

**Universidade Federal de Pernambuco
Centro de Ciências Biológicas
Programa de Pós-Graduação em Genética**

Aparecida Jayane Sampaio Miranda

**Caracterização genética e patrilinhagens do cromossomo Y no interior de
Pernambuco a partir de regiões Y-STRs**

Recife

2022

Aparecida Jayane Sampaio Miranda

**Caracterização genética e patrilinhagens do cromossomo Y no interior de
Pernambuco a partir de regiões Y-STRs**

Defesa de Mestrado apresentada ao Programa de Pós-Graduação em Genética da Universidade Federal de Pernambuco como parte dos requisitos exigidos para obtenção do título de Mestre em Genética.

Orientador: Valdir de Queiroz Balbino

Coorientadora: Anna Theresa de Souza Liberal

Recife

2022

Catálogo na Fonte:
Bibliotecária Elaine Cristina Barroso, CRB4/1728

Miranda, Aparecida Jayane Sampaio

Caracterização genética e patrilinhagens do cromossomo Y no interior de Pernambuco a partir de regiões Y-STRs / Aparecida Jayane Sampaio Miranda – 2022.

66 f. : il., fig., tab.

Orientador: Valdir de Queiroz Balbino

Coorientadora: Anna Theresa de Souza Liberal

Dissertação (mestrado) – Universidade Federal de Pernambuco. Centro de Biociências. Programa de Pós-graduação em Genética, Recife, 2022. Inclui referências e apêndices.

1. Genética de populações 2. Filogenia 3. Pernambuco I. Balbino, Valdir de Queiroz (orient). II. Liberal, Anna Theresa de Souza (coorient.) III. Título

576.58

CDD (22.ed.)

UFPE/CB – 2022-167

APARECIDA JAYANE SAMPAIO MIRANDA

**CARACTERIZAÇÃO GENÉTICA E PATRILINHAGENS DO CROMOSSOMO Y
NO INTERIOR DE PERNAMBUCO A PARTIR DE REGIÕES Y-STRS**

Dissertação apresentada ao Programa de Pós Graduação em Genética da Universidade Federal de Pernambuco, Centro Acadêmico de Biociências, como requisito para a obtenção do título de Mestre em Genética. Área de concentração: Genética Humana.

Aprovado em: 12/05/2022.

BANCA EXAMINADORA

Prof. Dr. Valdir de Queiroz Balbino (Orientadora)
Universidade Federal de Pernambuco

Prof^a. Dr^a. Paula Sandrin Garcia (Examinador Interno)
Universidade Federal de Pernambuco

Prof^a. Dr^a. Simone Silva do Santos Lopes (Examinador Externo)
Universidade Estadual da Paraíba

Prof^a. Dr^a. Rosane Silva (Examinador Externo)
Universidade Federal do Rio de Janeiro

Dedico este trabalho ao meu companheiro de jornada, Michael, e ao meu filho, Joaquim.

AGRADECIMENTOS

A todos que compõem o Programa de Pós-graduação em Genética da Universidade Federal de Pernambuco pelas oportunidades concedidas e aprendizado adquirido e a CAPES pelo financiamento da pesquisa e bolsa de estudos.

Ao meu orientador, Prof. Dr. Valdir Balbino, pela confiança, ensinamentos acadêmicos e momentos que culminaram em muito aprendizado interpessoal.

A minha coorientadora, Dra. Anna Theresa Liberal, pela contribuição na pesquisa e por fornecer doses extras de conhecimento nas reuniões do grupo e momentos no laboratório.

Aos que compõem o LABBE, pelos momentos oportunos de aprendizado, em especial a todos do grupo PROFORENSE, pela colaboração, partilha de conhecimentos e momentos descontraídos que aliviaram as tensões da caminhada. E, desse grupo, agradecimentos adicionais para Bárbara Natieli e Ellen Medeiros, amigas que a pós me proporcionou, que se fizeram presentes em minha vida sempre de forma tão solícita, compartilhando sentimentos.

A Keila Xavier, por ter vivido essa aventura comigo desde o começo e ter me ensinado tanto em um ano intenso de convivência diária que jamais esquecerei, um ano rico em boas lembranças.

A minha sogra, Socorro, por se mostrar solícita e sempre estar em torcida pelo meu sucesso.

A minha família, principalmente aos meus pais por sempre priorizassem o crescimento pessoal pelo incentivo aos estudos, por todo o apoio, dedicação e compreensão quanto as minhas ausências.

Ao meu amado esposo, Michael, que viveu essa jornada intensamente ao meu lado e foi imprescindível para que eu pudesse concluir com êxito, e ao meu querido filho Joaquim, que deixou a trajetória mais doce, um pouco mais desafiadora, e por isso mesmo ainda mais rica em aprendizados. Vocês dois dão razão aos meus dias e a minha busca constante em ser melhor.

Por fim, a todos que direta ou indiretamente contribuíram para a conclusão desse trabalho.

RESUMO

Para o Estado de Pernambuco existe uma sub-representação em estudos genético-populacionais com marcadores Y-STRs, bem como no banco de dados mundial YHRD, sabidamente referência para diversas análises. Esse fato, somado à importância antropológica em apresentar detalhes da história populacional local, justificam o presente trabalho, que objetivou caracterizar as linhagens paternas e fornecer dados de interesse forense local para uma amostra da população de Pernambuco, por meio de um conjunto de 23 marcadores microsatélites do cromossomo Y (Y-STRs). Para este fim, 172 amostras masculinas oriundas de três mesorregiões do interior do estado foram coletadas e genotipadas utilizando o sistema multiplex PowerPlex®Y23 (Promega Corporation). Os parâmetros forenses de frequências alélicas (FA), frequências haplotípicas (FH), diversidade gênica (DG), diversidade haplotípica (DH), capacidade de discriminação (CD) e probabilidade de coincidência haplotípica (PCH) foram calculados utilizando o programa Arlequin 3.5.2.2, bem como as comparações de distância genética pelo índice F_{st} , e a inferência dos haplogrupos foi realizada através da plataforma online Haplogroup predictor – HAPTEST. Os dados obtidos para os parâmetros analisados revelaram uma alta diversidade gênica para os *loci* presentes, alta capacidade de discriminar haplótipos entre indivíduos aleatórios na população e baixa probabilidade de coincidir um mesmo haplótipo para dois indivíduos não aparentados, a inferência de haplogrupos baseada em haplótipos dos Y-STRs corroborou com a estrutura de patrilinhagens indicada para os demais estados brasileiros, com predominância daquelas europeias, sendo condizente com a história local de ocupação humana, ao passo que as distâncias genéticas foram menores para populações europeias e americanas, e maiores para as africanas.

Palavras-chave: Parâmetros Forenses; Ancestralidade; Marcadores Uniparentais; YHRD.

ABSTRACT

For the state of Pernambuco, there is an underrepresentation in genetic-population studies with Y-STRs markers, as well as in the world database YHRD, known to be a reference for several analyses. This fact, added to the anthropological importance of presenting details of the local population history, justify the present work, which aimed to characterize the paternal lineages and provide data of local forensic interest for a sample of the population of Pernambuco, through a set of 23 microsatellite markers of the Y chromosome (Y-STRs). For this purpose, 172 male samples from three mesoregions of the interior of the state were collected and genotyped using the PowerPlex®Y23 multiplex system (Promega Corporation). The forensic parameters of allelic frequencies (FA), haplotypic frequencies (FH), gene diversity (DG), haplotypic diversity (DH), discrimination capacity (CD) and probability of haplotypic coincidence (PCH) were calculated using the Arlequin 3.5.2.2 program, as well as genetic distance comparisons by the Fst index, and the inference of haplogroups was performed through the online platform Haplogroup predictor – HAPTEST. The data obtained for the analyzed parameters reveal a high genetic diversity for the loci present, a high capacity to discriminate haplotypes between random individuals in the population and a low probability of matching the same haplotype for two unrelated individuals, the inference of haplogroups based on the haplotypes of the Y-STRs corroborated the patrilineage structure indicated for the other Brazilian states, with a predominance of European ones, being consistent with the local history of human occupation, while the genetic distances were smaller for European and American populations, and greater for African populations.

Key words: Forensic Parameters; Ancestry; Uniparental Markers; YHRD.

LISTA DE ILUSTRAÇÕES

Figura 1-	Posições relativas dos 23 loci Y-STR disponíveis no PowerPlex® Y23 System.....	17
Figura 2-	Filogenia simplificada do cromossomo Y humano com nomenclatura do haplogrupo basal (em negrito) e marcadores Y-SNP definindo indicados nos ramos.....	22
Figura 3-	Mapa do estado de Pernambuco destacando as três mesorregiões, pontos das coletas e que configuram o interior do estado.....	29
Gráfico 1-	Valores de diversidade gênica (DG) encontrados para os 23 marcadores Y-STRs nas amostras do Agreste.....	37
Gráfico 2-	Valores de diversidade gênica (DG) encontrados para os 23 marcadores Y-STRs nas amostras do Sertão.....	37
Gráfico 3-	Valores de diversidade gênica (DG) encontrados para os 23 marcadores Y-STRs nas amostras da mesorregião do São Francisco.....	38
Gráfico 4-	Comparação dos parâmetros forenses obtidos para os sistemas multiplex de 23 e 17 marcadores, PPY23 e Yfiler respectivamente, nas três mesorregiões de Pernambuco.....	39
Gráfico 5-	Representatividade dos haplogrupos presentes na população masculina do interior de Pernambuco.....	40
Figura 4-	Gráfico MDS da distância F_{st} evidenciando a proximidade genética entre as mesorregiões pernambucanas e demais populações brasileiras, especialmente europeias.....	42

LISTA DE TABELAS

Tabela 1-	Frequências alélicas, quantidade de alelos e diversidade gênica (DG) dos 23 marcadores Y-STR obtidos com o Powerplex®Y 23 nas amostras do Agreste.....	34
Tabela 2-	Frequências alélicas, quantidade de alelos e diversidade gênica (DG) dos 23 marcadores Y-STR obtidos com o Powerplex®Y 23 nas amostras do Sertão.....	35
Tabela 3-	Frequências alélicas, quantidade de alelos e diversidade gênica (DG) dos 23 marcadores Y-STR obtidos com o Powerplex®Y 23 nas amostras do São Francisco.....	36
Tabela 4-	Valores para os parâmetros forenses analisados nas três mesorregiões do interior de Pernambuco, com os dados obtidos para as análises de 23 e 17 marcadores, PPY23 e Yfiler respectivamente. DG: diversidade gênica média; Nhap: número de haplótipos obtidos; DH: diversidade haplotípica; CD: capacidade de discriminação e PCH: probabilidade de coincidência haplotípica.....	38
Tabela 5-	Relação dos haplogrupos presentes na população masculina do interior de Pernambuco. N indiv: nº absoluto de indivíduos; N% indiv: nº relativo de indivíduos; AG: Agreste; SE: Sertão e SF: São Francisco.....	40

LISTA DE ABREVIATURAS, SIGLAS E SÍMBOLOS

Item	Definição
AG	Agreste
AZFc	Fator de azoospermia C
CD	Capacidade de discriminação
CNVs	<i>Copy-Number Variants</i> (variantes de número de cópias)
DG	Diversidade gênica
DH	Diversidade haplotípica
FA	Frequência alélica
FH	Frequência haplotípica
Indels	inserções/deleções
LABBE-UFPE	Laboratório de Bioinformática e Biologia Evolutiva da Universidade Federal de Pernambuco
mtDNA	DNA mitocondrial
NRY	<i>Non-recombining Region of the Y</i> (região não-recombinante do cromossomo Y)
PAR	<i>Pseudoautosomal Regions</i> (regiões pseudo-autossômicas)
PCH	Probabilidade de coincidência haplotípica
PCR	<i>Polymerase Chain Reaction</i> (reação em cadeia da polimerase)
PPY23	PowerPlex Y23
SE	Sertão
SF	São Francisco
SNP	<i>Single Nucleotide Polymorphisms</i> (polimorfismos de nucleotídeo único)

STR	<i>Short Tandem Repeat</i> ; (repetições curtas em tandem)
TCLE	Termo de consentimento livre e esclarecido
YHRD	<i>Haplotype Reference Database</i> (Banco de dados de referência de haplótipos Y-STR)

SUMÁRIO

1. INTRODUÇÃO	14
2. REVISÃO DA LITERATURA.....	16
2.1 A variação genética humana individual e populacional.....	16
2.1.1 A variação individual como marcador para a identificação humana	17
2.1.2 A variação populacional na inferência da ancestralidade	18
2.2 O cromossomo Y humano	20
2.2.1 Aplicabilidade dos marcadores Y-STRs: Haplótipos e haplogrupos do cromossomo Y	22
2.3 A Ocupação humana no interior pernambucano e a diversidade da população brasileira.....	27
3 OBJETIVOS	31
3.1 Objetivo Geral	31
3.2 Objetivos específicos	31
4 MATERIAL E MÉTODOS.....	32
4.1 Local de estudo e amostragem	32
4.2 Coleta de dados e material biológico.....	32
4.3 Obtenção do DNA e reação de amplificação	33
4.4 Eletroforese capilar e genotipagem	33
4.5 Parâmetros forenses e análise estatística genético-populacional.....	34
5 RESULTADOS	35
6 DISCUSSÃO	45
7 CONCLUSÕES	50
REFERÊNCIAS.....	51
APÊNDICE A - HAPLÓTIPOS E FREQUÊNCIAS HAPLOTÍPICAS.....	59
APÊNDICE B – HAPLÓTIPOS DE CADA HAPLOGRUPO/MESORREGIÃO	62

1. INTRODUÇÃO

A variação genética entre indivíduos é de aproximadamente 0,1% de todo o genoma, o suficiente para gerar entre 4,1 e 5,0 milhões de *loci* informativos capazes de delinear metodologias investigativas pertinentes na busca por respostas em diversos estudos científicos. Toda essa informação genética fornece o alicerce para identificar um criminoso, uma vítima ou restos mortais desconhecidos, assim como absolver um inocente na prática forense; diagnosticar um paciente com uma doença rara ou entender os mecanismos da enfermidade em busca de um tratamento na prática médica; além de elucidar os detalhes da história populacional nas análises antropológicas e de ancestralidade, dentre diversas outras aplicações.

De toda essa variedade, apenas de 10 a 15% ocorre entre as grandes populações continentais. Apesar de razoavelmente homogênea, essa diferença é substancial e sua distribuição foi moldada pelos processos de longo prazo da história populacional, cuja investigação pode fornecer detalhes importantes sobre a ancestralidade.

As regiões conhecidas como microssatélites ou *Short Tandem Repeat* (STRs; repetições curtas em tandem) são sequências repetidas que geralmente ocorrem em regiões intergênicas nas quais elevados níveis de variabilidade são tolerados. Os microssatélites presentes no cromossomo Y humano são chamados de Y-STR, estando localizados na região não recombinante do cromossomo Y, uma região diferenciada que foge à recombinação homóloga e é transferida unilateralmente entre os indivíduos do sexo masculino. Dentre outras características marcantes, essas Y-STRs são eficientes para reconstruir a história da população humana, estabelecer relações genealógicas e determinar a linhagem masculina em aplicações forenses.

Diversas populações ao redor do mundo têm traços da sua história investigados utilizando as Y-STRs, que também são comuns na prática investigativa das polícias científicas. O banco de dados *Y-STR Haplotype Reference Database* (YHRD, Banco de dados de referência de haplótipos Y-STR) reúne esses dados em uma plataforma de referência, cuja aplicabilidade depende da representatividade de todas as partes do globo.

O Brasil é um país continental com uma população miscigenada resultante de diferentes processos de ocupação humana ao longo de sua história, com (principalmente) europeus, africanos e nativos americanos se dispersando e interagindo entre si de forma heterogênea em todo o vasto território brasileiro, tornando a demografia de cada região única, conforme demonstrado por diversos trabalhos. Apesar disso, na plataforma do YHRD, o Brasil consta atualmente com pouco mais de 1.500 haplótipos para a sua população (utilizando um conjunto de 23 Y-STRs), representativos de apenas seis estados brasileiros, não englobando o Estado de Pernambuco nesse conjunto.

Pernambuco foi uma das primeiras regiões do Brasil a serem colonizadas, contando com uma história de formação diversificada, tal qual o restante do continente brasileiro. No entanto, existem particularidades na formação populacional brasileira, e para o estado de Pernambuco, processos de povoamento diferentes ocorreram à medida que as terras se afastavam da capital. Assim, investigações em grupos geográficos específicos é de grande valia no somatório de informações concretas para elucidar detalhes da história local.

2. REVISÃO DA LITERATURA

2.1 A variação genética humana individual e populacional

Nos últimos anos vimos grandes esforços para catalogar a variação genética humana e, em compasso com o avanço tecnológico, cientistas buscam entender seus efeitos nas doenças, características complexas e evolução (Eichler *et al.*, 2007; Frazer *et al.*, 2009). Um dos mais importantes achados do Projeto Genoma Humano (McPherson *et al.*, 2001) foi a determinação da diversidade genética existente entre diferentes indivíduos, possibilitando, dentre outras aplicações, diferenciá-los com precisão e inferir suas origens individuais e populacionais, uma vez que a informação é transmitida através das gerações.

No sentido mais amplo, as variantes genéticas podem ser divididas em duas classes, segundo a composição de nucleotídeos: as variantes de nucleotídeo único e as variantes estruturais. A primeira classe foi inicialmente melhor explorada e abriga os polimorfismos de nucleotídeo único (SNPs, do inglês *Single Nucleotide Polymorphisms*) (Frazer *et al.*, 2009). Ao passo que a classe das variantes estruturais abrange uma mistura heterogênea de variantes que surgem por diferentes mecanismos mutacionais, necessitando de subclassificação adicional (Conrad e Hurles, 2007). Dentre as variantes estruturais multialélicas estão as matrizes de repetições em tandem de sequências duplicadas, a exemplo das regiões repetitivas mini- e microssatélites (Conrad e Hurles, 2007; Pang *et al.*, 2012).

Estudos revelaram que as variantes estruturais são onipresentes e comuns, tal qual os SNPs, sendo responsáveis por uma proporção significativa da variação fenotípica normal. Na prática, um genoma típico difere do genoma humano de referência em 4,1 milhões a 5,0 milhões de *loci* e apesar dos SNPs e *indels* (inserções/deleções) curtos predominarem em quantidade, as demais variantes estruturais ocupam uma maior proporção em tamanho na totalidade do genoma (Freeman *et al.*, 2006; Eichler *et al.*, 2007; Auton *et al.*, 2015).

