



Universidade do Estado do Rio de Janeiro
Centro Biomédico
Instituto de Biologia Roberto Alcântara Gomes

Silvia de Oliveira Loiola Martins

Estruturação gênica e ancestralidade da população brasileira: estudo da população nativa da Ilha de Marajó com o emprego de marcadores de DNA

Rio de Janeiro

2022

Silvia de Oliveira Loiola Martins

Estruturação gênica e ancestralidade da população brasileira: estudo da população nativa da Ilha de Marajó com o emprego de marcadores de DNA

Tese apresentada, como requisito parcial para obtenção do título de Doutor, ao Programa de Pós-Graduação em Biociências, da Universidade do Estado do Rio de Janeiro.

Orientador: Prof. Dr. Elizeu Fagundes de Carvalho

Coorientadora: Prof.^a Dra. Leonor Gusmão

Rio de Janeiro

2022

CATALOGAÇÃO NA FONTE
UERJ/REDE SIRIUS/BIBLIOTECA CB-A

M386 Martins, Silvia de Oliveira Loiola.
Estruturação gênica e ancestralidade da população brasileira: estudo da população nativa da Ilha de Marajó com o emprego de marcadores de DNA / Silvia de Oliveira Loiola Martins. - 2020.
197f.

Orientador: Prof. Dr. Elizeu Fagundes de Carvalho
Coorientadora: Prof.^a Dra. Leonor Gusmão

Doutorado (Tese) - Universidade do Estado do Rio de Janeiro, Instituto de Biologia Roberto Alcântara Gomes. Pós-graduação em Biociências.

1. Hereditariedade – Teses. 2. Marajó, Ilha de (PA) – Teses. 3. Genética de populações – Teses. 4. Cromossomo Y - Teses. 5. DNA mitocondrial – Teses. I. Carvalho, Elizeu Fagundes de. II. Gusmão, Leonor. III. Universidade do Estado do Rio de Janeiro. Instituto de Biologia Roberto Alcântara Gomes. IV. Título.

CDU 575.1

Bibliotecária: Ana Rachel Fonseca de Oliveira
CRB7/6382

Autorizo apenas para fins acadêmicos e científicos, a reprodução total ou parcial desta tese, desde que citada a fonte.

Assinatura

Data

Silvia de Oliveira Loiola Martins

Estruturação gênica e ancestralidade da população brasileira: estudo da população nativa da Ilha de Marajó com o emprego de marcadores de DNA

Tese apresentada, como requisito parcial para obtenção do título de Doutor, ao Programa de Pós-Graduação em Biociências, da Universidade do Estado do Rio de Janeiro.

Aprovada em 11 de fevereiro de 2022.

Coorientadora: Prof^a Dra. Leonor Gusmão

Instituto de Biologia Roberto Alcântara Gomes – UERJ

Banca Examinadora:

Prof. Dr. Elizeu Fagundes de Carvalho (Orientador)

Instituto de Biologia Roberto Alcântara Gomes – UERJ

Prof.^a Dra. Tatiana de Almeida Simão

Instituto de Biologia Roberto Alcântara Gomes – UERJ

Dra. Fernanda Saloum de Neves Manta

Instituto Oswaldo Cruz

Dra. Verônica Gomes

Universidade do Porto

Rio de Janeiro

2022

DEDICATÓRIA

Dedico este trabalho:

A todos os que se voluntariaram para que ele pudesse ser realizado.

A todos que contribuíram para a minha formação acadêmica.

A Deus, que me capacitou a chegar até aqui.

AGRADECIMENTOS

Agradeço sobretudo a Deus por sempre me sustentar, abençoando todas as áreas da minha vida. Porque dEle, por meio dEle e para Ele são todas as coisas. Dependendo sempre dEle.

Aos meus pais, Itamar (*in memorian*) e Sueli, por todo amor, cuidado e incentivo. E por terem me direcionado ao caminho do estudo e do conhecimento.

À minha família como um todo por todo apoio e incentivo e todos os amigos que se alegram e me ajudam com suas orações.

Ao meu marido Joel, por ser um grande companheiro e incentivador, com sua calma e paciência enormes! Com você divido a vida, os momentos alegres que você tem o dom de potencializar e os momentos difíceis, nos quais você me traz calma, demonstrando apoio incondicional e o seu amor.

Ao Prof.^o Dr. Elizeu Fagundes de Carvalho, por todo o apoio e oportunidade de aprendizado através das atividades no Laboratório de Diagnósticos por DNA, que me trouxe amadurecimento e crescimento profissional.

