full-length isolates.* s.l. : Arch Virol, 1998. Vol. 143.
120. **Tanaka, Y., et al.** *Specific mutations in enhancer II/core promoter of hepatitis B virus subgenotypes C1/C2 increase the risk of hepatocellular carcinoma.* s.l. : J Hepatol, 2006. Vol. 45.
121. **Nwokediuko, S. C.** *Chronic Hepatitis B: Management Challenges in Resource-Poor Countries.* s.l. : Hepat Mon., 2011, Vol. 10.
122. **Paraná, R. e Almeida, D.** *HBV epidemiology in Latin America.* s.l. : J Clin Virol, 2005, Vol. 34.
123. **Roman, S., et al.** *Hepatitis B virus infection in Latin America: a genomic medicine approach.* 23, s.l. : World J Gastroenterol, 2014, Vol. 20.
124. **Campos, R. H., Mbayed, V. A. e Pineiro Y Leone, F. G.** *Molecular epidemiology of hepatitis B virus in Latin America.* s.l. : J Clin Virol, 2005, Vol. 34.
125. **Fay, O. H. e Latin American Regional Study Group.** *Hepatitis B in Latin America: epidemiological patterns and eradication strategy.* 1, s.l. : Vaccine, 1990, Vol. 8.
126. **Kramvis, A.** *Genotypes and genetic variability of hepatitis B virus.* s.l. : Intervirology, 2014, Vol. 57.
127. **Devesa, M. e Pujol, F. H.** *Hepatitis B virus genetic diversity in Latin America.* s.l. : Virus Res, 2007, Vol. 127.
128. **Pujol, F., et al.** *Hepatitis B virus American genotypes: Pathogenic variants?* 44, s.l. : Clinics and Research in Hepatol and Gastrol, 2020.

129. **Mojsiejcuk, L., et al.** *Long-term evolution of hepatitis B virus genotype F: Strong association between viral diversification and the prehistoric settlement of Central and South America.* 6, s.l. : J Viral Hepat, 2020, Vol. 27.
130. **Wolf, J. M., et al.** *Evolutionary history of hepatitis B virus genotype H.* s.l. : Journal of Medical Virology, 2020.
131. **Flichman, D., et al.** *Full-length genome characterization of hepatitis B virus genotype H strain isolated from serum samples collected from two chronically infected patients in Argentina.* J Clin Microbiol, 2009, Vol. 47.
132. **Ural, O., et al.** *First case of hepatitis B virus genotype H infection in Turkey.* 3, s.l. : Mikrobiyol Bul, 2013, Vol. 47.
133. **Yamada, N., et al.** *Acute hepatitis B of genotype H resulting in persistent infection.* 11, s.l. : World Journal of Gastroenterology, 2014, Vol. 20.
134. **Mojsiejczuk, L., et al.** *Long-term evolution of hepatitis B virus genotype F: Strong association between viral diversification and the prehistoric settlement of Central and South America.* s.l. : Journal of Viral Hepatitis, 2019.
135. **Paraskevis, D., et al.** *Dating the origin of hepatitis B virus reveals higher substitution rate and adaptation on the branch leading to F/H genotypes.* s.l. : Mol Phylogenet Evol, 2015, Vol. 93.
136. **Souza, B. F. C. D., et al.** *A novel hepatitis B virus species discovered in capuchin monkeys sheds new light on the evolution of primate hepadnaviruses.* 6, s.l. : J Hepatol, 2018, Vol. 68.
137. **Sasha, M., et al.** *Analysis of the complete genome of hepatitis B virus subgenotype C2 isolate NHB17965 from a HBV infected patient.* National Institute of Biotechnology, Bangladesh : s.n., 2018.
138. **Spitz, N., et al.** *Complete genome sequence of a hepatitis B virus isolate of genotype D2, subtype adrq+, from Brazil.* s.l. : Arch Virol, 2017.
139. **Fonseca, D. E. F., et al.** *Full genome and phylogenetic analysis of hepatitis B virus genotype F in Mexican isolates.* s.l. : Archives of Virology, 2018. Vol. 163.