É validado na literatura que a maior parte da variação genética total é atribuída às diferenças intrapopulacionais, e não interpopulacional. De aproximadamente 0,1% do DNA que varia entre os indivíduos, apenas 10-15%

corresponde à variação entre as grandes populações continentais que se agrupam em: 1) África, 2) Europa, Oriente Médio e Ásia do Sul e Central, 3) Ásia Oriental, 4) Oceania e 5) Américas. O restante refere-se às variações individuais intrínsecas dessas populações humanas (Rosenberg *et al.*, 2002; Excoffier, 2003; Jorde e Wooding, 2004; Serre e Pääbo, 2004; Tishkoff e Kidd, 2004; Witherspoon *et al.*, 2007; Jobling *et al.*, 2016).

2.1.1 A variação individual como marcador para a identificação humana

O uso do DNA para a identificação humana (tipagem genética) foi sugerida no início da década de 80 por Botstein *et al.* (1980), ao demonstrarem a existência de pequenas variações individuais no material genético. A metodologia utilizava a determinação de polimorfismos em fragmentos gerados por enzimas de restrição. Em comparação à análise de grupos sanguíneos precedente, a nova técnica demonstrava a capacidade de determinar maior variabilidade, entretanto não detinha de potencial para abordar a individualidade (Martin *et al.*, 2001).

O caminho para conseguir individualizar as pessoas pelo seu DNA teve início com a descoberta de regiões repetitivas e hipervariáveis, denominadas minissatélites por Alec Jeffrey, em 1985 (Gill *et al.*, 1985), permitindo estabelecer a identidade de uma pessoa por meio de um conjunto dessas pequenas variações presente em seu genoma. A identificação humana através desse tipo de marcador genético ficou conhecida como *DNA fingerprinting* (impressão digital de DNA) e fazia uso de uma técnica que utilizava sondas para hibridização. Porém problemas estatísticos e de padronização geravam incompatibilidade nos resultados obtidos, consequentemente houve um decaimento no uso dos minissatélites (Martin *et al.*, 2001; Schlötterer, 2004; Butler, 2010).

O advento da técnica de reação em cadeia da polimerase (PCR, do inglês *Polymerase Chain Reaction*) em 1983, com sua alta sensibilidade para detecção e multiplicação de regiões alvo de forma facilitada e temporalmente menos custosa, tornou viável a análise de amostras contendo pequenas quantidades de DNA, característica importante para os laboratórios forenses (Butler, 2007; Caracedo, 2017). Dentre os marcadores moleculares baseados na PCR, os SNP e os microssatélites se destacam atualmente na genotipagem de DNA.

De forma geral, os microssatélites, ou STRs, atendem à maioria dos casos solicitados de identificação humana devido às características que apresentam e que serão revisadas a seguir. Os SNPs assumem papel de destaque em alguns casos especiais, a exemplo da análise de amostras com elevados níveis de degradação e para individualização de gêmeos monozigóticos, sendo considerados marcadores informativos adicionais (Weber-Lehmann *et al.*, 2014; Pontes *et al.*, 2015).

As STRs são sequências repetidas com um núcleo central (tema de repetição) de dois a sete nucleotídeos, que geralmente ocorrem em regiões intergênicas nas quais elevados graus de variabilidade são tolerados (Willems *et al.*, 2016). Esses marcadores se popularizaram como marcadores forenses por excelência devido, principalmente, à sua abundância e elevada variabilidade no genoma humano, sendo suficientemente informativos para questões gerais que requeiram a genotipagem para a identificação e individualização de amostras biológicas (Budowle e Daal, 2008; Butler, 2010; Primorac e Schanfield, 2014).

A amplificação das STRs caracteriza-se ainda pela simplicidade e robustez no processo, a possibilidade de testar vários *loci* simultaneamente por sistemas multiplexes e pela geração de produtos com comprimento reduzido, geralmente inferiores a 350 pb (Budowle e Daal, 2008; Primorac e Schanfield, 2014). Adicionalmente, existe uma base de dados bem estabelecida e diversos *loci* STR foram testados e validados em populações de vários países, permitindo um alto poder comparativo (Butler, 2007).

Os sistemas STR multiplex possibilitam a definição de perfis genéticos, que podem ser individualizantes (a exemplo dos autossômicos), ou específicos para grupos de indivíduos (marcadores de linhagem) (Primorac e Schanfield, 2014). Dentre aqueles capazes de diferenciar agrupamentos de indivíduos, estão as STRs localizadas no cromossomo Y humano, denominadas Y-STRs.

2.1.2 A variação populacional na inferência da ancestralidade

A história e a demografia da espécie humana, com uma evolução recente e alta mobilidade, explicam a reduzida variação genética entre os grandes grupos populacionais anteriormente citada (Bertorelle e Barbujani, 2004). Esse padrão de variação foi moldado pelos processos de longo prazo da história populacional, ao

qual cada variante tem sua própria história evolutiva específica, com o estado alélico ancestral estabelecendo a direção da mudança (Kittles e Weiss, 2003; Conrad e Hurles, 2007). Em contrapartida, variantes raras mais provavelmente derivam de mutações recentes e podem, portanto, agrupar indivíduos segundo uma ancestralidade recente em comum (Bergström *et al.*, 2020).

Existem limites imprecisos e arbitrários entre as populações, que nunca são verdadeiramente “puras” em seu sentido genético (que está atrelado aos conceitos de isolamento reprodutivo e barreiras reprodutivas) por consequência da contínua migração e mistura das populações humanas ao longo da história (Jorde e Wooding, 2004; Liphardt, 2012). Assim, algumas variantes estão ligadas a determinados grupos populacionais e são tidos como bons marcadores por tornar as distinções possíveis, uma vez que a distribuição da variação pelo genoma entre as populações é assimétrica (Keita *et al.*, 2004; Jobling *et al.*, 2015).

A comparação de frequências alélicas nas populações permite avaliar o grau de diferenciação entre elas, bem como relacionar indivíduos ao seu continente de origem com alta precisão, com base no padrão de uma pequena parte da sua variação individual, desde que haja dados genéticos suficientes (Jorde e Wooding, 2004; Serre e Pääbo, 2004; Tishkoff e Kidd, 2004; Witherspoon *et al.*, 2007; Rotimi e Jorde, 2010). A precisão pode ser maior com a utilização de marcadores com frequências de alelos muito diferentes em populações de diferentes regiões ou variantes funcionais que podem estar sob seleção diferencial. Todavia essa mesma precisão diminui para regiões geográficas intermediárias e indivíduos de ascendência mista (Tishkoff e Kidd, 2004).

Conrad e Hurles (2007) previram corretamente que os SNPs e as STRs permaneceriam os principais marcadores para delimitar grupos e atribuir indivíduos a grupos específicos em virtude de seu número e simplicidade, dentre outras características específicas. Apesar disso, os *indels* também se mostraram eficientes e ganharam espaço à medida que os estudos se intensificaram (Zaumsegl *et al.*, 2013). É válido ressaltar que em média, as STRs aleatórias (sem especificidade a uma dada população) são aproximadamente quatro a doze vezes mais informativas para inferir ancestralidade do que SNPs aleatórios, preferivelmente usados em confirmações pós-triagem por STRs (Liu *et al.*, 2005).

O termo ancestralidade abriga múltiplos significados e é conveniente que seja bem esclarecido em um estudo. A ancestralidade biogeográfica ou global, por exemplo, refere-se à origem ancestral média de um indivíduo em todo seu genoma e pode ser resolvida utilizando marcadores genéticos autossômicos que fazem distinção entre grupos ancestrais, resultando em um mosaico de segmentos com representação proporcional de populações ancestrais (Shiver e Kittles, 2004; Royal *et al.*, 2010; Martin *et al.*, 2018).

Em contrapartida, a linhagem ancestral faz referência às linhas de descendência materna e paterna e representam, portanto, uma pequena fração de toda a ancestralidade que um indivíduo carrega. As linhagens são obtidas pelo uso de marcadores haploides, representados, na espécie humana, pelo DNA mitocondrial (mtDNA) e pela região não-recombinante do cromossomo Y (NRY; do inglês *Non-recombining Region of the Y*), herdados materna- e paternalmente, respectivamente. Esses marcadores fornecem um alto nível de informações de ancestralidade devido à falta de recombinação, altas taxas de mutação e distribuições geográficas mais restritas em comparação com as regiões autossômicas (Shiver e Kittles, 2004; Royal *et al.*, 2010; Calafell e Larmuseau, 2017).

Dentre os marcadores uniparentais acima citados, a região NRY do cromossomo Y tem se mostrado a mais popular por combinar genealogia genética e história familiar, devido as suas características particulares (Calafell e Larmuseau, 2017), conforme será mais detalhado a seguir.

2.2 O cromossomo Y humano

O cromossomo Y humano contém uma diversidade genética funcional notavelmente baixa dentre todos os cromossomos humanos. Com uma história evolutiva incomum, o Y teria evoluído de um cromossomo autossômico, em um processo de degeneração caracterizado pelo decaimento maciço da maioria de genes ativos presentes em seus ancestrais, conforme revelado por vários modelos evolutivos, porém preservando genes com funções específicas do sexo heterogamético (Lahn *et al.*, 2001; Bachtrög, 2013; Jobling e Tyler-Smith, 2017).

O cromossomo Y possui aproximadamente 60 Mb de tamanho, destes cerca de 23 Mb é formado por eucromatina e contém 43 genes de codificação de

proteínas, além de diversos pseudogenes (Kido e Lau, 2015; Jobling e Tyler-Smith, 2017). As regiões pseudo-autossômicas de homologia com o X (*PARs*; do inglês *Pseudoautosomal Regions*) são os únicos segmentos que se recombina e é compartilhado livremente entre os sexos. Além das *PARs*, a porção eucromática abriga ainda regiões de cópia única e regiões repetitivas, enquanto que a porção heterocromática consiste apenas de regiões altamente repetitivas e não abriga genes (Lahn *et al.*, 2001; Kayser e Ballantyne, 2014).

A anteriormente citada região NRY compreende aproximadamente 95% do cromossomo Y que não passa pela recombinação homóloga durante a meiose (Kayser e Ballantyne, 2014). Essa região é transmitida apenas para os descendentes masculinos e tende a ser estática devido à essa ausência de recombinação e, conseqüentemente, tem baixas taxas de variações, que são atribuídas à maior rotatividade das células germinativas masculinas em comparação com às femininas (Lahn *et al.*, 2001). Por conseguinte, espera-se que os homens de uma mesma parentalidade paterna compartilhem uma região NRY idêntica, preservando um registro de sua história ancestral, e quaisquer mudanças presentes entre gerações subsequentes ocorram devido às mutações (Jobling e Tyler-Smith, 2003).

Existem forças evolutivas únicas atuando sobre o cromossomo sexual masculino humano devido as suas particularidades, a exemplo da provável maior susceptibilidade à deriva genética e do reduzido tamanho populacional efetivo do Y em comparação aos demais cromossomos (sendo $\frac{1}{4}$ do tamanho populacional de qualquer autossômico e $\frac{1}{3}$ do cromossomo X), o que contribui para a menor diversidade de sequências nele observada. Adicionalmente, assume-se a reduzida ação da seleção natural do cromossomo Y em decorrência da ausência de recombinação, influenciando na fixação ou eliminação de mutações (Jobling e Tyler-Smith, 2003; Bachrog, 2013).

As regiões microssatélites presentes na região NRY, as Y-STRs, são marcadores moleculares comprovadamente eficientes para reconstruir a história evolutiva da população humana, estabelecer relações genealógicas e determinar a linhagem masculina em aplicações forenses (Kareem *et al.*, 2015; Jobling e Tyler-Smith, 2017). A figura 1 evidencia a localização das 23 Y-STRs presentes

no kit comercial PowerPlex® Y23 System (Promega, Fitchburg, Wisconsin, USA), de uso comum na prática forense.

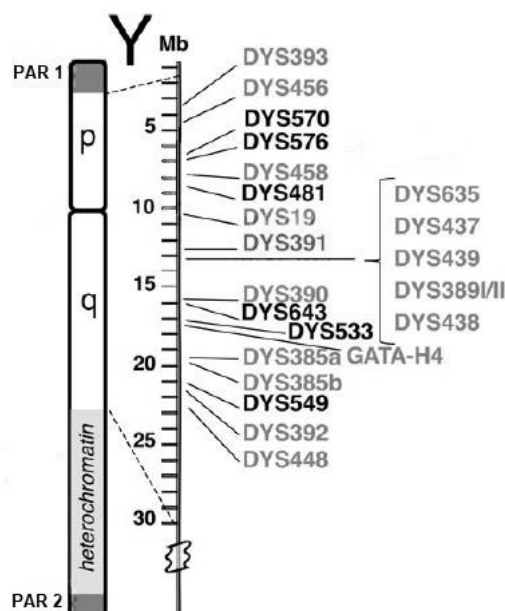


Figura 1 - Posições relativas dos 23 *loci* Y-STR disponíveis no PowerPlex® Y23 System. Os seis novos marcadores não presentes no antecessor Yfiler (17 *loci*) são mostrados em negrito. PAR1 e PAR2 são regiões pseudo-autossômicas nas extremidades do cromossomo Y que se recombinam com o cromossomo X. Fonte: Kareem *et al.* (2015) (modificado).

2.2.1 Aplicabilidade dos marcadores Y-STRs: Haplótipos e haplogrupos do cromossomo Y

Os marcadores Y-STRs são analisados em conjunto, e o resultado do teste gera uma combinação de números que representa os valores dos alelos dos marcadores DYS (nomenclatura que antecede os marcadores, sendo D para DNA, Y para cromossomo e S para segmento) em uma porção do DNA. Essas combinações de alelos Y-STRs ao longo do cromossomo formam um haplótipo, usado para identificar o haplogrupo de um indivíduo. Os haplogrupos são formados por um haplótipo que se correlaciona com outros através de polimorfismos binários de mutação lenta, geralmente os SNPs e consistem em informações genéticas comuns aos integrantes do grupo, remetendo ao seu ancestral comum mais recente (MRCA) (Jobling e Tyler-Smith, 2003, 2017; Mahal e Matsoukas, 2018; Navarro-López *et al.*, 2021).

A taxa de mutação mais rápida das Y-STRs em comparação aos Y-SNPs permite que a primeira seja eficiente na discriminação entre linhagens paternas não relacionadas, enquanto que o último seja usado para estabelecer a árvore filogenética e traçar a origem de cada haplogrupo (Navarro-López *et al.*, 2021).

A diversidade e conhecimento prévio da taxa de mutação dos *loci* contribui para que a análise de haplótipos possa alcançar a correta interpretação dos perfis genéticos. Para tal, existem Y-STRs de lenta, média e rápida mutação, atendendo a aplicações específicas (Ballantyne *et al.*, 2010; Calafell e Larmuseau, 2017).

Na área forense o uso das Y-STRs é bem estabelecido e normalmente aplicado em casos de crimes sexuais, onde a contribuição genética masculina é evidenciada mesmo em pequenas quantidades (Roewer, 2019). Além de identificar o sexo masculino em uma mistura, o método também permite identificar a linhagem masculina ou a origem biogeográfica paterna onde não há correspondência de perfis STR autossômicos em amostras de cadáveres não identificados ou doadores de vestígios desconhecidos (Kayser e Ballantyne, 2014; Navarro-López *et al.*, 2021). Os principais parâmetros estatísticos de interesse para validação forense dos *loci* estão descritos a seguir:

- A diversidade gênica (DG) retrata a probabilidade de dois indivíduos escolhidos ao acaso na população terem alelos divergentes e é um parâmetro equivalente a heterozigosidade esperada nos STRs autossômicos (Butler, 2005). Quanto maior a diversidade gênica, mais diverso, e consequentemente mais informativo, será o marcador;
- A diversidade haplotípica (DH) retrata a probabilidade de dois indivíduos ao acaso na população terem haplótipos divergentes (Excoffier, 2005);
- A probabilidade de coincidência haplotípica (PCH) representa a probabilidade de dois homens selecionados ao acaso possuírem haplótipos idênticos e é calculado por $1 - DH$ (Piglionica, 2013);
- Por fim, a capacidade de discriminação (CD) revela a eficiência do conjunto de marcadores em especificar linhagens de haplótipos de indivíduos não aparentados e é obtido pela razão entre número de haplótipos diferentes encontrados na amostragem e o número total de haplótipos (indivíduos) (Purps *et al.*, 2014).

Os *loci* Y-STRs contidos nos kits comerciais geralmente possuem moderada mutabilidade, sendo possível compará-los e selecionar os que

satisfazem a necessidade de resgatar registros históricos de parentalidade e genealogia genética. Isto representa uma vantagem técnica importante, uma vez que a recombinação existente entre os segmentos diploides das STR autossômicas pode reduzir a sensibilidade e a precisão desse método em casos onde a relação de parentesco está separada por algumas meioses, diminuindo a fração compartilhada do genoma autossômico (Calafell e Larmuseau, 2017).

Esses marcadores também demonstram forte ligação geográfica, uma vez que ainda se observa nas sociedades modernas uma alta taxa de migração feminina ocasionada pelo casamento patrilocal, com a mulher se deslocando para o território nativo do homem. Como consequência, mais homens vivem na região onde nasceram, ou próximos a esta, aprimorando a diferenciação genética sobre a distância geográfica do cromossomo Y (Jobling e Tyler-Smith, 2003; Ballantyne *et al.*, 2010; Calafell e Larmuseau, 2017). Dessa forma, as Y-STRs fornecem informações sobre eventos migratórios e ocupações territoriais formadoras das populações atuais nos estudos populacionais (Navarro-López *et al.*, 2021).

A região e o grupo étnico dos ancestrais paternos podem ser determinados com segurança, desde que existam dados de referência local representativos. Assim, o estabelecimento de bancos de dados populacionais acessíveis é um ponto crucial, sendo o maior deles o *Y Chromosome Haplotype Reference Database* (YHRD; Banco de Dados de Referência de Haplótipos do Cromossomo Y), que é um banco de dados de referência com perfis Y-STR anônimos de projetos de amostragem guiados por pesquisa acadêmica que se tornou uma plataforma online amplamente aceita para estudos populacionais revisados por pares (Roewer, 2019).

Outro fator importante para que as Y-STRs possam discriminar e agrupar amostras nas linhagens paternas é que estas sejam previamente estabelecidas pela filogenia dos SNPs (Navarro-López *et al.*, 2021). Para tal, o número crescente de Y-SNPs caracterizados nas duas últimas décadas contribuiu para estabelecer as relações filogenéticas e traçar a origem dos haplogrupos de Y através da reconstrução de eventos demográficos (Cruciani *et al.*, 2011).