À Prof^a Dra. Leonor Gusmão, por todo o aprendizado, orientação e atenção, além do apoio e paciência sempre presentes durante o desenvolvimento deste projeto; e por ser uma pessoa inspiradora, demonstrando generosidade e desejo de promover o crescimento das pessoas.

À Prof^a. Dr^a. Dayse Aparecida da Silva, por todo apoio, ajuda, amizade e conversas sempre incentivadoras, que me trouxeram crescimento profissional e acadêmico.

Ao Professor Dr. Luiz Marcelo de Lima Pinheiro pela parceria neste projeto.

Ao Professor Dr. César Amaral por todo incentivo e aprendizado.

A todos os meus amigos técnicos e da área administrativa do LDD – Jaque, Martinha, Flávio, Paty, Laís, João, Cris, Andréa, Henrique e Ed (*in memorian*). Não tenho palavras para descrever o quanto são importantes no meu dia a dia! Obrigada por tudo...ajuda técnica, parceria, amizade, companhia, risadas e conversas!

À Filipa, por toda a atenção e contribuição no desenvolvimento desse trabalho.

A todos os alunos do LDD com quem aprendi e a quem pude ensinar. Especialmente aos alunos Ju Jannuzzi, July, Masinda, Rodrigo e Allan por compartilharem conhecimentos, por toda a ajuda e pela companhia em tantos momentos!

A todos que, de alguma maneira, contribuíram para a realização deste projeto, o meu muitoobrigada!

RESUMO

MARTINS, Silvia de Oliveira Loiola. **Estruturação gênica e ancestralidade da população brasileira**: estudo da população nativa da Ilha de Marajó com o emprego de marcadores de DNA. 2022. 197 f. Tese (Doutorado em Biociências) – Instituto de Biologia Roberto Alcântara Gomes, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2022.

A população brasileira, altamente miscigenada, é resultante de três componentes ancestrais, de origens continentais europeia, africana e nativa americana. Diferentes padrões dessa mistura podem ser vistos ao longo do Brasil e trabalhos realizados com marcadores de linhagem e autossômicos têm demonstrado um predomínio da ancestralidade européia, mais acentuada nas regiões sul e sudeste, seguida pela ancestralidade africana, em maior proporção no nordeste, e pelo componente nativo, mais evidente na região Norte. Além das diferenças observadas entre as regiões geopolíticas do Brasil, existem evidências de subestrutura genética em uma mesma região. Na região Norte já foi visto que há diferenças entre populações urbanas, com maior contribuição europeia, e populações rurais, que têm maior ancestralidade nativa. Este trabalho tem como objetivo o estudo da população da Ilha de Marajó, no estado do Pará, na Região Norte, com o emprego de marcadores de DNA de origem uniparental (cromossomo Y e DNA mitocondrial - mtDNA) e com marcadores autossômicos informativos de ancestralidade (AIMs). Para investigar o *background* genético da população Ilha de Marajó, selecionou-se uma amostra de indivíduos não aparentados, com avós ali nascidos, dos quais 97 homens foram analisados para os 27 loci Y-STRs (*Short Tandem Repeats*) do kit *Yfiler Plus* e 47 Y-SNPs (*Single Nucleotide Polymorphisms*); a região controle do mtDNA de 101 indivíduos também foi sequenciada e 160 amostras foram genotipadas para o multiplex com 46 marcadores do tipo inserção-deleção (Indels). Resultados dos haplótipos Y-STR indicam uma diversidade haplotípica menor que o observado para populações urbanas de outras regiões (0,9959) e da região Norte, incluindo os Nativos americanos de São Gabriel da Cachoeira. As distâncias genéticas foram baixas entre o Marajó e todas as outras populações do Norte, mas alta com São Gabriel da Cachoeira, indicando uma predominância de linhagens masculinas europeias, o que foi confirmado pela proporções de haplogrupos determinados por Y-SNPs, de 71 % de linhagens européias, 17 % de africanas e 12 % de nativas. Esses dados foram corroborados por distâncias genéticas F_{ST} , demonstrando a proximidade das linhagens masculinas brasileiras com o continente Europeu. Na análise do mtDNA foi observada uma diversidade haplotípica de 0,9897, inferior ao valor observado para outras populações urbanas brasileiras. A composição de haplogrupos revelou alta contribuição de linhagens maternas nativas americanas (62,9 %). Haplogrupos africanos representaram 36,1 % das amostras e apenas 1 % de origem europeia. A distribuição das proporções de ancestralidade biparental, com os AIM-Indels, foi de 40,3 % nativa americana, 35,4 % europeia e 24,3 % africana. Esses valores não se distanciam muito dos valores médios para os marcadores uniparentais nessa população, indicando baixo influxo de ancestralidade europeia recente, no âmbito de três gerações. Os resultados obtidos permitirão a criação de uma base de dados para a população miscigenada da Ilha de Marajó, sendo esta a primeira vez que foi avaliado o perfil de ancestralidade genética dessa população.