140. **Wang, X., et al.** *A simple and efficient strategy for the de novo construction of greater-than-genome-length hepatitis B virus replicons.* Beijing : Journal of Virological Methods, 2014. Vol. 207.
141. **Sozzi, V., et al.** *In vitro studies identify a low replication phenotype for hepatitis B virus genotype H generally associated with occult HBV and less severe liver disease.* s.l. : Virology, 2018. Vol. 519.
142. **Chotiyaputta, W. e Lok, A. S.** *Hepatitis B virus variants.* s.l. : Nat Rev Gastroenterol Hepatol, 2009. Vol. 6.
143. **Cao, J., Luo, S. e Xiong, Y.** *The Variability of Amino Acid Sequences in Hepatitis B Virus.* 1, s.l. : Virol Sin., 2019, Vol. Vol. 34.
144. **Thakur, V., et al.** *Profile, spectrum and significance of HBV genotypes in chronic liver disease patients in the Indian subcontinent.* s.l. : J Gastroenterol Hepatol, 2002, Vol. 17.
145. **Escobedo-Melendez, G., et al.** *High prevalence of occult hepatitis B virus genotype H infection among children with clinical hepatitis in west Mexico.* 6, s.l. : Mem Inst Oswaldo Cruz, 2014, Vol. 109.
146. **Tacke, F., et al.** *Acute hepatitis B virus infection by genotype F despite successful vaccination in an immune-competent German patient.* 4, s.l. : J Clin Virol, 2007, Vol. 38.
147. **Limeres, M. J., et al.** *Impact of hepatitis B virus genotype F on in vitro diagnosis: detection efficiency of HBsAg from Amerindian subgenotypes F1b and F4.* 9, s.l. : Arch Virol, 2019, Vol. 164.
148. **Morozov, V. e Groudinin, M. P. M.** *Homologous recombination between different genotypes of hepatitis B virus.* 2, 2000 : Gene, Vol. 260.
149. **Lazarevic, I.** *Clinical implications of hepatitis B virus mutations: recent advances.* 24, 2014 : World J Gastroenterol, Vol. 20.
150. **Araujo, N. M., Teles, S. A. e Spitz, N.** *Comprehensive analysis of clinically significant hepatitis B virus mutations in relation of genotype, subgenotype and geographic region.* s.l. : Front Microbiol, 2020, Vol. 11.

151. **Sugauchi, F., et al.** *Hepatitis B virus genotypes and clinical manifestation among hepatitis B carriers in Thailand.* s.l. : J Gastroenterol and Hepatol, 2002. Vol. 17.
152. **Akuta, N. e Kumada, H.** *Influence of hepatitis B virus genotype on the response to antiviral therapies.* s.l. : J Antimicrobial Chemo, 2005. Vol. 55.
153. **Liu, C. J. e Kao, J. H.** *Global perspective on the natural history of chronic hepatitis B: Role of hepatitis B virus genotypes A to J.* s.l. : Semin. Liver Dis., 2013. Vol. 33.
154. **Nakajima, A., et al.** *Full-Length Sequence of Hepatitis B Virus Belonging to Genotype H Identified in Japanese Patient with Chronic Hepatitis.* s.l. : Jpn J. Infect Dis, 2005. Vol. 58.
155. **Armstrong, R. A.** *When to use the Bonferroni correction.* Birmingham : School of Life and Health Sciences, 2014. Vol. 34.
156. **Garrison, E. e Marth, G.** *Haplotype-based variant detection from short-read sequencing.* 2012.
157. **Cingolani, P., et al.** *A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3.* s.l. : Landes Bioscience, 2012. Vol. 92.
158. **Hung, J. H., et al.** *Endoplasmic reticulum stress stimulates the expression of cyclooxygenase-2 through activation of NF- κ B and pp38 mitogen-activated protein kinase.* Journal of Biol Chemistry, 2004. Vol. 279.
159. **Wang, H. C., et al.** *Hepatitis B virus pre-S mutants, endoplasmic reticulum stress and hepatocarcinogenesis.* s.l. : Cancer Sci, 2006, Vol. 97.
160. **Ceres, P., Stray, S. J. e Zlotnick, A.** *Hepatitis B virus capsid assembly is enhanced by naturally occurring mutation F97L.* 17, s.l. : Journal of Virology, 2004, Vol. 78.
161. **Yuan, T. T., Tai, P. C. e Shih, C.** *Subtype-independent immature secretion and subtype-dependent replication deficiency of a highly frequent, naturally occurring mutation of human hepatitis B virus core antigen.* J. Virol. 1999, Vol. 73.

162. **Wang, Y., Zeng, L. I. e Chen, W.** *HBV X gene point mutations are associated with the risk of hepatocellular carcinoma: A systematic review and meta-analysis.* 6, s.l. : Mol Clin Oncol, 2016, Vol. 4.
163. **Al-Qahtani, A. A., et al.** *Hepatitis B virus (HBV) X gene mutations and their association with liver disease progression in HBV-infected patients.* 62, s.l. : Oncotarget, 2017, Vol. 8.
164. **Kao, J. H. e Chen, D. S.** *Global control of hepatitis B virus infection.* s.l. : Lancet Infect Dis, 2002, Vol. 2.
165. **Tipples, G. A., et al.** *Mutation in HBV RNA-dependent DNA polymerase confers resistance to lamivudine in vivo.* s.l. : Hepatology, 1996, Vol. 3.
166. **Tsubota, A.** *How do naturally occurring YMDD-motif mutants influence the clinical course of lamivudine-naïve patients with chronic hepatitis B virus infection?* s.l. : J Gastroenterol Hepatol, 2006, Vol. 12.