Em 2002, o *Y Chromosome Consortium* (https://isogg.org/wiki/Y_Chromosome_Consortium) publicou uma única árvore de parcimônia, mostrando as relações entre 153 haplogrupos, com base em marcadores binários, sendo 18 clados principais nomeados de A-R, e

desenvolveu dois sistemas de nomenclatura padronizados. O primeiro é baseado em um sistema alfanumérico para definir os subclados hierárquicos dentro dos haplogrupos (ex. E1, E1a, E1a1) enquanto o segundo nomina os haplogrupos com base na mutação terminal que os definiu (ex. E-M75, correspondente a E2) (YCC, 2002).

Nos anos seguintes, atualizações e revisões foram publicadas, como a de Karafet *et al.* (2008), que incluíram novos haplogrupos à árvore, dentre eles, dois ramos principais (S e T). Van Oven *et al.* (2014) apresentaram uma filogenia dos 20 haplogrupos cuja árvore de referência visa a capacidade de discriminação global ideal com base em um conjunto fortemente reduzido que inclui apenas os Y-SNPs de maior resolução (Figura 2).

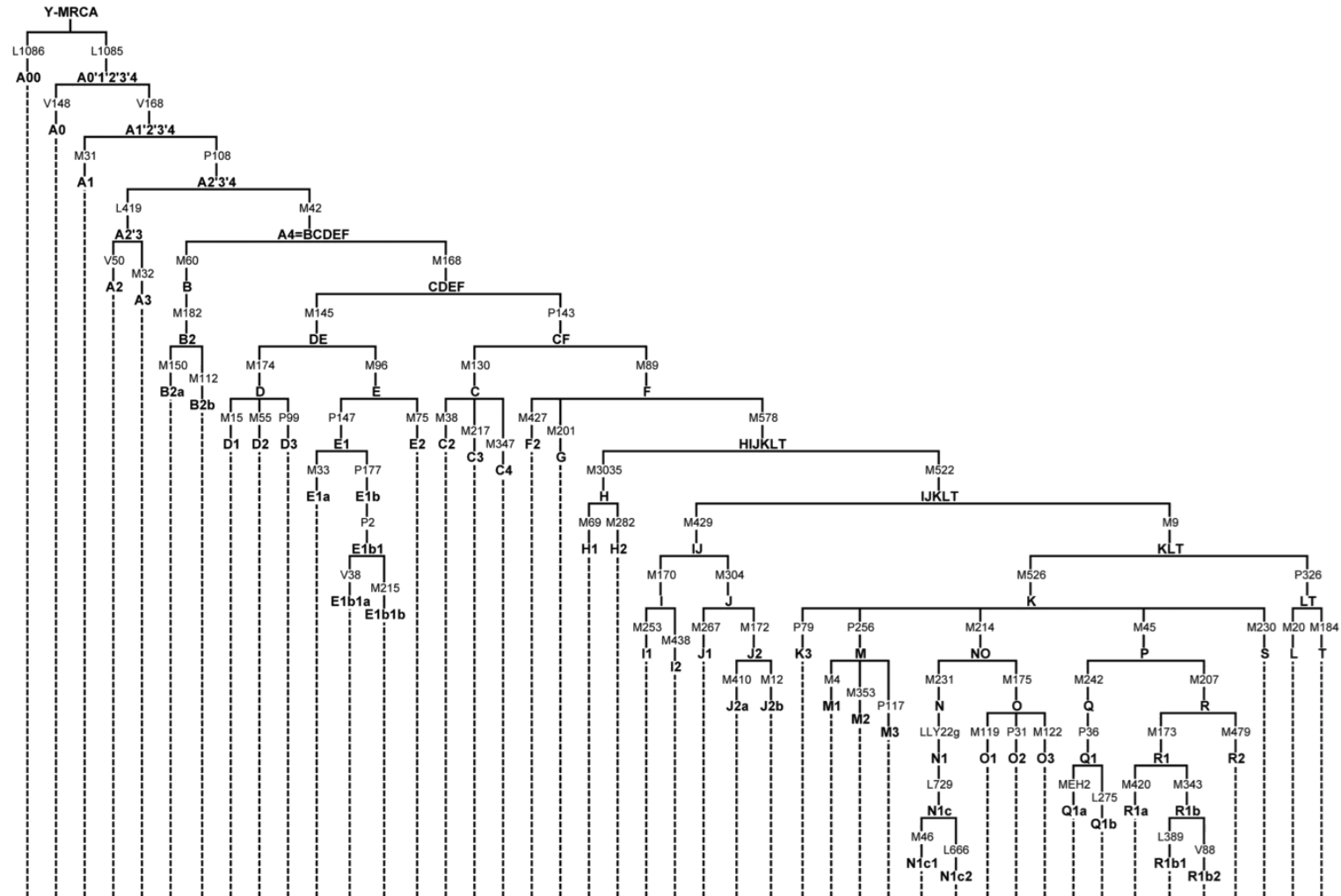


Figura 2 - Filogenia simplificada do cromossomo Y humano com nomenclatura do haplogrupo basal (em negrito) e marcadores Y-SNP definindo indicados nos ramos. O ancestral patrilinear comum mais recente de todos os humanos modernos (Y-MRCA) é indicado na parte superior. Os clados A e B de enraizamento mais profundo (à esquerda) são específicos para os africanos, apoiando assim a origem africana dos humanos modernos. Fonte: Van Oven *et al.* (2014).

2.3 A Ocupação humana no interior pernambucano e a diversidade da população brasileira

Os primeiros grupos humanos a ocupar o Nordeste brasileiro em épocas pleistocênicas pertenciam a grupos mongoloides, assim como os demais habitantes das Américas. Esses paleoíndios se instalaram em comunidades ribeirinhas às margens do Rio São Francisco ou em ocupações de caçadores-pescadores-coletores nas margens pernambucanas do rio, há pelo menos 8.000 anos (Martin, 1998).

Mais adiante no tempo, a miscigenação da população brasileira teve início com a chegada dos portugueses numa terra onde só havia nativos americanos. Nos anos que seguiram à chegada destes europeus, povos de outras nações também entraram no país. A princípio, e em maior escala devido ao comércio de escravos, os africanos, posteriormente italianos, espanhóis, alemães e imigrantes da Ásia e Oriente médio, em decorrência da abertura dos portos do país, a partir do final do século XIX, resultando na formação de uma das populações mais miscigenadas do mundo (IBGE 2007; 2008; Pena *et al.*, 2011; Souza *et al.*, 2019).

No Nordeste do Brasil colônia, os índios ali presentes foram praticamente extintos devido a resistência que ofereciam às obras missionárias, inicialmente pelos jesuítas, bem como a expulsão territorial para instalação das fazendas de gado no interior. Os índios foram então paulatinamente se afastando para as áreas mais afastadas do Vale do São Francisco (Martin, 1998). O interior de Pernambuco passou a abrigar uma sociedade formada pela camada populacional “excedente” do litoral da província de Pernambuco, funcionando como espaço de fuga de índios e negros que não se encaixavam na sociedade açucareira da época, mas também pelo interesse dos colonos em obter novas terras e títulos e da igreja em busca de expansão para a catequese (Jesus; Possamai, 2006).

A região se desenvolveu aos passos do avanço econômico pelo cultivo principalmente do algodão, além de outras culturas, com a influência do rio São Francisco e suas bacias hidrográficas locais. A pecuária também se estabelece nesse contexto e em certo período se interioriza mais em direção ao Agreste e posteriormente rumo ao Sertão (Valença Ferraz, 2004).

Essa sociedade seguiu baseada numa tradicional relação entre o patrimônio privado e instituição da família, onde o casamento fornecia os meios (filhos; herdeiros legítimos) de manter a riqueza (bens) adquirida ao longo do tempo. Aos homens, especificamente, cabia a manutenção do vínculo de proteção e laços familiares, o que lhes tornavam “reféns” do lugar e perpetuar a condição de proprietários de terra, com registros históricos do século XIX relatando tais costumes (Valença Ferraz, 2004). Mais do que os bens, os herdeiros do sexo masculino recebiam também o registro da linhagem paterna a qual pertencia através do seu DNA.

A diversidade genética do Brasil foi revisada por Manta *et al.* (2013) e Gontijo *et al.* (2018) para inferir padrões de variação e diferenças entre comunidades com histórias particulares utilizando diversos tipos de marcadores moleculares. Segundo as revisões, apesar da contribuição diversificada, há um consenso acerca da origem predominantemente tri-híbrida da população brasileira, sabidamente formada por europeus, africanos e nativos americanos.

Estes, por sua vez, dispersaram-se e interagiram entre si de forma heterogênea em todo o vasto território brasileiro, tornando a demografia de cada região única, com uma mistura genética de grau e extensão diversa (Pena *et al.*, 2011; Manta *et al.*, 2013; Gontijo *et al.*, 2018; Souza *et al.*, 2019). Isso gerou, por exemplo, uma maior influência de raiz ameríndia no Norte, forte presença africana pela escravidão no Nordeste e maior imigração europeia no Sul (Pena *et al.*, 2011; Palha *et al.*, 2012).

As análises com marcadores uniparentais revelaram origens principalmente europeias para as linhagens masculinas do cromossomo Y, e principalmente africanas e ameríndias para as linhagens femininas do mtDNA (Alves-Silva *et al.*, 2000; Carvalho-Silva *et al.*, 2001; Souza *et al.*, 2019). Segundo Gontijo *et al.* (2018), os marcadores de linhagem não descrevem a completa complexidade do *background* de cada indivíduo, mas trazem informações específicas cujos marcadores autossômicos não o fazem.

No que diz respeito à população do Nordeste, considera-se, com base na análise de *indels*, que essa população reflete a brasileira como um todo, com prevalência de ascendência europeia, seguida da africana e ameríndia (Manta *et al.*, 2013). A proporção de ascendência europeia é a menor registrada (Pena *et*

al., 2011), porém há uma maior proporção de ancestralidade africana dentre as demais regiões brasileiras (Manta *et al.*, 2013; Souza *et al.*, 2019).

Similarmente, a análise do mtDNA e de regiões NRY no Nordeste também revela predominância de linhagens maternas africanas e paternas europeias, como era esperado (Alves-Silva, 2000; Palha *et al.*, 2012; Resque *et al.*, 2016; Schaan *et al.*, 2020). Todavia há uma menor distância genética, inferida pelo índice F_{st} , das patrilinhagens nordestinas com àquelas das populações da África Subsaariana e maior distância com populações de Portugal e França em comparação aos demais estados brasileiros, aproximando-se mais em direção à Europa Central, com maiores contribuições vindo possivelmente da Itália (Resque *et al.*, 2016) ou da Península Ibérica na Europa Ocidental (Schaan *et al.*, 2020).

Apesar da literatura corroborante, existem limitações de amostragem que devem ser consideradas, como no trabalho de Manta *et al.* (2013), por exemplo, onde o Nordeste é representada apenas pelos estados de Pernambuco e Alagoas, cuja amostragem foi obtida em instituições de ensino e hospitais públicos e, por conseguinte, refletem uma porção limitada e possivelmente enviesada da população nordestina. Similarmente, Pena *et al.* (2011) representaram a região com outras duas fontes de amostragens, os estados do Ceará e Bahia, cujas coletas de amostras biológicas ocorreram em uma universidade pública e um banco de sangue, nesta ordem.

Na investigação das patrilinhagens, os trabalhos mais recentes e de maior amplitude amostral também continham tais limitações. Resque *et al.* (2016) analisaram 41 Y-SNPs fazendo uso da amostragem de Palha *et al.* (2012), que representaram o Nordeste apenas pelos estados do Maranhão e Ceará. No estudo de Palha *et al.* (2012), também foram incluídas mais 30 amostras de Pernambuco. Schaan *et al.* (2020) analisaram 45 Y-SNPs em 280 nordestinos, dentre estes, apenas nove amostras eram procedentes de indivíduos pernambucanos.

O trabalho de Jannuzzi *et al.* (2020) utilizou 141 amostras pernambucanas para análise com o kit comercial de 26 marcadores, sendo essa a investigação com melhor representatividade de Pernambuco encontrado até o momento, bem como com maior número de marcadores caracterizados. No entanto, as amostras foram registradas no YHRD para Recife (número de acesso YHRD: YA004379), e pode não representar o interior do estado. Adicionalmente, a inferência de

haplogrupos utilizando marcadores Y-SNPs foi realizada para uma fração da amostra total e a região Nordeste foi representada apenas pelas amostras do Maranhão.

Diante da necessidade de ampliar as possibilidades de uma melhor representação desse estado em bancos de dados genéticos para fins diversos na pesquisa, bem como investigar informações adicionais sobre a composição genética dessa região, uma das primeiras a serem colonizadas no país, o conjunto de 23 marcadores Y-STRs (PowerPlex Y23) provaram uma maior eficiência para as amostras brasileiras em comparação aos demais kits comerciais disponíveis com números menores de *loci* Y-STRs (Palha *et al.*, 2012). No trabalho de Palha *et al.* (2012) a capacidade de discriminação do kit aumentava e o compartilhamento de haplótipos entre as amostras reduzia à medida que mais *loci* foram inseridos. De fato, Roewer (2019) considera que os kits com 17 ou menos regiões Y-STRs são de baixa resolução, ao passo que a partir de 23 Y-STRs é possível obter alta resolução nas análises.

Por fim, o banco de dados YHRD consta atualmente com pouco mais de 1.500 haplotipos para a população brasileira utilizando o conjunto de 23 Y-STRs, amostrando apenas seis estados brasileiros. O estado de Pernambuco não está nesse conjunto (YHRD, 2021), o que ressalva a necessidade de análises com conjuntos de marcadores de alta resolução nesse estado.

3 OBJETIVOS

3.1 Objetivo Geral

Caracterizar as linhagens paternas de uma amostra da população de Pernambuco oriundas de três mesorregiões do estado, por meio de um conjunto de 23 marcadores microssatélites do cromossomo Y (Y-STRs).

3.2 Objetivos específicos

1. Identificar os diferentes haplótipos do cromossomo Y e calcular os parâmetros estatísticos forenses referentes aos *loci* Y-STRs;
2. Comparar a eficiência do kit comercial de 23 marcadores (PPY23) com o seu antecessor de 17 (YFiler);
3. Inferir, a partir dos haplótipos gerados, os haplogrupos e, consequentemente, determinar as principais etnias contribuintes nas populações dessas mesorregiões pernambucanas;
4. Avaliar a distância genética da população do interior de Pernambuco em relação a outras brasileiras, bem como as populações relacionadas ao contexto histórico de ocupação local;
5. Alimentar o banco de dados da população pernambucana com os marcadores genéticos supracitados e depositar os haplótipos para o banco de dados do cromossomo Y.

4 MATERIAL E MÉTODOS

4.1 Local de estudo e amostragem

A pesquisa foi desenvolvida no Laboratório de Bioinformática e Biologia Evolutiva da Universidade Federal de Pernambuco (LABBE-UFPE), cuja realização foi previamente aprovada pelo Comitê de Ética em Pesquisa da Universidade Federal de Pernambuco (protocolo número 2.625.726; CAAE 84801518.1.0000.5208). A amostragem total foi composta por 172 indivíduos do sexo masculinos, o que equivale a aproximadamente 0,005% da população total de três mesorregiões de Pernambuco: São Francisco (SF), Sertão (SE) e Agreste (AG), e representa aproximadamente 0,01% dos homens, consistindo de indivíduos saudáveis, não aparentados, naturais e residentes das unidades amostradas.

As coletas foram realizadas em cidades variadas e proporcionalmente mais representativas, por serem consideradas pontos estratégicos para obtenção de uma amostragem mais distribuída para as mesorregiões (SF= 32; SE= 56 e AG= 84) (Figura 3).



Figura 3 - Mapa do estado de Pernambuco destacando as três mesorregiões pontos das coletas e que configuram o interior do estado.

4.2 Coleta de dados e material biológico

Todas as amostras foram coletadas seguindo as recomendações éticas. Os indivíduos foram abordados em locais públicos e convidados a participar da pesquisa. Após assinatura do termo de consentimento livre e esclarecido (TCLE),

o material biológico foi colhido através da realização de uma punção digital, utilizando-se uma lanceta individual e estéril, e as gotas de sangue capilar (aproximadamente 100 µL) foram depositadas em um cartão absorvente (FTA, Whatman Bioscience), cuja composição contém substâncias químicas capazes de protegerem as moléculas de DNA da degradação de nucleases, além de preservar o papel da ação bacteriana (Butler, 2005).

4.3 Obtenção do DNA e reação de amplificação

O DNA genômico foi obtido dos leucócitos do sangue periférico de forma direta, sem extração de DNA. Discos de papel medindo 1,2 mm são obtidos dos cartões FTA, através de cortes realizados com o auxílio de um picotador (Harris Micro Punch - Whatman Bioscience) e, em seguida, depositados nos poços de placas de reação (MicroAmp® Optical 96-Well Reaction Plate (ThermoFisher Scientific) contendo os reagentes. Os 23 *loci* STR do cromossomo Y foram amplificados utilizando-se o sistema multiplex PowerPlex® Y23 (Promega Corporation), contendo os seguintes *loci*: DYS19, DYS385a/b, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS481, DYS533, DYS549, DYS570, DYS576, DYS635, DYS643 e Y-GATA-H4.

A amplificação obedeceu às recomendações descritas no protocolo do fabricante, no entanto o volume final foi otimizado para 12,5µL, sendo 8,75µl de água (Amplification Grade), 2,5µl de PowerPlex® Y23 5X Master Mix e 1,25µl de PowerPlex® Y23 10X Primer Pair Mix.

4.4 Eletroforese capilar e genotipagem

Os produtos amplificados foram separados e detectados no equipamento ABI 3500 Genetic Analyzer (Applied Biosystems, ThermoFisher Scientific), utilizando o polímero POP-4 (Applied Biosystems, Thermo Fischer Scientific) e capilar de 36 cm (Applied Biosystems, Thermo Fischer Scientific). As amostras foram preparadas em placa de 96 poços conforme recomendações do fabricante: 1,0µl do produto resultante da PCR, 9,5µl de formamida HI-DI (Applied Biosystems) e 0,5µl de WEN ILS500 (Promega Corporation)

A desnaturação foi realizada no termociclador à 95°C por três minutos e em seguida no freezer a -10°C por mais três minutos. Por fim, os perfis genotípicos dos Y-STRs foram obtidos através do software GeneMapper® ID-X v.1.4 (Applied Biosystems, ThermoFisher Scientific). A tipagem ou classificação dos alelos foi efetuada por comparação com o *Allelic Ladder* fornecido com o kit.

4.5 Parâmetros forenses e análise estatística genético-populacional

O programa Arlequin v. 3.5.2.2 (Excoffier; Lischer, 2010) foi utilizado para as análises das frequências alélicas (FA), frequências haplotípicas (FH); diversidade gênica (DG), diversidade haplotípica (DH), capacidade de discriminação (CD) e probabilidade de coincidência haplotípica (PCH).

A inferência dos haplogrupos foram realizados através da plataforma online Haplogroup predictor – HAPEST (www.hprg.com/hapest5). Os dados dos haplótipos foram organizados por ordem do marcador FTDNA e submetidos à versão do programa de 27 haplogrupos do software Haplogroup Predictor. O *locus* DYS549 não foi incluído na análise, uma vez que não faz parte dos marcadores FTDNA. Para os *loci* com duplicações, apenas um dos alelos integrou o haplótipo na inferência do respectivo haplogrupo.

Os resultados da avaliação da distância genética pelo índice Fst, foram gerados pelo *software* Arlequin v. 3.5.2.2. Os seguintes critérios, comumente adotados nas publicações, foram levados em conta nos dados analisados: 1) foi considerado, na análise do DYS389, o valor resultante da subtração entre o comprimento dos alelos de DYS389II e de DYS389I; 2) o DYS385 foi removido da análise; 3) os alelos nulos, intermediários e duplicados foram colocados como dados ausentes. O gráfico MDS da distância Fst foi gerado utilizando o *software* Vegan v. 2.5-6. O valor de stress do gráfico foi *Stress*: 0.004327538.

5 RESULTADOS

A análise de 23 marcadores microssatélites presentes no cromossomo Y com o kit comercial PowerPlex Y23 (PPY23) em 172 amostras provenientes das mesorregiões Agreste, Sertão e São Francisco, no interior do estado de Pernambuco, revelou a presença de 170 haplótipos distintos, sendo 168 únicos e com frequência haplotípica (F_h) igual a 0,0179 no Sertão, 0,0119 no Agreste e 0,0313 no São Francisco. Dois haplótipos apareceram duas vezes na amostragem, sendo compartilhados por dois indivíduos do Sertão ($F_h = 0,0357$) e dois do Agreste ($F_h = 0,0238$) (Apêndice A).

As frequências alélicas, a quantidade de alelos e a diversidade gênica para os *loci* em cada uma das três mesorregiões são mostrados nas tabelas 1, 2 e 3. No total 184 alelos diferentes foram encontrados na amostragem, sendo 39 alelos duplos para o *locus* DYS385, característica já descrita para esse marcador (Purps *et al.*, 2014). Outros alelos duplicados apareceram também nos marcadores DYS448 e DYS389 II, ambos no Sertão, ao passo que alelos intermediários, que compreende as repetições fracionárias, foram observados apenas no marcador DYS458 em todas as regiões.

Tabela 1 - Frequências alélicas, quantidade de alelos e diversidade gênica (DG) dos 23 marcadores Y-STR obtidos com o Powerplex®Y 23 nas amostras do Agreste.

Alelos	Marcadores																				YGATAH4	DYS385	
	DYS576	DYS389I	DYS448	DYS389II	DYS19	DYS391	DYS481	DYS549	DYS533	DYS438	DYS437	DYS570	DYS635	DYS390	DYS439	DYS392	DYS643	DYS393	DYS458	DYS456			
7																	0,0119				10, 14	0,0238	
8																	0,0238				10, 15	0,0238	
9						0,0476			0,0119	0,0833							0,1548				11, 11	0,0119	
10						0,5476		0,0238	0,0595	0,3452					0,0595		0,3452				0,0238	11, 12	0,0119
11		0,0119				0,3810		0,1667	0,3214	0,1190					0,3929	0,4286	0,1190				0,3452	11, 13	0,0357
12		0,2143				0,0119		0,4524	0,4762	0,4167					0,4286	0,0714	0,2381	0,1786			0,5833	11, 14	0,2262
13		0,6071			0,1310	0,0119		0,2500	0,1190	0,0238					0,0833	0,4048	0,0833	0,6310	0,0119	0,0714	0,0476	11, 15	0,0714
14	0,0119	0,1548			0,5119			0,0952	0,0119	0,0119	0,3929				0,0238	0,0714	0,0238	0,1667		0,0833		12, 12	0,0238
15	0,0595	0,0119			0,2976			0,0119			0,4881	0,0119			0,0119	0,0238		0,0238				12, 13	0,0119
16	0,2143				0,0476						0,1190	0,1310							0,1429	0,4643		12, 13	0,0119
16.2																			0,2262	0,2976		12, 14	0,0238
17	0,2619				0,0119							0,3452	0,0119						0,0119			12, 15	0,0119
17.2																			0,3095	0,0714		12, 16	0,0119
18	0,2500		0,0595									0,2381							0,0238			12, 19	0,0119
18.2																			0,2024			13, 13	0,0357
19	0,1667		0,4524									0,1310							0,0357			13, 14	0,0595
20	0,0238		0,3214									0,0714	0,0833	0,0119					0,0238			13, 15	0,0595
21	0,0119		0,1310				0,1310					0,0357	0,2500	0,0952					0,0119	0,0119		13, 16	0,0238
22			0,0238				0,3690					0,0238	0,1071	0,1310								13, 17	0,0476
23			0,0119				0,1667					0,0119	0,4762	0,2738								13, 18	0,0238
24							0,0595						0,0714	0,3571								13, 19	0,0119
25							0,0595						0,1071									14, 14	0,0238
26							0,0833							0,1071								14, 15	0,0119
27				0,0119			0,0357							0,0238								14, 16	0,0119
28				0,0238			0,0357															14, 17	0,0119
29				0,1071			0,0714															15, 15	0,0238
30				0,5000			0,0238															15, 16	0,0119
31				0,2262																		15, 19	0,0119
32				0,1190																		15, 20	0,0119
32				0,0119																		16, 16	0,0357
																						16, 17	0,0119
																						17, 17	0,0119
																						17, 18	0,0357
																						17, 20	0,0119
																						19, 20	0,0119
Σ frequências	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
N indivíduos	84	84	84	84	84	84	84	84	84	84	84	84	84	84	84	84	84	84	84	84	84	84	
N alelos	8	5	6	7	5	5	9	6	6	6	3	9	6	7	6	5	8	4	10	6	4	34	
DG	0,8003	0,5680	0,6787	0,6804	0,6374	0,5591	0,8075	0,7037	0,6598	0,6936	0,6004	0,7920	0,6954	0,7682	0,6586	0,6495	0,7871	0,5482	0,7983	0,6867	0,5442	0,9334	

Tabela 2 - Frequências alélicas, quantidade de alelos e diversidade gênica (DG) dos 23 marcadores Y-STR obtidos com o Powerplex®Y 23 nas amostras do Sertão.

Alelos	Marcadores																				
	DYS576	DYS389I	DYS448	DYS389II	DYS19	DYS391	DYS481	DYS549	DYS533	DYS438	DYS437	DYS570	DYS635	DYS390	DYS439	DYS392	DYS643	DYS393	DYS458	DYS456	YGATAH4
8									0,0179												
9						0,0536			0,0179	0,1429							0,0714				
10						0,4643		0,0179	0,0536	0,3036					0,0893		0,6071				0,0714
11		0,0357				0,4464		0,0893	0,2143	0,1250					0,3750	0,3214	0,1250				0,4107
12		0,1607			0,0179	0,0357		0,3929	0,6071	0,3393					0,3750	0,0893	0,1786	0,2143			0,4643
13	0,0179	0,5893			0,1429			0,4643	0,0536	0,0893					0,1250	0,5357	0,0179	0,6250		0,0536	0,0536
14	0,0179	0,2143			0,5179			0,0357	0,0357		0,3750				0,0357	0,0536		0,1250		0,0536	
15	0,0357				0,2143						0,5357	0,0179						0,0357	0,1429	0,4821	
16	0,0714				0,0893						0,0714	0,0179							0,1429	0,2679	
16.2																		0,0179			
17	0,1964				0,0179						0,0179	0,4464							0,3750	0,1250	
17.2																			0,0179		
18	0,4286		0,0357									0,1964	0,0179						0,2321	0,0179	
19	0,1607		0,5179				0,0179					0,1786	0,0179						0,0536		
19, 20			0,0179																		
20	0,0714		0,2679				0,0357					0,0357	0,0893						0,0179		
21			0,1607				0,0357					0,0536	0,1607								
22							0,4107					0,0357	0,1429	0,0714							
23							0,1429					0,0179	0,3929	0,3750							
24							0,1071						0,1607	0,4464							
25							0,1250						0,0179	0,0893							
26							0,0357							0,0179							
27				0,0357			0,0357														
28				0,1250			0,0536														
28, 29				0,0179																	
29				0,3571																	
30				0,3571																	
31				0,1071																	
Σ frequências	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
N indivíduos	56	56	56	56	56	56	56	56	56	56	56	56	56	56	56	56	56	56	56	56	56
N alelos	8	4	5	6	6	4	10	5	7	5	4	9	8	5	5	4	5	4	8	6	4
DG	0,7532	0,5903	0,6442	0,7292	0,6688	0,5916	0,7896	0,6318	0,5883	0,7623	0,5773	0,7370	0,7786	0,6584	0,7065	0,6097	0,5890	0,5565	0,7747	0,6864	0,6188

Tabela 3 - Frequências alélicas, quantidade de alelos e diversidade gênica (DG) dos 23 marcadores Y-STR obtidos com o Powerplex®Y 23 nas amostras do São Francisco.

Alelos	Marcadores																				YGATAH4	DYS385	
	DYS576	DYS389I	DYS448	DYS389II	DYS19	DYS391	DYS481	DYS549	DYS533	DYS438	DYS437	DYS570	DYS635	DYS390	DYS439	DYS392	DYS643	DYS393	DYS458	DYS456			
8									0,0313								0,0313				11, 13	0,0313	
9						0,0938		0,0313		0,0625							0,0938				11, 14	0,1875	
10						0,4375		0,0313	0,0625	0,3750					0,0938		0,5000				0,0313	11, 15	0,0938
11						0,4688		0,0313	0,3438	0,1250					0,4063	0,4063	0,0313	0,0313		0,0313	0,4063	12, 12	0,0313
12			0,1563					0,5313	0,4375	0,4063					0,3125	0,0625	0,3125	0,1563			0,4688	12, 13	0,0313
13			0,5625		0,1875			0,3438	0,1250	0,0313					0,1563	0,4688	0,0313	0,5313		0,0938	0,0938	12, 14	0,0625
14			0,2813		0,3750			0,0313			0,4375				0,0313	0,0313		0,2188		0,1250		12, 15	0,0313
15	0,0938				0,3125						0,4375					0,0313		0,0625	0,2188	0,4375		12, 16	0,0313
16	0,1250				0,0938						0,1250	0,1563							0,2500	0,2500		13, 14	0,0625
17	0,1875				0,0313							0,4063							0,2813	0,0625		13, 15	0,0625
17.2																			0,0313			13, 17	0,0938
18	0,3750			0,1250								0,1875							0,1875			13, 18	0,0938
19	0,1875			0,3750								0,0625							0,0313			13, 19	0,0313
20	0,0313			0,3750								0,0938	0,0313									14, 14	0,0313
21				0,0625								0,0313	0,2500	0,0625								14, 15	0,0313
22							0,4063					0,0313	0,1563	0,0938								14, 16	0,0313
23				0,0625			0,2813					0,0313	0,4688	0,1875								15, 18	0,0313
24							0,1250						0,0938	0,5625								16, 17	0,0313
25							0,0625							0,0938									
27							0,0313																
28					0,0938		0,0625																
29					0,3750		0,0313																
30					0,2813																		
31					0,2188																		
32					0,0313																		
Σ frequências	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
N indivíduos	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	
N alelos	6	3	5	5	5	3	7	6	5	5	3	8	5	5	5	5	6	5	6	6	4	18	
DG	0,7883	0,5988	0,7177	0,7460	0,7399	0,5988	0,7540	0,6149	0,6915	0,6956	0,6210	0,7843	0,7056	0,6472	0,7258	0,6290	0,6613	0,6613	0,7984	0,7399	0,6250	0,9456	

Os valores de diversidade gênica (DG) excederam 0,5 para todos os 23 marcadores do PPY23, e foram em média 0,6932 no Agreste, 0,6687 no Sertão e 0,7041 no São Francisco. Como era esperado, o marcador DYS385 que amplifica simultaneamente dois alelos apresentou maior valor de DG para todas as três representações populacionais (SE = 0,8532, AG = 0,9334 e SF = 0,9456). O DYS481 foi o segundo marcador mais polimórfico no Sertão (0,7896) e no Agreste (0,8075) e o DYS458 na mesorregião do São Francisco (0,7984) (Gráficos 1, 2 e 3).

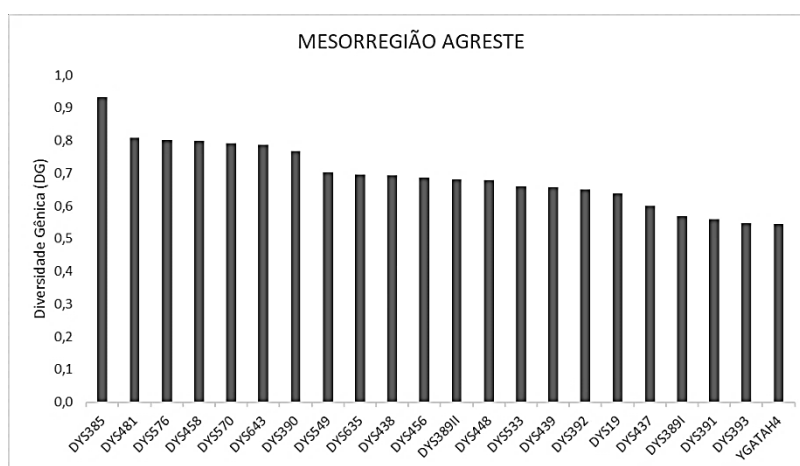


Gráfico 1 - Valores de diversidade gênica (DG) encontrados para os 23 marcadores Y-STRs nas amostras do Agreste.

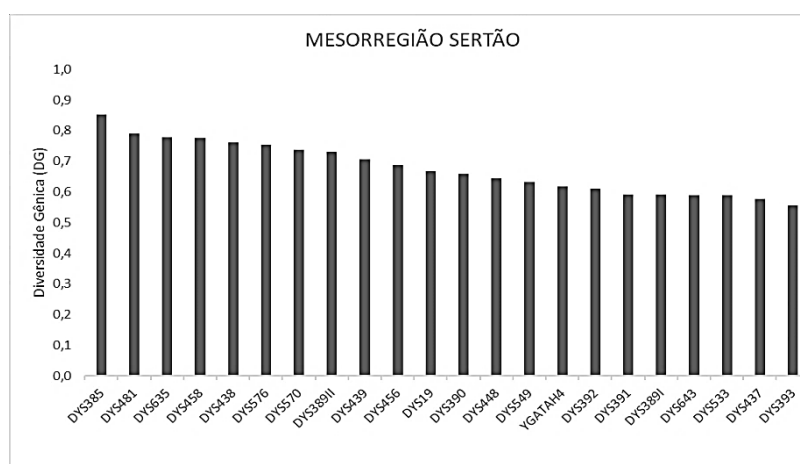


Gráfico 2 - Valores de diversidade gênica (DG) encontrados para os 23 marcadores Y-STRs nas amostras do Sertão.

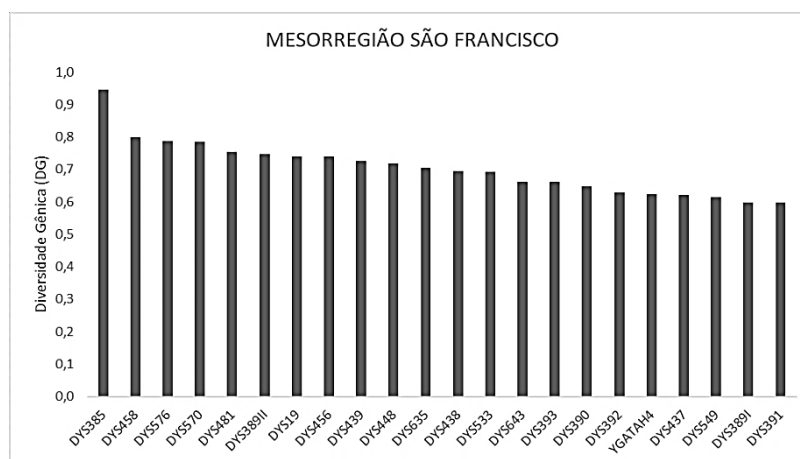


Gráfico 3 - Valores de diversidade gênica (DG) encontrados para os 23 marcadores Y-STRs nas amostras do São Francisco.

Na análise dos parâmetros forenses, a diversidade haplotípica (DH) mostrou valores iguais a 0,9997 no Agreste, 0,9993 no Sertão e 1,0 no São Francisco para o PPY23. Similarmente, a capacidade de discriminação (CD) obtida no presente trabalho foi de 0,9881 no Agreste, 0,9821 no Sertão e 1,0 no São Francisco (Tabela 4). Em contrapartida, a probabilidade de coincidência haplotípica (PCH) revelou valores de 0,0003 no Agreste, 0,0006 no Sertão e 0,0 no São Francisco.

Tabela 4 - Valores para os parâmetros forenses analisados nas três mesorregiões do interior de Pernambuco, com os dados obtidos para as análises de 23 e 17 marcadores, PPY23 e Yfiler respectivamente. DG: diversidade gênica média; Nhap: número de haplótipos obtidos; DH: diversidade haplotípica; CD: capacidade de discriminação e PCH: probabilidade de coincidência haplotípica.

	População		
	Agreste	Sertão	S. Francisco
N	84	56	32
DG (PPY23)	0,6932	0,6687	0,7041
Nhap	83	55	32
DH	0,9997	0,9994	1,0000
CD	0,9881	0,9821	1,0000
PCH	0,0003	0,0006	0,0000
DG (Yfiler)	0,6688	0,6754	0,6997
Nhap	81	55	32
DH	0,9991	0,9994	1
CD	0,9643	0,9821	1
PCH	0,0009	0,0006	0

Ao comparar os dados obtidos para os parâmetros forenses com o kit comercial YFiler, contendo 17 dos 23 marcadores analisados no PPY23, é possível observar uma variação sutil ou nenhuma variação nos valores de DG, DH e CD para os três grupos populacionais (Gráficos 4a, b, c e d). A diferença mais acentuada foi o compartilhamento de outros dois haplótipos entre os indivíduos do Agreste, totalizando três compartilhamentos para o conjunto de *loci* do YFiler e apenas um nessa mesorregião para o PPY23 (Tabela 4).

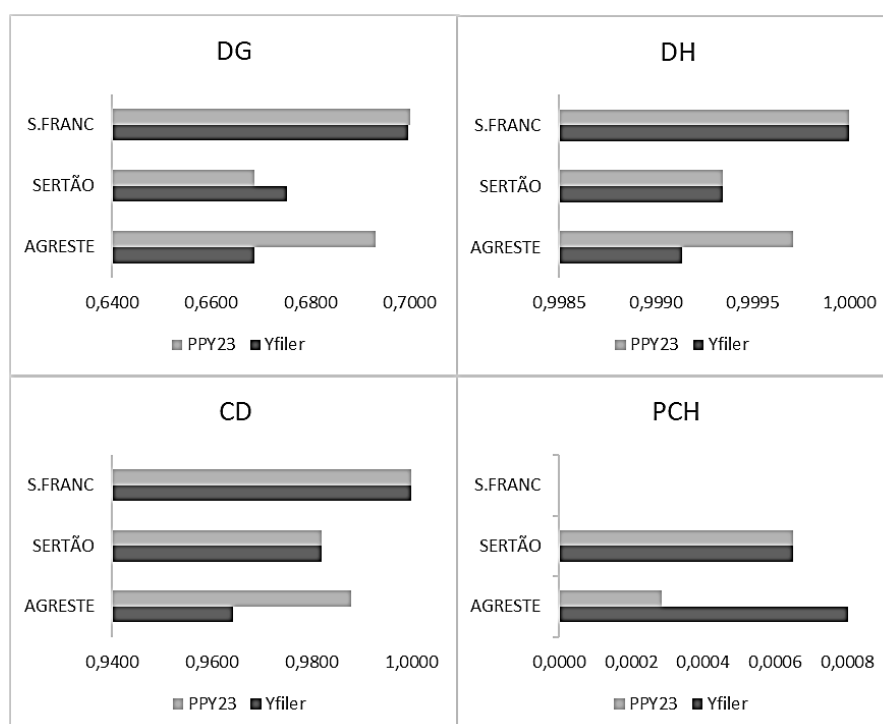


Gráfico 4 - Comparação dos parâmetros forenses obtidos para os sistemas multiplex de 23 e 17 marcadores, PPY23 e Yfiler respectivamente, nas três mesorregiões de PE. DG: diversidade gênica média; DH: diversidade haplotípica; CD: capacidade de discriminação e PCH: probabilidade de coincidência haplotípica.

Com auxílio da ferramenta online Haplogroup predictor – HAPEST foi realizada a inferência dos haplogrupos com base em dados Y-STR de cada amostra. Assim, os 170 haplótipos foram classificados em seu respectivo haplogrupo, sendo reunidos nos 13 haplogrupos a seguir: R1b, J1, E1b1b, E1b1a, Q, G2a, I1, I2b1, I2a1, R1a, J2b, I2a (xI2a1) e L (Tabela 5; Gráfico 5). O haplogrupo R1b foi o mais prevalente em todas as três mesorregiões, totalizando 46,51% da amostragem total, seguido do haplogrupo E1b1b e do J1, ambos

presentes em 11,63% do total de indivíduos amostrados nas três regiões. O haplogrupo L foi o mais raro, aparecendo apenas em um indivíduo do Sertão (0,58%).

Tabela 5 - Relação dos haplogrupos presentes na população masculina do interior de Pernambuco. N indiv: nº absoluto de indivíduos; N% indiv: nº relativo de indivíduos; AG: Agreste; SE: Sertão e SF: São Francisco.

Haplogrupo	N indiv	N% indiv	AG (%)	SE (%)	SF (%)
R1b	80	46,51	46,43	48,21	43,75
J1	20	11,63	11,90	14,29	6,25
E1b1b	20	11,63	10,71	10,71	15,63
E1b1a	9	5,23	10,71	0,00	0,00
Q	9	5,23	3,57	7,14	6,25
G2a	9	5,23	5,95	3,57	6,25
I1	7	4,07	4,76	3,57	3,13
I2b1	5	2,91	2,38	1,79	6,25
I2a1	5	2,91	1,19	3,57	6,25
R1a	3	1,74	1,19	1,79	3,13
J2b	2	1,16	1,19	0,00	3,13
I2a (xI2a1)	2	1,16	0,00	3,57	0,00
L	1	0,58	0,00	1,79	0,00
Total	172	100,0	100,0	100,0	100,00

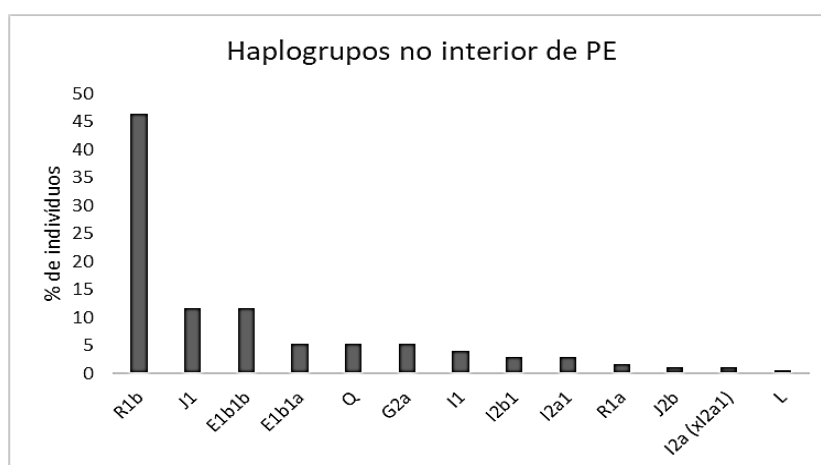


Gráfico 5 - Representatividade dos haplogrupos presentes na população masculina do interior de Pernambuco.

As comparações populacionais foram feitas entre as três mesorregiões do interior de Pernambuco com amostras brasileiras e estrangeiras com dados publicados que incluem as 23 Y-STRs no YHRD. As amostras brasileiras foram oriundas dos estados do Amazonas (duas amostragens, uma delas referente à população de nativos americanos de São Gabriel da Cachoeira), São Paulo e Rio de Janeiro. Além dessas, amostras de origem europeia, africana e de populações das américas foram comparadas (Tabela 6; Figura 4).

As maiores distâncias F_{st} para o interior de Pernambuco foram para as duas amostragens africanas (0,16348 – 0,25002), e as menores distâncias foram para as populações europeias (Portugal, Holanda e Espanha) e amostras de indivíduos de países americanos com descendência considerada europeia (Costa Rica, Paraguai, Argentina e Estados Unidos da América), além das demais populações brasileiras, exceto para a de nativos americanos em São Gabriel da Cachoeira.

Tabela 6 - Distância genética pelo índice F_{st} entre as mesorregiões do Agreste (AG), Sertão (SE) e São Francisco (SF), em Pernambuco, com populações brasileiras e de outros países. BR: Brasil (área sombreada).

Origem	Etnia	AG	SE	SF
BR: Amazonas		0,00158	0,00104	-0,0021
BR: S.G. da Cachoeira	Nativo americano	0,07473	0,06827	0,06339
BR: Rio de Janeiro		-0,0006	0,00803	0,00097
BR: São Paulo		-0,0013	0,00279	-0,0036
Estados Unidos	Americano Africano	0,06552	0,09124	0,07301
Estados Unidos	Americano Asiático	0,06319	0,07955	0,06271
Estados Unidos	Americano Europeu	0,01398	0,00736	0,0152
Estados Unidos	Americano Hispânico	0,00575	0,00457	-0,0007
Bolívia	Mestiço	0,09094	0,08358	0,08003
Bolívia	Nativo americano	0,12838	0,13498	0,12557
Argentina	Europeu	-5E-05	-0,0012	-0,0042
Peru		0,04771	0,03648	0,03899
Paraguai		0,00722	0,00089	0,00318
Panama		0,0104	0,01746	0,00688
Costa Rica		0,00372	0,00175	-0,0048
Espanha		0,01657	0,00667	0,01279
Portugal (Central)		-0,0011	0,00422	-0,0005
Portugal (Norte)		0,00333	-0,0004	-0,0024
Portugal (Sul)		0,00444	0,0036	0,00287
Holanda		0,01243	0,01212	0,01513
África do Sul	Africano	0,16348	0,20222	0,18408
Zimbábue		0,20454	0,25002	0,22902

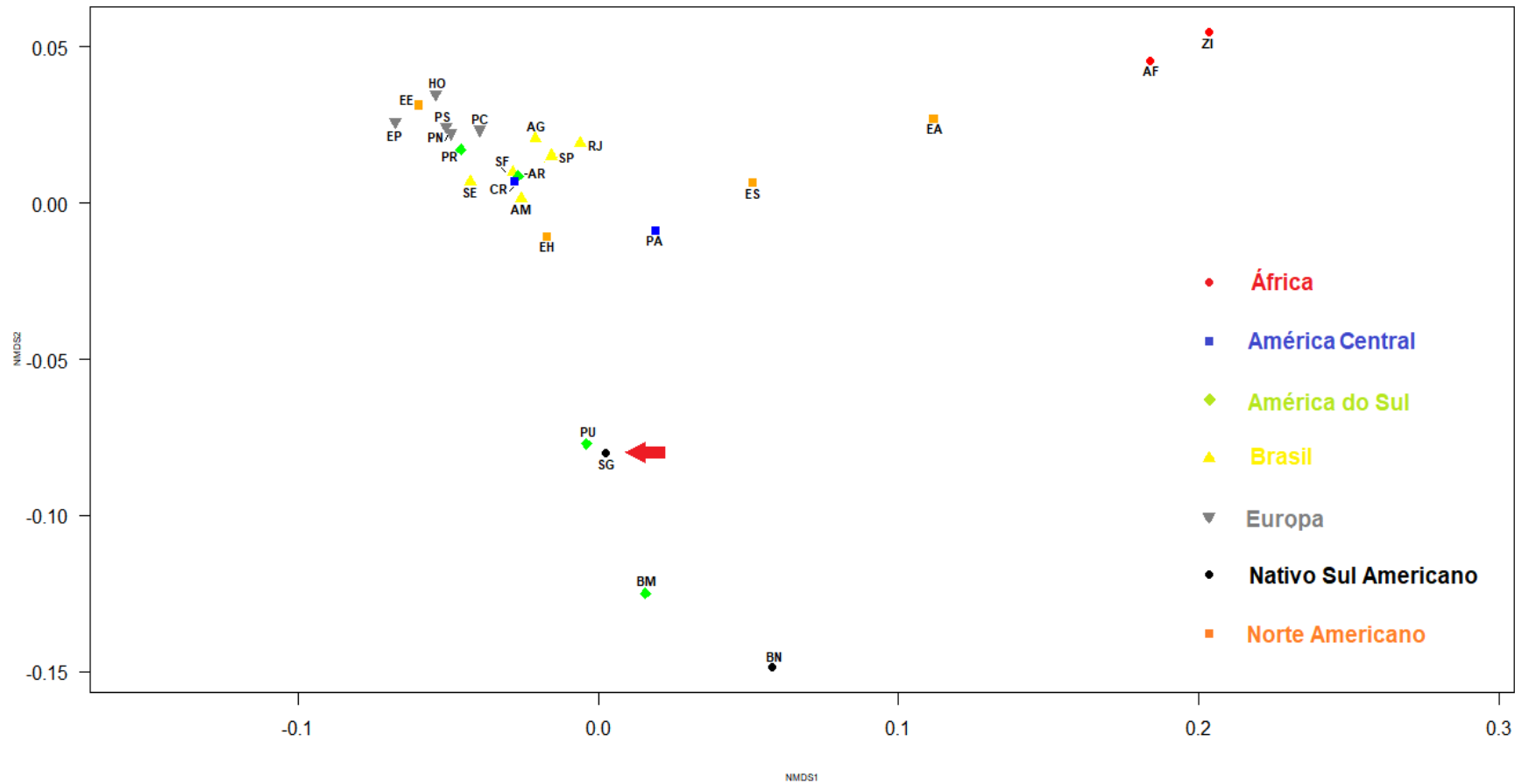


Figura 4 - Gráfico MDS da distância F_{st} evidenciando a proximidade genética entre as mesorregiões pernambucanas e demais populações brasileiras, especialmente europeias. Seta indicando a população de nativos americanos da Amazônia. AG = Agreste; SE = Sertão; SF = São Francisco; AM = Amazonas; SP = São Paulo; RJ = Rio de Janeiro; AR = Argentina; CR = Costa Rica; HO = Holanda; EE = EUA Americano europeu; EP = Espanha; PS = Portugal (Sul); PN = Portugal (Norte); PC = Portugal (Central); PR = Paraguai; EH = EUA Americano hispânico; PA = Panamá; ES = EUA Americano asiático; EA = EUA Americano africano; AF = África do Sul; ZI = Zimbabwe; PU = Peru; SG = São Gabriel da Cachoeira; BM = Bolívia (mestiço) e BN = Bolívia (Nativo americano).

6 DISCUSSÃO

O uso das Y-STRs pode beneficiar a resolução de questões específicas ligadas ao âmbito cível e criminal, entretanto, para que esses marcadores sejam utilizados pelas polícias científicas ou pelos laboratórios forenses, é necessário que estejam validados em determinada população local, haja vista a necessidade de conhecer o comportamento desses marcadores nas diferentes populações a fim de obter um resultado altamente confiável, que minimizem a possibilidade de uma identificação errônea.

Assim, todos os parâmetros analisados para os marcadores evidenciaram a eficiência do kit PowerPlex Y23 na fração da população pernambucana, revelando uma alta diversidade gênica para os *loci* presentes, alta capacidade de discriminar haplótipos entre indivíduos aleatórios na população e baixa probabilidade de coincidir um mesmo haplótipo para dois indivíduos não aparentados. Somando a escassez de dados utilizando o conjunto de 23 Y-STRs para Pernambuco no banco de dados YHRD (YHRD, 2021), a análise em 172 indivíduos do interior do estado mostra relevância na validação do kit para uso pela polícia científica brasileira.

Os resultados obtidos a partir da análise do conjunto de 23 marcadores moleculares presentes no PPY23 em amostras de três mesorregiões pernambucanas permitiram individualizar a linhagem de 97,67% dos homens amostrados, ao passo que o kit comercial YFiler, com 17 marcadores, possibilitou diferenciar 95,34% das linhagens desses indivíduos não aparentados.

Um maior número de haplótipos únicos obtidos pelo PPY23 *versus* YFiler também foi encontrado por Purps *et al.* (2014) e Lazim *et al.* (2020). Esse resultado seria consequência do maior número de marcadores do kit, bem como da escolha preferencial de marcadores com maior poder discriminatório (Purps *et al.*, 2014). Dentre os cinco *loci* com maior diversidade gênica, incluindo o DYS385, três são específicos do PPY23, a saber: DYS481, DYS570 e DYS576. Esses *loci* também estiveram entre os marcadores mais informativos nos trabalhos de Purps *et al.* (2014), Ambrosio (2019), Lazim *et al.* (2020) e Da Fré *et al.* (2015).

Os *loci* DYS570 e DYS576 são constituídos de repetições simples de tetranucleotídeos [(TTTC)₁₄₋₂₄ e (AAAG)₁₃₋₂₂ respectivamente] e considerados marcadores de rápida mutação, com taxas mutacionais na ordem de grandeza de 10^{-2} , estando entre os mais elevados dentre os disponíveis no YHRD. Essa classe de marcadores fornecem uma diferenciação masculina relativamente aumentada, capaz de levar a análise forense para além da diferenciação de linhagens masculinas, podendo distinguir membros de uma mesma linhagem paterna e alcançar a identificação do indivíduo masculino (YHRD, 2021; Ballantyne *et al.*, 2012; 2010).

Em contrapartida, o *locus* DYS481 é uma repetição simples de trinucleotídeos [(CTT)_N] que indicou alta informatividade nas análises de D'Amato *et al.* (2010) devido a presença de frequências alélicas diferenciais entre grupos populacionais, mais especificamente entre o inglês europeu e outros grupos, sugerindo potencial desse marcador para inferência de ancestralidade. Outros marcadores caracterizam-se por apresentarem variações alélicas com mais frequência e, sabidamente, algumas áreas do genoma humana são mais propensas a alterações no DNA. Os marcadores DYS448, DYS389 II e DYS458 apresentaram variações alélicas nesse trabalho.

A ocorrência de alelos adicionais pode ser consequência de uma duplicação do *locus* seguida por um evento mutacional que modifica o número de repetições (Jannuzzi *et al.*, 2020). Nesse contexto o *locus* DYS448, que apareceu com uma duplicação, está inserido na região do fator de azoospermia C (AZFc), no braço longo do cromossomo Y (Yq11.23). Essa região contém genes específicos para a espermatogênese e é uma área susceptível a rearranjos cromossômicos, podendo inclusive ocasionar efeitos fenotípicos (Turrina *et al.*, 2015; Ravasini *et al.*, 2021). Vários casos de duplicação ou exclusão foram relatados para o marcador DYS448 (Dente *et al.*, 2019; Aliferi *et al.*, 2018; Turrina *et al.*, 2015; Ambrosio, 2019; Della Rocca *et al.*, 2020; Jannuzzi *et al.*, 2020).

Dente *et al.*, (2019) mostraram que entre as variações alélicas analisadas, as duplicações no marcador DYS448 eram particularmente comuns em uma amostra de indivíduos portugueses de descendência africana. Ao passo que as duplicações encontradas por Aliferi *et al.* (2018), somadas a outras citadas por

eles, sugeriu um fenômeno específico presente em aproximadamente 1% (0,0095) dos indivíduos negros da população britânica. Concomitantemente, Della Rocca *et al.* (2020) encontraram 24 (de um total de 30) duplicações para o DYS448 em indivíduos africanos, todas estavam associadas ao haplogrupo E1a, sendo um padrão relativamente comum na África.

Outro *locus* que apresentou uma duplicação nesse trabalho foi o DYS389 II, fato esse também registrado na literatura. No trabalho de Valverde *et al.* (2012) um indivíduo residente, porém não-autóctone do país Basco, apresentou a mesma duplicação nesse marcador, cujos os alelos 28 e 29 coincidiram com os encontrados aqui em um indivíduo do Sertão. Dois indivíduos de São Paulo (Brasil) também apresentaram duplicações para o marcador em questão no trabalho de Soares-Vieira *et al.* (2008). Em um deles a duplicação ocorreu simultaneamente com o *locus* DYS389 I.

Adicionalmente, as variantes alélicas do tipo alelos intermediários em DYS458 foram observados aqui e em diversos outros trabalhos, tendo vários registros no banco de dados YHRD, sendo relativamente comum. Ferri *et al.* (2008) caracterizaram ao nível molecular os alelos fracionários 0.2 (ex. 17.2; 18.2; 20.2, etc.) para esse marcador, bem como sua distribuição mundial. Os resultados sugeriram uma origem comum para a variante 0.2, com origem provável no Oriente Médio ou África Oriental, sendo um polimorfismo estável com perspectivas para estudos populacionais. Todos os cromossomos contendo a variante por eles investigados se agruparam no haplogrupo J1. Da mesma forma, todos os sete haplótipos contendo as variações 0.2 (2/SE, 4/AG e 1/SF) aqui identificadas para o interior de Pernambuco estão reunidas nesse mesmo haplogrupo (J1).

Quanto à inferência de haplogrupos, estes são inferidos pelo *software* com base na associação de frequências de haplótipos presente em cada haplogrupo e disponíveis nas bases de dados (Jannuzzi *et al.*, 2020). Apesar disto, diversos trabalhos estão em consenso sobre a predominância do haplogrupo R1b na população brasileira, bem como a presença dos demais também encontrados para o interior de Pernambuco (este trabalho), com algumas confirmações por Y-

SNPs (Francez *et al.*, 2012; Resque *et al.*, 2016; Bottino *et al.*, 2019; Jannuzzi *et al.*, 2020; Schaan *et al.*, 2020).

O haplogrupo E é o mais mutacionalmente diverso, sendo também o terceiro mais comum na África, seguido dos dois mais basais e restritos a essa região, os haplogrupos A e B (Karafet *et al.*, 2008; Chiaroni *et al.*, 2009). Os haplogrupos G, I, J e L geralmente são geograficamente mais restritos (Europa, Oriente Médio e partes da Ásia Ocidental com várias extensões, especialmente para a Arábia e a Índia). O haplogrupo Q tem origem asiática, porém é bem estabelecido nas Américas e caracteriza a ancestralidade de nativos americanos (Chiaroni *et al.*, 2009). Enquanto o Haplogrupo R povoou a Ásia Central e Ocidental e uma grande parte da Europa, com as sublinhagens R1b atingindo as maiores frequências na Europa Ocidental, com frequências consideráveis também na Ásia e África, ao passo que o R1a prevalece na Europa Oriental e Ásia Central e do Sul (ISOGG, 2021; Myres *et al.*, 2011; Underhill *et al.*, 2015).

A prevalência de linhagens do cromossomo Y mais frequentes em países Europeus, com contribuições africanas e uma pequena representação de nativos americanos (haplogrupo Q; 5,23%), inferidas para o interior do estado de Pernambuco condiz com a história local revisada, com os primeiros povos (nativos americanos) sendo praticamente extintos no Brasil colônia pelos portugueses, que almejavam ocupar terras para agricultura e pecuária, além de uma porção menor de africanos que buscavam abrigo no interior do estado após fugir da condição escrava nas lavouras na capital pernambucana. Assim, a inferência de haplogrupos com base nos haplótipos de 23 Y-STRs revelaram um pouco da miscigenação já bem caracterizada para a população brasileira e de acordo com sua história, e agora também para amostras do interior de Pernambuco.

Para a análise de distância genética, Resque *et al.* (2016) mostram que as populações brasileiras estavam a uma menor distância F_{st} com os europeus e distâncias significativamente maiores com amostras africanas e nativo americanas. As amostras nordestinas, representadas pelos estados de Ceará e Maranhão, divergiram significativamente de Portugal e França e se aproximaram mais das populações do centro-europeu, com maiores contribuições vindo da

Itália. Para Schaan *et al.* (2020), as amostras nordestinas, representadas pelos estados do Ceará, Maranhão e Pernambuco (para esse último, nove amostras), estiveram mais próximas das populações da Península Ibérica e outras populações da Europa Ocidental (Espanha, Portugal, França, Itália, Alemanha e Holanda).

O trabalho de Jannuzzi *et al.* (2020) mostrou uma proximidade maior de todas as populações brasileiras com os europeus (com Pernambuco em uma posição mais distante dentre as demais), principalmente da Península Ibérica, em seguida os africanos e por fim os nativos americanos. Os autores constataram que as amostras de 11 populações de estados brasileiros mostraram alta homogeneidade, revelada por valores de distância *Fst* extremamente baixos.

As amostras do interior de Pernambuco analisadas mostraram baixa distância entre si e maior proximidade com as populações europeias e americanas, de colonização histórica europeia, em consenso com os achados mais recentes. Todavia, entre africanos e nativos americanos, as amostras brasileiras, incluindo pernambucanas, estavam mais próximas dos africanos nos trabalhos de Resque *et al.* (2016) e Jannuzzi *et al.* (2020), ao passo que para as amostras do interior de Pernambuco (esse trabalho) houve menor distância genética com populações nativas sul-americanas.

As proporções de sublinhagens do haplogrupo Q que caracteriza as origens nativo-americanas foram 1,92% (N = 416; Brasil) em Jannuzzi *et al.* (2020), 2,5% (N = 280; Nordeste) em Schaan *et al.* (2020) e 3,1% (N = 1217; Brasil) em Resque *et al.* (2016). Para o interior de Pernambuco essa proporção foi de 5,23%, o que resgata raízes dos primeiros indígenas como parte importante e ainda atuante na história local.

7 CONCLUSÕES

1. A análise dos 170 haplótipos distintos obtidos para o interior de Pernambuco e os valores de DG, DH, CD e PCH obtidos demonstrou que o sistema multiplex PowerPlex-Y23 é eficiente para a diferenciação de indivíduos masculinos não aparentados, naturais e filhos de pais naturais do interior de Pernambuco, condizentes com sua validação para uso forense no interior do estado.

2. Os dados obtidos revelaram um maior poder discriminatório para esse kit, em comparação com seu antecessor de 17 marcadores Y-STRs, tornando-o mais indicado para a prática investigativa.

3. A inferência de haplogrupos baseada em haplótipos dos Y-STRs corroborou com a estrutura de patrinhagens já indicada para os demais estados brasileiros, bem como com a história local de ocupação de terras interioranas, com uma maior representatividade de haplogrupos de predominância europeia, especialmente o R1b, seguido daqueles africanos e com uma menor contribuição ameríndia.

4. Há pouca distância genética entre o interior de Pernambuco e as demais regiões brasileiras, e maior proximidade destas com as populações europeias e americanas, de colonização histórica europeia. Em segundo plano, estão as populações nativas sul-americanas, e por fim as africanas.

5. Os haplótipos obtidos pelo PPY23 nas amostras do interior de Pernambuco foram depositados na plataforma YHRD, a fim de melhorar a representatividade do Estado naquela base de dados.

REFERÊNCIAS

- ALIFERI, A.; THOMSON, J.; MCDONALD, A. *et al.* UK and Irish Y-STR population data – A catalogue of variant alleles. **Forensic Science International: Genetics**, v. 34, p. e1-e6, 2018.
- ALVES-SILVA, J.; SANTOS, M. S.; GUIMARÃES, P. E. *et al.* The ancestry of Brazilian mtDNA lineages. **The American Journal of Human Genetics**, v. 67, n. 2, p. 444-461, 2000.
- AMBROSIO, I. B. **Avaliação de linhagens patrilíneas de 23 Y-STRs nas populações dos estados de São Paulo e Rio de Janeiro**. 2019. 178p. Tese (Doutorado em Biotecnologia) – Instituto de Química, Universidade Estadual Paulista Júlio de Mesquita Filho, São Paulo, 2019.
- AUTON, A.; 1000 GENOMES PROJECT CONSORTIUM; *et al.* A global reference for human genetic variation. **Nature**, v. 526, n. 7571, p. 68-74, 2015.
- BACHTROG, D. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. **Nature Reviews Genetics**, v. 14, n. 2, p. 113-124, 2013.
- BALLANTYNE, K. N.; GOEDBLOED, M.; FANG, R. *et al.* Mutability of Y-chromosomal microsatellites: rates, characteristics, molecular bases, and forensic implications. **The American Journal of Human Genetics**, v. 87, n. 3, p. 341-353, 2010.
- BALLANTYNE, K. N.; KEERL, V.; WOLLSTEIN, A. *et al.* A new future of forensic Y-chromosome analysis: rapidly mutating Y-STRs for differentiating male relatives and paternal lineages. **Forensic Science International: Genetics**, v. 6, n. 2, p. 208-218, 2012.
- BERGSTRÖM, A.; MCCARTHY, S. A.; HUI, R. A. *et al.* Insights into human genetic variation and population history from 929 diverse genomes. **Science**, v. 367, n. 6484, 2020.
- BERTORELLE, G.; BARBUJANI, G. Geographic structure of human genetic variation: medical and evolutionary implications. **Encyclopedia of Genetics, Genomics, Proteomics and Bioinformatics**, 2004.
- BOTSTEIN, D.; WHITE, R. L.; SKOLNICK, M. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. **American journal of human genetics**, v. 32, n. 3, p. 314, 1980.
- BOTTINO, C. G.; SILVA, R.; MOURA-NETO, R. S. Analysis of 124 SNP loci included in HID Ampliseq identity panel in a small population of Rio de Janeiro,

Brazil. **Forensic Science International: Genetics Supplement Series**, v. 7, n. 1, p. 243-244, 2019.

BUDOWLE, B.; VAN DAAL, A. Forensically relevant SNP classes. **Biotechniques**, v. 44, n. 5, p. 603-610, 2008.

BUTLER, J. M. **Forensic DNA typing: biology, technology, and genetics of STR markers**. 2^aed. Elsevier, USA, 2005.

BUTLER, J. M. Short tandem repeat typing technologies used in human identity testing. **Biotechniques**, v. 43, n. 4, p. Sii-Sv, 2007.

BUTLER, J. M. **Fundamentals of forensic DNA typing**. Academic press. Elsevier, EUA, 2010.

CALAFELL, F.; LARMUSEAU, M. H. The Y chromosome as the most popular marker in genetic genealogy benefits interdisciplinary research. **Human genetics**, v. 136, n. 5, p. 559-573, 2017.

CARACEDO, A. From Hemogenetics to Forensic Genomics. *In* Ferrara SD (ed). **P5 Medicine and Justice** (pp. 438-447). Springer, Cham, 2017.

CARVALHO-SILVA, D. R.; SANTOS, F. R.; ROCHA, J.; PENA, S. D. The phylogeography of Brazilian Y-chromosome lineages. **The American Journal of Human Genetics**, v. 68, n. 1, p. 281-286, 2001.

CASE, L. K.; TEUSCHER, C. Y genetic variation and phenotypic diversity in health and disease. **Biology of sex differences**, v. 6, n. 1, p. 1-9, 2015.

CHIARONI, J.; UNDERHILL, P. A.; CAVALLI-SFORZA, L. L. Y chromosome diversity, human expansion, drift, and cultural evolution. **Proceedings of the National Academy of Sciences**, v. 106, n. 48, p. 20174-20179, 2009.

CONRAD, D. F.; HURLES, M. E. The population genetics of structural variation. **Nature genetics**, v. 39, n. 7, p. S30-S36, 2007.

CRUCIANI, F.; TROMBETTA, B.; MASSAIA, A. *et al.* A revised root for the human Y chromosomal phylogenetic tree: the origin of patrilineal diversity in Africa. **The American Journal of Human Genetics**, v. 88, n. 6, p. 814-818, 2011.

DA FRÉ, N. N.; RODENBUSCH, R.; GASTALDO, A. Z. *et al.* Genetic data and de novo mutation rates in father-son pairs of 23 Y-STR loci in Southern Brazil population. **International journal of legal medicine**, v. 129, n. 6, p. 1221-1223, 2015.

D'AMATO, M. E.; EHRENREICH, L.; CLOETE, K. *et al.* Characterization of the highly discriminatory loci DYS449, DYS481, DYS518, DYS612, DYS626, DYS644 and DYS710. **Forensic science international: genetics**, v. 4, n. 2, p. 104-110, 2010.

DELLA ROCCA, C.; CANNONE, F.; D'ATANASIO, E. *et al.* Ethnic fragmentation and degree of urbanization strongly affect the discrimination power of Y-STR haplotypes in central Sahel. **Forensic Science International: Genetics**, v. 49, n. 102374, 2020.

DENTE, Â.; AMORIM, A.; COSTA, H. A. *et al.* Study of Y chromosome markers with forensic relevance in Lisbon immigrants from African countries—allelic variants study. **Forensic Science International: Genetics Supplement Series**, v. 7, n. 1, p. 906-907, 2019.

EICHLER, E. E.; NICKERSON, D. A.; ALTSHULER, D. *et al.* Completing the map of human genetic variation: A plan to identify and integrate normal structural variation into the human genome sequence. **Nature**, v. 447, n. 7141, p. 161- 165, 2007.

EMMEROVA, B.; EHLER, E.; COMAS, D. *et al.* Comparison of Y-chromosomal haplogroup predictors. **Forensic Science International: Genetics Supplement Series**, v. 6, p. e145-e147, 2017.

EXCOFFIER, L. Human diversity: our genes tell where we live. **Current Biology**, v. 13, n. 4, p. R134-R136, 2003.

EXCOFFIER, L. Arlequin 3.5 An integrated software package for population **Genetics Data analysis**, user manual, 2005. Disponível em: [<http://cmpg.unibe.ch/software/arlequin3>].

EXCOFFIER, L.; LISCHER, H. E. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. **Molecular ecology resources**, v. 10, n. 3, p. 564-567, 2010.

FERRI, G.; ROBINO, C.; ALU, M. *et al.* Molecular characterisation and population genetics of the DYS458. 2 allelic variant. **Forensic Science International: Genetics Supplement Series**, v. 1, n. 1, p. 203-205, 2008.

FRANCEZ, P. A. D. C.; RAMOS, L. P. V.; PALHA, T. D. J. B. F.; SANTOS, S. E. B. D. Haplotype diversity of 17 Y-STR loci in an admixed population from the Brazilian Amazon. **Genetics and molecular biology**, v. 35, p. 45-52, 2012.

FRAZER, K. A.; MURRAY, S. S.; SCHORK, N. J.; TOPOL, E. J. Human genetic variation and its contribution to complex traits. **Nature Reviews Genetics**, v. 10, n. 4, p. 241-251, 2009.

FREEMAN, J. L.; PERRY, G. H.; FEUK, L. *et al.* Copy number variation: new insights in genome diversity. **Genome research**, v. 16, n. 8, p. 949-961, 2006.

GONTIJO, C. C.; MENDES, F. M.; SANTOS, C. A. *et al.* Ancestry analysis in rural Brazilian populations of African descent. **Forensic Science International: Genetics**, v. 36, p. 160-166, 2018.

IBGE, INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Brasil 500 anos de povoamento**. Centro de Documentação e Disseminação de Informações, Rio de Janeiro. 232p, 2007.

IBGE, INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **O vento do Oriente: uma viagem através da imigração japonesa no Brasil**. Centro de Documentação e Disseminação de Informações, Rio de Janeiro. 96p, 2008.

ISOGG, INTERNATIONAL SOCIETY OF GENETIC GENEALOGY. 2021. Disponível em: [<https://isogg.org/tree/index.html>].

JANNUZZI, J.; RIBEIRO, J.; ALHO, C. *et al.* Male lineages in Brazilian populations and performance of haplogroup prediction tools. **Forensic Science International: Genetics**, v. 44, n. 102163, 2020.

JESUS, M. S.; POSSAMAI, P. C. O avanço da fronteira interna: a ocupação do sertão no século XVII. **Anais** Eletrônicos do VII Encontro Internacional da ANPHLAC, 2006. Disponível em: [http://antigo.anphlac.org/sites/default/files/mirian_silva.pdf].

JOBLING, M. A.; TYLER-SMITH, C. The human Y chromosome: an evolutionary marker comes of age. **Nature Reviews Genetics**, v. 4, n. 8, p. 598-612, 2003.

JOBLING, M. A.; RASTEIRO, R.; WETTON, J. H. In the blood: the myth and reality of genetic markers of identity. **Ethnic and Racial Studies**, v. 39, n. 2, p. 142-161, 2016.

JOBLING, M. A.; TYLER-SMITH, C. Human Y-chromosome variation in the genome-sequencing era. **Nature Reviews Genetics**, v. 18, n. 8, p. 485-497, 2017.

JORDE, L. B.; WOODING, S. P. Genetic variation, classification and 'race'. **Nature genetics**, v. 36, n. 11, p. S28-S33, 2004.

KARAFET, T. M.; MENDEZ, F. L.; MEILERMAN, M. B. *et al.* New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. **Genome research**, v. 18, n. 5, p. 830-838, 2008.

KAREEM, M. A.; HUSSEIN, A. O.; HAMEED, I. H. Y-chromosome short tandem repeat, typing technology, locus information and allele frequency in different population: A review. **African Journal of Biotechnology**, v. 14, n. 27, p. 2175-2178, 2015.

KAYSER, M.; BALLANTYNE, K. N. Y Chromosome in Forensic Science. In: Primorac D and Schanfield M (eds) **Forensic DNA Applications: An Interdisciplinary Perspective**. CRC Press, Boca Raton, pp. 105-128, 2014.

KEITA, S. O. Y.; KITTLES, R. A.; ROYAL, C. D. *et al.* Conceptualizing human variation. **Nature genetics**, v. 36, n. 11, p. S17-S20, 2004.

KIDO, T.; LAU, Y. F. C. Roles of the Y chromosome genes in human cancers. **Asian journal of andrology**, v. 17, n. 3, p. 373, 2015.

KITTLES, R. A.; WEISS, K. M. Race, ancestry, and genes: implications for defining disease risk. **Annual review of genomics and human genetics**, v. 4, n. 1, p. 33-67, 2003.

LAHN, B. T.; PEARSON, N. M.; JEGALIAN, K. The human Y chromosome, in the light of evolution. **Nature Reviews Genetics**, v. 2, n. 3, p. 207-216, 2001.

LAZIM, H.; ALMOHAMMED, E. K.; HADI, S.; SMITH, J. population genetic diversity in an iraqi population and gene flow across the Arabian Peninsula. **Scientific reports**, v. 10, n. 1, p. 1-11, 2020.

LIPPHARDT, V. Isolates and crosses in human population genetics; or, a contextualization of German race science. **Current Anthropology**, v. 53, n. S5, p. S69-S82, 1012.

LIU, N.; CHEN, L.; WANG, S. *et al.* Comparison of single-nucleotide polymorphisms and microsatellites in inference of population structure. **BMC Genetics**, BioMed Central, v. 6, n. 1, p. 1-5, 2005.

MAHAL, D. G.; MATSOUKAS, I. G. The geographic origins of ethnic groups in the Indian subcontinent: exploring ancient footprints with Y-DNA haplogroups. **Frontiers in genetics**, n. 9, 2018.

MALSOM, S.; FLANAGAN, N.; MCALISTER, C.; DIXON, L. The prevalence of mixed DNA profiles in fingernail samples taken from couples who co-habit using autosomal and Y-STRs. **Forensic Science International: Genetics**, 3(2): 57-62, 2009.

MANTA, F. S. N.; PEREIRA, R.; VIANNA, R. *et al.* Revisiting the genetic ancestry of Brazilians using autosomal AIM-Indels. **PloS one**, v. 8, n. 9, p. e75145, 2013.

MARTIN, E. R.; TUNC, I.; LIU, Z.; *et al.* Properties of global-and local-ancestry adjustments in genetic association tests in admixed populations. **Genetic epidemiology**, v. 42, n. 2, p. 214-229, 2018.

MARTIN, G. O povoamento pré-histórico do vale do São Francisco (Brasil). **Clio-Série arqueológica-UFPE**, v. 1, n. 13, p. 9-41, 1998.

MARTIN, P. D.; SCHMITTER, H.; SCHNEIDER, P. M. A brief history of the formation of DNA databases in forensic science within Europe. **Forensic science international**, v. 119, n. 2, p. 225-231, 2001.

MCPHERSON, J. D.; MARRA, M.; HILLIER, L. D. *et al.* A physical map of the human genome. **Nature**, v. 409, n. 6822, p. 934-941, 2001.

MYRES, N. M.; ROOTSI, S.; LIN, A. A. *et al.* A major Y-chromosome haplogroup R1b Holocene era founder effect in Central and Western Europe. **European Journal of Human Genetics**, v. 19, n. 1, p. 95-101, 2011.

NAVARRO-LÓPEZ, B.; GRANIZO-RODRÍGUEZ, E.; PALENCIA-MADRID, L. *et al.* Phylogeographic review of Y chromosome haplogroups in Europe. **International Journal of Legal Medicine**, p. 1-10, 2021.

PADHUKASAHASRAM, B. Inferring ancestry from population genomic data and its applications. **Frontiers in genetics**, v. 5, n. 204, 2014.

PALHA, T.; GUSMAO, L.; RIBEIRO-RODRIGUES, E. *et al.* Disclosing the genetic structure of Brazil through analysis of male lineages with highly discriminating haplotypes. **PLoS One**, v. 7, n. 7, p. e40007, 2012.

PANG, A. W. C.; MIGITA, O.; MACDONALD, J. R. *et al.* Mechanisms of formation of structural variation in a fully sequenced human genome. **Human mutation**, v. 34, n. 2, p. 345-354, 2013.

PENA, S. D.; DI PIETRO, G.; FUCHSHUBER-MORAES, M. *et al.* The genomic ancestry of individuals from different geographical regions of Brazil is more uniform than expected. **PloS one**, 6(2): e17063, 2011.

PIGLIONICA, M.; BALDASSARRA, S. L.; GIARDINA, E. *et al.* Population data for 17 Y-chromosome STRs in a sample from Apulia (Southern Italy). **Forensic Science International: Genetics**, v. 7, n. 1, p. e3-e4, 2013.

PONTES, M. L.; FONDEVILA, M.; LARÉU, M. V.; MEDEIROS, R. SNP markers as additional information to resolve complex kinship cases. **Transfusion Medicine and Hemotherapy**, v. 42, n. 6, p. 385-388, 2015.

PRIMORAC, D.; SCHANFIELD, M. **Forensic DNA applications: an interdisciplinary perspective**. CRC Press, New York, 2014.

PURPS, J.; SIEGERT, S.; WILLUWEIT, S. *et al.* A global analysis of Y-chromosomal haplotype diversity for 23 STR loci. **Forensic Science International: Genetics**, v. 12, p. 12-23, 2014.

RAO, M.; BALAKRISHNAN, V. S. The genetic basis of kidney disease risk in African Americans: MYH9 as a new candidate gene. **American Journal of Kidney Diseases**, v. 53, n. 4, p. 579-583, 2009.

RAVASINI, F.; D'ATANASIO, E.; BONITO, M. *et al.* Sequence read depth analysis of a monophyletic cluster of Y chromosomes characterized by structural rearrangements in the AZFc region resulting in *DYS448* deletion and *DYF387S1* duplication. **Frontiers in Genetics**, v. 12, 2021.

RESQUE, R.; GUSMAO, L.; GEPPERT, M. *et al.* Male lineages in Brazil: intercontinental admixture and stratification of the European background. **PLoS One**, v. 11, n. 4, p. e0152573, 2016.

ROEWER, L. Y chromosome STR typing in crime casework. **Forensic science, medicine and pathology**, v. 5, n. 2, p. 77-84, 2009.

ROEWER, L. Y-chromosome short tandem repeats in forensics - Sexing, profiling, and matching male DNA. Wiley Interdisciplinary Reviews: **Forensic Science**, v. 1, n. 4, p. e1336, 2019.

ROSENBERG, N. A.; PRITCHARD, J. K.; WEBER, J. L. *et al.* Genetic structure of human populations. **Science**, v. 298, n. 5602, p. 2381-2385, 2002.

ROSENBERG, N. A.; LI, L. M.; WARD, R.; PRITCHARD, J. K. Informativeness of genetic markers for inference of ancestry. **The American Journal of Human Genetics**, v. 73, n. 6, p. 1402-1422, 2003.

ROTIMI, C. N.; JORDE, L. B. Ancestry and disease in the age of genomic medicine. **New England Journal of Medicine**, v. 363, n. 16, p. 1551-1558, 2010.

ROYAL, C. D.; NOVEMBRE, J.; FULLERTON, S. M. *et al.* Inferring genetic ancestry: opportunities, challenges, and implications. **The American Journal of Human Genetics**, v. 86, n. 5, p. 661-673, 2010.

SCHAAN, A. P.; GUSMÃO, L.; JANNUZZI, J. *et al.* New insights on intercontinental origins of paternal lineages in Northeast Brazil. **BMC evolutionary biology**, v. 20, n. 1, p. 1-9, 2020.

SCHLÖTTERER, C. The evolution of molecular markers - just a matter of fashion?. **Nature reviews genetics**, v. 5, n. 1, p. 63-69, 2004.

SERRE, D.; PÄÄBO, S. Evidence for gradients of human genetic diversity within and among continents. **Genome research**, v. 14, n. 9, p. 1679-1685, 2004.

SHRIVER, M. D.; KITTLES, R. A. Genetic ancestry and the search for personalized genetic histories. **Nature Reviews Genetics**, v. 5, n. 8, p. 611-618, 2004.

SMART, A.; BOLNICK, D. A.; TUTTON, R. Health and genetic ancestry testing: time to bridge the gap. **BMC medical genomics**, v. 10, n. 1, p. 1-9, 2017.

SOARES-VIEIRA, J. A.; BILLERBECK, A. E.; IWAMURA, E. S. *et al.* Population and mutation analysis of Y-STR loci in a sample from the city of São Paulo (Brazil). **Genetics and Molecular Biology**, v. 31, n. 3, p. 651-656, 2008.

SOUZA, A. M. D.; RESENDE, S. S.; SOUSA, T. N. D.; BRITO, C. F. A. D. A systematic scoping review of the genetic ancestry of the Brazilian population. **Genetics and molecular biology**, v. 42, p. 495-508, 2019.

TISHKOFF, S. A.; KIDD, K. K. Implications of biogeography of human populations for 'race' and medicine. **Nature genetics**, v. 36, n. 11, p. S21-S27, 2004.

TURRINA, S.; CARATTI, S.; FERRIAN, M. DE LEO, D. Deletion and duplication at DYS448 and DYS626 loci: unexpected patterns within the AZFc region of the Y-chromosome. **International journal of legal medicine**, v. 129, n. 3, p. 449-455, 2015.

UNDERHILL, P. A.; POZNIK, G. D.; ROOTSI, S. *et al.* The phylogenetic and geographic structure of Y-chromosome haplogroup R1a. **European Journal of Human Genetics**, v. 23, n. 1, p. 124-131, 2015.

VALENÇA FERRAZ, T. **A formação da sociedade no sertão pernambucano: trajetória de núcleos familiares**. Dissertação (Mestrado em História) – Departamento de História, Universidade Federal de Pernambuco, Recife, 2004.

VALVERDE, L.; KÖHNEMANN, S.; ROSIQUE, M. *et al.* 17 Y-STR haplotype data for a population sample of Residents in the Basque Country. **Forensic Science International: Genetics**, v. 6, n. 4, p. e109-e111, 2012.

VAN OVEN, M.; VAN GEYSTELEN, A.; KAYSER, M. *et al.* Seeing the wood for the trees: a minimal reference phylogeny for the human Y chromosome. **Human mutation**, v. 35, n. 2, p. 187-191, 2014.

WEBER-LEHMANN, J.; SCHILLING, E.; GRADL, G. *et al.* Finding the needle in the haystack: differentiating “identical” twins in paternity testing and forensics by ultra-deep next generation sequencing. **Forensic Science International: Genetics**, v. 9, p. 42-46, 2014.

WILLEMS, T.; GYMREK, M. POZNIK, G. D.; TYLER-SMITH, C.; 1000 GENOMES PROJECT CHROMOSOME Y GROUP. Population-scale sequencing data enable precise estimates of Y-STR mutation rates. **The American Journal of Human Genetics**, v. 98, n. 5, p. 919-933, 2016.

WITHERSPOON, D. J.; WOODING, S.; ROGERS, A. R. *et al.* Genetic similarities within and between human populations. **Genetics**, v. 176, n. 1, p. 351-359, 2007.

YCC, Y CHROMOSOME CONSORTIUM. A nomenclature system for the tree of human Y-chromosomal binary haplogroups. **Genome research**, v. 12, n. 2, p. 339-348, 2002.

YHRD, Y CHROMOSOME HAPLOTYPE REFERENCE DATABASE. 2021. Disponível em: [https://yhrd.org/tools/national_database/Brazil].

ZAUMSEGEL, D.; ROTHSCILD, M. A.; SCHNEIDER, P. M. A 21 marker insertion deletion polymorphism panel to study biogeographic ancestry. **Forensic Science International: Genetics**, v. 7, n. 2, p. 305-312, 2013.

APÊNDICE A - HAPLÓTIPOS E FREQUÊNCIAS HAPLOTÍPICAS

SERTÃO	Fh	H307	0,017857
H24	0,017857	H317	0,017857
H26	0,017857	H327	0,017857
H32	0,017857	H334	0,017857
H42	0,017857	H355	0,017857
H51	0,017857	H360	0,017857
H53	0,017857	H366	0,017857
H55	0,017857	H378	0,017857
H58	0,017857	H380	0,017857
H59	0,017857	H381	0,017857
H63	0,017857	H388	0,017857
H67	0,017857	H392	0,017857
H70	0,017857	H402	0,017857
H71	0,017857	H408	0,017857
H82	0,017857		
H83	0,017857		
H84	0,017857		
H85	0,017857		
H90	0,017857		
H98	0,017857		
H112	0,017857		
H129	0,017857		
H134	0,017857		
H142	0,017857		
H151	0,017857		
H152	0,017857		
H165	0,017857		
H169	0,017857		
H179	0,017857		
H184	0,017857		
H188	0,017857		
H195	0,017857		
H203	0,017857		
H204	0,017857		
H205	0,035714		
H229	0,017857		
H240	0,017857		
H246	0,017857		
H250	0,017857		
H285	0,017857		
H304	0,017857		
H305	0,017857		

AGRESTE Fh

H1	0,011905	H215	0,011905
H7	0,011905	H218	0,011905
H21	0,011905	H224	0,011905
H22	0,011905	H233	0,011905
H28	0,011905	H238	0,011905
H30	0,011905	H241	0,011905
H35	0,011905	H242	0,011905
H37	0,011905	H243	0,011905
H38	0,011905	H248	0,011905
H48	0,011905	H255	0,011905
H52	0,011905	H260	0,011905
H59	0,011905	H261	0,011905
H72	0,011905	H267	0,011905
H73	0,011905	H268	0,011905
H76	0,011905	H271	0,011905
H78	0,011905	H277	0,011905
H79	0,011905	H285	0,011905
H91	0,011905	H291	0,011905
H97	0,011905	H299	0,011905
H105	0,011905	H303	0,011905
H106	0,011905	H306	0,011905
H114	0,011905	H321	0,011905
H117	0,011905	H325	0,011905
H118	0,011905	H339	0,011905
H123	0,011905	H348	0,011905
H125	0,011905	H352	0,011905
H128	0,011905	H358	0,011905
H139	0,011905	H359	0,011905
H141	0,011905	H361	0,011905
H145	0,011905	H372	0,011905
H150	0,011905	H377	0,02381
H155	0,011905	H387	0,011905
H157	0,011905	H394	0,011905
H160	0,011905	H397	0,011905
H164	0,011905	H404	0,011905
H167	0,011905	H413	0,011905
H171	0,011905	H414	0,011905
H181	0,011905	H417	0,011905
H187	0,011905		
H189	0,011905		
H190	0,011905		
H192	0,011905		
H207	0,011905		
H210	0,011905		
H211	0,011905		

S.FRANC	Fh
H2	0,03125
H10	0,03125
H17	0,03125
H49	0,03125
H50	0,03125
H60	0,03125
H65	0,03125
H68	0,03125
H101	0,03125
H109	0,03125
H119	0,03125
H120	0,03125
H136	0,03125
H143	0,03125
H172	0,03125
H173	0,03125
H189	0,03125
H194	0,03125
H209	0,03125
H231	0,03125
H234	0,03125
H237	0,03125
H239	0,03125
H264	0,03125
H278	0,03125
H315	0,03125
H357	0,03125
H369	0,03125
H383	0,03125
H403	0,03125
H407	0,03125
H412	0,03125

APÊNDICE B – HAPLÓTIPOS DE CADA HAPLOGRUPPO/MESORREGIÃO

H	HG	DYS576	DYS389I	DYS448	DYS389II	DYS19	DYS391	DYS481	DYS549	DYS533	DYS438	DYS437	DYS570	DYS635	DYS390	DYS439	DYS392	DYS643	DYS393	DYS458	DYS385	DYS456	YGATAH4
H22	R1b	19	13	19	29	14	11	22	12	12	12	15	16	23	24	12	13	10	13	17 11//14	15	12	
H30	R1b	17	12	19	28	14	11	21	13	12	12	15	18	23	25	12	13	11	13	17 11//14	16	12	
H37	R1b	16	11	19	27	15	10	21	13	12	12	15	17	23	24	12	13	10	13	18 14//14	16	12	
H72	R1b	16	13	19	29	15	11	22	13	11	12	15	19	23	23	12	14	10	12	16 12//12	16	13	
H78	R1b	15	13	19	29	14	11	23	10	12	12	15	17	23	24	12	13	10	13	17 11//14	15	12	
H91	R1b	18	13	19	29	14	11	22	14	13	12	15	17	23	24	12	14	10	13	17 11//15	17	11	
H97	R1b	17	15	18	31	14	11	22	13	12	12	14	16	23	24	12	13	10	13	18 11//14	15	11	
H105	R1b	17	14	19	30	14	10	22	12	12	12	14	18	23	24	11	13	10	13	16 11//15	15	12	
H114	R1b	18	13	19	29	14	10	22	12	11	12	15	19	24	24	12	13	10	13	17 11//14	15	12	
H118	R1b	18	13	19	30	14	10	22	12	12	12	15	17	23	24	12	13	9	14	17 11//14	16	12	
H123	R1b	19	13	19	29	14	11	22	13	12	12	15	18	24	25	12	13	11	14	17 11//14	17	12	
H125	R1b	16	13	19	29	14	10	22	11	13	12	15	16	23	24	11	13	11	13	18 10//15	16	12	
H139	R1b	17	13	19	29	15	11	23	13	13	12	15	17	23	24	12	13	9	13	17 11//14	15	12	
H141	R1b	17	14	19	30	14	11	22	12	12	12	15	16	23	24	12	13	10	13	19 11//15	15	12	
H145	R1b	18	13	19	29	15	11	22	14	13	12	15	16	23	24	11	13	10	13	18 11//15	15	12	
H150	R1b	18	13	18	29	14	11	22	13	12	13	15	16	23	25	12	12	10	13	16 12//16	15	12	
H157	R1b	17	14	19	31	14	10	22	12	12	12	15	17	23	25	12	15	12	13	17 12//14	15	13	
H160	R1b	20	13	19	29	14	11	22	12	12	12	15	17	23	26	12	13	11	12	16 11//12	15	12	
H164	R1b	18	13	19	29	14	11	22	14	11	10	15	17	23	24	12	13	10	13	19 11//14	16	13	
H181	R1b	18	13	20	29	15	11	22	14	12	11	15	17	23	23	12	13	10	13	17 11//14	15	12	
H192	R1b	16	13	19	29	14	11	23	11	12	12	15	17	24	24	13	13	10	13	18 11//15	15	12	
H207	R1b	18	13	18	29	14	11	22	13	12	13	15	16	23	25	12	12	10	13	16 12//15	15	12	
H211	R1b	19	13	19	30	14	10	21	13	12	12	15	18	23	24	12	13	10	13	17 11//14	15	12	
H215	R1b	17	13	19	29	14	11	22	13	13	12	15	17	23	24	11	13	10	13	16 11//14	16	12	
H242	R1b	16	13	19	28	14	10	23	12	12	12	15	17	23	23	11	13	10	13	17 11//14	16	12	
H248	R1b	18	14	18	30	15	13	22	14	13	12	15	17	24	25	12	13	11	13	18 11//13	15	12	
H255	R1b	19	13	19	29	13	11	22	12	12	12	15	17	23	24	14	13	9	13	17 11//14	15	12	
H260	R1b	17	13	19	29	14	11	21	12	13	12	15	19	23	23	13	13	10	13	18 11//13	16	11	
H261	R1b	18	14	19	30	14	10	22	14	12	10	15	17	23	22	13	13	9	13	17 10//14	15	12	
H277	R1b	19	13	19	29	14	11	22	13	12	12	15	17	23	24	11	13	10	12	17 11//14	14	12	
H303	R1b	20	12	18	28	14	11	21	14	12	12	14	17	23	23	12	14	10	13	18 11//14	16	11	
H321	R1b	18	12	19	26	15	11	22	14	11	12	15	17	24	24	11	13	10	13	16 12//14	16	12	
H339	R1b	17	13	19	29	15	10	23	13	13	12	15	17	23	24	12	13	9	13	17 11//14	15	12	
H352	R1b	19	13	19	29	14	11	22	13	12	12	15	17	24	25	12	13	11	13	16 11//14	16	12	
H359	R1b	19	13	19	29	14	10	24	12	12	12	15	17	23	24	12	13	10	13	16 12//12	17	12	
H361	R1b	16	13	19	29	14	10	22	11	13	12	15	16	23	24	11	13	11	13	18 10//15	17	12	
H387	R1b	18	12	19	27	14	12	22	15	12	12	15	16	23	24	11	13	10	14	17 11//14	15	12	
H413	R1b	17	14	19	28	14	11	22	13	12	12	15	17	23	23	11	13	10	13	17 11//15	15	13	
H417	R1b	17	13	19	29	15	11	23	12	12	12	14	18	23	25	11	14	10	13	18 10//14	16	12	
H7	J1	16	13	20	30	14	10	24	11	11	11	15	19	20	23	12	11	9	12 17.2	13//19	16	11	
H35	J1	19	13	21	30	15	10	25	13	11	10	14	18	21	23	13	11	9	12 17.2	13//15	15	11	
H38	J1	18	13	20	29	14	11	25	12	12	10	14	19	21	23	11	11	9	12 18.2	13//18	13	12	

(Continuação: Agreste)

H	HG	DYS576	DYS389I	DYS448	DYS389II	DYS19	DYS391	DYS481	DYS549	DYS533	DYS438	DYS437	DYS570	DYS635	DYS390	DYS439	DYS392	DYS643	DYS393	DYS458	DYS385	DYS456	YGATAH4
H59	J1	16	14	20	29	14	10	22	13	12	9	14	17	21	23	11	13	10	13	18 14//17	15	10	
H155	J1	17	13	20	29	14	11	25	12	12	10	14	18	21	23	11	11	9	12 18.2	13//18	13	12	
H189	J1	18	13	20	30	14	10	23	12	11	9	15	16	21	22	11	11	8	12	15 13//15	15	11	
H243	J1	17	12	20	29	13	10	25	12	11	10	14	17	21	24	14	11	9	12 16.2	12//19	15	10	
H285	J1	19	13	21	29	14	10	24	12	12	9	15	15	22	26	11	11	11	12	20 13//13	15	11	
H299	J1	18	13	20	30	14	10	23	12	11	9	15	16	21	22	11	11	8	12	15 13//16	15	11	
H397	J1	18	13	20	29	14	11	25	12	12	10	14	18	21	23	11	11	9	12 18.2	13//13	13	12	
H1	E1b1b	17	12	20	29	13	10	22	12	12	10	14	20	22	24	15	11	12	13	15 17//18	17	11	
H21	E1b1b	19	14	20	30	13	9	29	11	11	10	14	23	21	23	10	11	12	13	18 13//14	15	12	
H73	E1b1b	17	12	20	29	13	10	22	12	12	10	14	18	23	23	13	11	13	13	16 17//17	16	12	
H128	E1b1b	18	14	20	31	13	10	23	12	12	10	14	21	21	23	10	11	12	13	13 17//20	15	11	
H171	E1b1b	19	12	20	29	15	10	23	13	11	10	14	18	20	24	12	11	12	13	15 16//16	15	11	
H187	E1b1b	19	14	20	30	13	10	28	12	11	10	14	22	21	24	10	11	12	13	18 13//14	15	12	
H210	E1b1b	19	14	20	30	13	9	28	11	11	10	14	22	22	23	10	11	12	13	18 13//14	16	12	
H325	E1b1b	17	13	21	30	13	10	21	12	11	10	14	18	20	23	11	11	13	14	16 16//16	16	11	
H404	E1b1b	18	13	19	31	15	10	24	13	11	10	14	20	20	24	11	11	12	13	15 19//20	17	12	
H52	E1b1a	14	13	21	31	15	10	28	11	11	11	14	19	21	21	12	11	12	14	16 15//20	15	11	
H76	E1b1a	16	13	20	31	15	10	27	12	11	11	14	21	22	21	12	11	13	13	17 15//15	16	12	
H117	E1b1a	16	13	21	31	15	11	28	11	11	11	14	20	21	21	11	11	14	13	16 16//17	15	12	
H233	E1b1a	17	13	20	30	16	10	26	12	12	11	14	19	23	21	12	11	12	14	18 15//16	16	11	
H241	E1b1a	16	13	21	30	15	10	26	11	11	11	14	18	21	21	12	11	13	14	16 17//18	20	11	
H267	E1b1a	16	13	21	30	15	10	27	12	11	11	14	17	20	21	12	11	13	15	16 17//18	16	11	
H271	E1b1a	16	14	20	31	15	9	29	12	11	10	14	18	21	22	12	12	12	14	16 13//17	13	11	
H358	E1b1a	16	13	21	32	15	10	28	11	11	11	14	18	21	21	12	11	14	13	18 16//16	15	12	
H372	E1b1a	17	13	19	31	14	10	26	12	12	14	16	18	17	20	13	14	7	13	16 15//19	15	11	
H218	Q	18	13	19	29	13	10	26	12	12	9	14	17	21	22	11	13	12	13	16 13//13	16	12	
H306	Q	17	13	19	29	15	10	21	11	10	10	14	20	22	23	11	14	11	14	17 11//13	14	12	
H394	Q	19	12	20	29	13	10	23	12	12	11	14	17	23	24	13	15	9	12	18 13//14	16	12	
H28	G2a	15	12	22	29	15	10	21	11	10	10	16	18	21	22	11	11	12	14	17 13//15	16	11	
H190	G2a	15	12	23	29	15	10	22	12	10	10	16	17	22	21	11	11	12	15	17 13//17	15	11	
H224	G2a	17	13	21	29	14	9	21	11	10	9	15	17	23	22	11	11	10	12	17 13//17	15	11	
H238	G2a	15	12	22	29	15	10	21	11	10	10	16	18	20	22	11	11	12	14	17 13//16	15	11	
H348	G2a	15	12	21	30	16	10	21	13	9	10	16	18	21	22	12	12	11	14	17 14//15	14	12	
H79	I1	17	12	20	29	14	10	24	12	11	10	16	21	22	23	12	11	12	12	15 14//14	14	11	
H106	I1	16	12	20	28	14	11	27	12	11	10	16	19	21	23	11	11	12	13	15 13//14	15	11	
H377	I1	16	12	20	28	14	10	26	12	11	10	16	19	23	22	11	11	12	13	15 13//15	14	11	
H268	I2b1	18	13	20	29	16	11	26	12	14	10	15	20	20	23	11	12	13	14	15 15//15	16	11	
H291	I2b1	18	14	20	31	15	10	28	13	12	10	14	18	22	23	12	12	13	14	15 14//16	13	12	
H48	I2a1	21	13	21	28	17	10	23	10	12	10	14	20	22	23	11	11	12	13	17 12//13	14	12	
H167	R1a	17	13	20	30	16	11	23	12	13	12	14	19	23	25	10	11	10	13	16 11//11	16	12	
H414	J2b	16	12	19	28	15	10	23	13	11	9	16	18	21	24	11	11	9	13	15 13//17	13	11	

SERTÃO

H	HG	DYS576	DYS389I	DYS448	DYS389II	DYS19	DYS391	DYS481	DYS549	DYS533	DYS438	DYS437	DYS570	DYS635	DYS390	DYS439	DYS392	DYS643	DYS393	DYS458	DYS385	DYS456	YGATAH4
H32	R1b	16	13	19	29	15	12	22	13	12	11	15	18	23	23	12	13	10	13	16 11//14	15	12	
H42	R1b	18	13	19	29	14	11	22	13	12	13	15	17	24	23	12	13	10	13	17 11//14	17	12	
H58	R1b	19	13	19	29	14	11	22	13	13	12	15	17	23	24	12	13	10	12	19 11//14	15	12	
H63	R1b	18	14	19	30	14	11	22	12	12	13	15	17	23	25	11	13	10	12	16 11//14	15	13	
H67	R1b	18	13	19	29	15	10	22	13	12	12	15	17	23	23	13	13	10	13	17 12//13	17	12	
H70	R1b	19	14	19	31	14	11	22	13	12	12	15	17	24	24	13	13	11	13	18 11//15	16	12	
H71	R1b	17	13	19	29	13	11	22	12	12	12	15	17	23	24	13	13	10	13	16 11//14	15	12	
H82	R1b	20	13	21	29	14	11	22	14	12	12	15	17	24	24	12	13	10	13	17 11//14	15	11	
H98	R1b	18	13	19	29	14	11	25	12	13	12	15	17	24	24	11	13	10	14	18 11//14	17	12	
H142	R1b	18	13	19	30	14	11	22	12	12	13	15	17	23	24	13	13	10	14	17 12//14	16	12	
H151	R1b	18	13	19	30	14	10	23	13	12	12	15	18	23	24	11	13	10	13	18 11//14	15	12	
H165	R1b	17	14	18	31	14	11	22	13	13	12	14	17	24	24	12	13	10	13	18 10//13	17	11	
H169	R1b	17	13	19	29	14	11	19	13	12	12	15	18	24	24	12	13	10	13	18 11//14	15	12	
H179	R1b	18	13	18	30	15	11	21	12	12	12	14	19	23	24	11	13	9	13	17 11//13	18	11	
H188	R1b	18	13	19	30	14	11	22	12	12	12	15	17	23	24	12	13	10	13	17 11//14	16	11	
H205	R1b	18	13	19	30	14	10	22	13	12	12	15	17	23	23	10	13	10	13	17 11//14	15	11	
H250	R1b	19	12	19 28//29		14	11	22	13	11	10	15	19	23	24	12	13	10	13	17 11//14	16	11	
H304	R1b	19	11	19	27	14	12	22	13	12	13	15	17	25	23	14	13	10	13	16 11.2//14		17	11
H305	R1b	18	12	19	28	14	11	22	13	12	11	15	18	24	24	13	13	10	15	17 11//14	15	12	
H327	R1b	19	13	19	29	14	11	22	13	12	12	15	17	23	22	13	13	10	13	17 11//14	16	12	
H334	R1b	18	13	19	29	14	11	24	13	12	12	16	17	23	24	11	13	10	13	19 11//14	16	11	
H355	R1b	16	13	19	30	14	11	20	13	12	12	15	19	23	25	12	13	10	13	16 11//14	16	11	
H381	R1b	17	14	19	30	15	11	24	12	14	12	15	18	23	24	12	13	9	13	17 11//14	15	12	
H388	R1b	19	14	19	30	14	11	22	13	12	12	15	17	23	24	12	13	11	13	17 11//14	16	11	
H392	R1b	18	13	19	30	14	11	22	12	12	13	15	17	23	24	13	13	10	14	17 11//14	16	12	
H408	R1b	18	13	19	29	14	11	23	14	12	12	15	18	24	24	12	13	10	14	17 11//14	15	12	
H59	J1	16	14	20	29	14	10	22	13	12	9	14	17	21	23	11	13	10	13	18 14//17	15	10	
H129	J1	15	13	19	28	13	10	22	12	11	9	14	18	21	23	12	13	10	13	17 13//15	15	11	
H152	J1	18	13	20	31	15	10	25	12	11	10	14	17	20	23	11	11	9	12 17.2	13//18	15	11	
H184	J1	19	11	19	27	15	10	28	12	11	11	14	20	20	25	12	11	10	13	18 9//12	15	10	
H195	J1	17	12	20	29	13	10	26	12	11	10	14	17	21	24	14	11	9	12 16.2	12//19	15	10	
H240	J1	19	13	20	30	14	10	22	13	11	9	15	17	23	23	11	11	10	12	18 13//17	15	12	
H285	J1	19	13	21	29	14	10	24	12	12	9	15	15	22	26	11	11	11	12	20 13//13	15	11	
H402	J1	15	13	19	28	13	10	23	13	11	9	14	18	21	23	12	13	10	13	18 13//15	16	11	
H51	E1b1b	18	14	20	30	13	9	28	11	12	10	14	22	21	24	10	11	12	13	18 13//14	16	12	
H83	E1b1b	17	13	21	29	15	9	22	13	12	9	14	19	23	23	11	11	10	12	15 13//17	15	12	
H84	E1b1b	18	13	21	31	13	10	25	12	10	10	14	19	22	23	12	11	12	13	18 16//16	15	11	
H203	E1b1b	17	14	20	30	13	9	28	11	11	10	14	21	21	24	10	11	12	13	18 13//14	16	13	
H246	E1b1b	18	13	21	31	16	10	25	13	8	10	14	22	23	24	11	11	10	12	15 13//18	15	10	
H380	E1b1b	18	14	20	30	15	10	23	12	12	9	14	23	23	23	12	11	10	12	19 15//20	15	11	
H90	Q	16	14	20	31	13	10	25	12	11	11	14	17	22	24	11	14	12	13	17 13//15	17	12	
H204	Q	18	12	21	28	14	10	27	12	12	10	14	19	18	25	11	12	10	13	17 11//17	15	12	

(Continuação: Sertão)

H	HG	DYS576	DYS389I	DYS448	DYS389II	DYS19	DYS391	DYS481	DYS549	DYS533	DYS438	DYS437	DYS570	DYS635	DYS390	DYS439	DYS392	DYS643	DYS393	DYS458	DYS385	DYS456	YGATAH4
H307	Q	17	14	19	30	12	11	20	11	12	10	14	19	21	23	11	14	11	13	16 10//13	13	12	
H317	Q	20	12	20	29	14	10	24	13	10	11	14	16	22	23	12	14	10	13	15 13//18	15	12	
H24	G2a	13	12 19//20		30	17	10	27	13	12	10	17	18	20	22	11	12	12	12	15 13//15	16	11	
H85	G2a	18	12	21	28	16	10	23	13	9	11	16	17	20	22	11	11	11	14	16 14//15	17	12	
H229	I1	14	12	20	28	15	10	25	12	11	10	16	19	21	23	11	11	12	13	15 13//17	14	12	
H378	I1	17	12	20	29	14	10	24	12	11	10	16	21	22	23	12	11	12	12	15 14//14	13	11	
H26	I2b1	18	13	20	30	15	10	25	11	12	10	15	19	19	24	11	12	12	15	16 15//15	15	11	
H134	I2a1	17	14	21	29	16	10	21	11	12	10	15	20	22	22	12	11	12	13	17 11//12	13	13	
H366	I2a1	20	13	21	28	15	10	23	10	12	10	14	18	22	23	11	11	12	13	18 12//12	14	12	
H112	R1a	17	13	20	30	16	11	23	13	12	11	14	19	24	25	10	11	10	13	15 11//12	15	12	
H53	I2a (xI2a1)	18	13	20	29	16	11	26	12	14	10	15	21	21	23	11	12	13	14	15 15//15	16	11	
H360	I2a (xI2a1)	20	13	20	30	15	10	24	12	10	10	14	18	20	24	11	12	11	14	17 15//16	14	11	
H55	L	18	13	19	29	14	11	23	12	11	9	15	17	22	23	12	11	11	12	17 13//16	16	11	

SÃO FRANCISCO

H	HG	DYS576	DYS389I	DYS448	DYS389II	DYS19	DYS391	DYS481	DYS549	DYS533	DYS438	DYS437	DYS570	DYS635	DYS390	DYS439	DYS392	DYS643	DYS393	DYS458	DYS385	DYS456	YGATAH4
H2	R1b	17	14	18	30	14	11	22	13	12	12	14	17	23	24	12	13	10	13	17 12//15	14	11	
H17	R1b	18	13	19	29	14	10	23	13	12	13	15	17	23	24	13	13	10	13	16 11//14	15	12	
H65	R1b	18	13	19	29	15	11	22	13	12	12	15	17	23	24	12	13	10	13	16 11//15	17	12	
H68	R1b	18	14	19	30	14	11	22	14	13	12	15	16	24	24	12	13	10	13	17 11//14	11	11	
H109	R1b	18	13	19	29	15	11	22	12	11	12	15	16	23	24	12	13	10	14	16 11//14	16	12	
H172	R1b	18	13	19	29	14	11	22	13	12	12	15	17	24	24	12	13	10	13	17 11//15	15	11	
H173	R1b	19	13	19	29	15	11	22	13	12	12	15	17	24	24	12	13	10	14	16 11//15	15	12	
H209	R1b	16	13	19	29	14	11	24	12	12	12	15	17	23	23	12	13	10	13	16 14//14	16	12	
H264	R1b	18	12	19	31	15	10	23	13	12	12	15	18	23	23	12	13	10	13	17 11//14	16	13	
H278	R1b	18	14	18	30	13	11	22	13	12	12	15	16	23	24	13	13	10	13	18 12//14	16	11	
H315	R1b	17	13	19	29	14	11	22	13	12	12	15	17	23	24	11	13	10	14	17 11//14	15	12	
H369	R1b	16	13	18	29	15	11	22	12	12	12	14	17	23	24	13	13	10	13	17 12//16	15	11	
H407	R1b	19	13	18	31	15	11	22	13	11	12	14	17	23	24	13	13	10	13	18 11//13	16	12	
H412	R1b	18	13	19	29	14	11	23	12	13	12	15	18	23	25	12	13	10	13	18 12//14	15	12	
H189	J1	18	13	20	30	14	10	23	12	11	9	15	16	21	22	11	11	8	12	15 13//15	15	11	
H194	J1	19	14	20	32	14	10	24	12	11	10	14	19	21	23	11	11	9	12 17.2	13//17	15	11	
H10	E1b1b	17	13	20	30	14	9	25	12	11	10	14	16	22	25	13	11	12	13	16 16//17	15	12	
H119	E1b1b	17	13	20	31	15	10	25	13	8	10	14	21	21	24	11	11	10	12	15 13//18	15	10	
H136	E1b1b	19	14	20	30	13	9	29	11	11	10	14	23	21	24	10	11	12	13	18 13//14	16	12	
H239	E1b1b	17	13	20	30	13	10	22	12	12	11	14	20	22	24	11	11	12	13	16 15//18	16	13	
H357	E1b1b	19	14	20	30	13	9	27	12	11	10	14	22	20	24	10	11	12	14	18 13//14	17	12	
H120	Q	20	12	20	29	13	10	23	12	11	11	14	17	23	24	12	15	9	12	19 14//15	16	12	
H237	Q	18	14	19	31	13	11	24	12	12	11	15	17	22	24	14	14	11	11	15 14//16	15	12	
H143	G2a	15	12	23	29	15	11	22	12	10	10	16	17	21	21	11	13	12	15	17 13//18	15	11	
H403	G2a	15	12	23	30	15	10	22	12	10	10	16	17	22	21	11	11	12	15	17 13//19	15	11	
H50	I1	15	13	20	29	14	10	24	12	11	10	16	20	21	23	11	11	12	13	15 13//15	14	11	
H49	I2b1	16	14	20	31	16	10	28	12	11	10	14	18	23	22	11	12	12	14	15 13//17	13	11	
H60	I2b1	16	14	20	31	15	10	28	12	11	10	14	18	23	22	11	12	12	14	15 13//17	13	11	
H231	I2a1	18	13	21	28	16	10	23	9	12	10	15	20	21	23	11	11	12	13	18 12//12	14	12	
H234	I2a1	19	13	21	28	17	10	23	10	12	10	14	19	22	23	11	11	13	13	17 12//13	14	12	
H101	R1a	18	13	20	31	16	10	23	12	13	11	14	18	23	25	10	11	10	14	15 11//14	15	13	
H383	J2b	17	12	19	28	14	11	23	13	13	9	16	18	21	24	11	11	9	12	16 13//18	13	11	