Palavras-chave: Ancestralidade. Ilha de Marajó. Cromossomo Y. mtDNA. AIMs. Indels.

ABSTRACT

MARTINS, Silvia de Oliveira Loiola. **Genetic structure and ancestry of the Brazilian population:** study of the native population of the Island of Marajó using DNA markers. 2022. 197 f. Tese (Doutorado em Biociências) – Instituto de Biologia Roberto Alcântara Gomes, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2022.

Brazilian population is considered highly admixed and different patterns of the mixture between European, African and Native-american ancestry can be observed across the country. Predominance of European ancestry, more pronounced in the South and Southeast regions, followed by African ancestry, in greater proportion in the northeast, and finally the native component, more evident in the North region, have been demonstrated through data from autosomal and lineage markers. The Island of Marajó is in the north of Pará state, Northern Brazil. At the time of Portuguese colonization, it is believed that 30 different indigenous groups inhabited the Island. Currently it is composed by an admixed population, as a result of interethnic mating between European colonizers, Native Americans and African slaves. To investigate the genetic background of the Marajó population, a sample of unrelated individuals was selected, with matrilineal and / or patrilineal local ancestry for at least three generations. A total of 97 males were genotyped for the 27 Y chromosome STR included in the Yfiler Plus kit and 47 Y-SNPs. Additionally, 101 samples were sequenced for the complete mtDNA control region and 160 were genotyped for 46 autosomal insertion-deletion ancestry informative markers in a single multiplex reaction. The haplotype diversities obtained for Y-chromosome and mtDNA markers are lower (0.9959 and 0,9897, respectively) when compared with other Brazilian admixed populations, probably due to the geographic isolation of the island. The mtDNA haplogroup composition revealed a high contribution of Native American maternal ancestry of approximately 63 %, followed by 36 % of African haplogroups and only 1 % where from European origin. Conversely, 71% of male haplogroups, defined by Y-SNPs, were of European origin, 17 % sub-Saharan African and Native American ancestry was responsible for 12 %. In the study of the autosomal ancestry of the individuals, a 40,3 % of Native American ancestry was found, followed by 35.4 % of European and 24,3 % of African contributions. These values are close to the mean values for uniparental markers in this population, indicating a low influx of recent European ancestry within three generations. The results obtained will allow the creation of a database for the admixed population of Ilha de Marajó, this being the first time that the genetic ancestry profile of this population has been evaluated.

Keywords:: Ancestry. Marajó Island. Y chromosome. mtDNA. AIMs. Indels.

LISTA DE FIGURAS

Figura 1 –	Exemplos de diferentes tipos de marcadores genéticos: SNP, Indel e STR.....	29
Figura 2 –	Diferentes tipos de genomas e suas propriedades de transmissão e de recombinação.....	34
Figura 3 –	Representação da estrutura circular do genoma mitocondrial.....	39
Figura 4 –	Representação simplificada dos haplogrupos de mtDNA na <i>PhyloTree</i> e suas ramificações.....	42
Figura 5 –	O cromossomo Y.....	43
Figura 6 –	Árvore filogenética de haplogrupos do cromossomo Y.....	52
Figura 7 –	Representação esquemática do processo de miscigenação.....	54
Figura 8 –	Rotas do tráfico de escravos africanos para o Brasil.....	66
Figura 9 –	O Território da Ilha de Marajó.....	70
Figura 10 –	A Ilha de Marajó e seus 16 municípios, com destaque para os municípios de origem de 218 indivíduos da população autóctone.....	77
Figura 11 –	Marcadores SNP do cromossomo Y usados neste trabalho, em destaque na árvore filogenética do cromossomo Y.....	89
Figura 12 –	Eletroferograma do perfil genético da amostra controle (007) obtido através da tipagem com o <i>kit Yfiler Plus</i>	105
Figura 13 –	Gráfico MDS das distâncias genéticas F_{ST} entre as populações brasileiras e populações de referência africana, europeia e nativa americana - haplótipos <i>Yfiler Plus</i>	121
Figura 14 –	Gráfico MDS das distâncias genéticas F_{ST} entre as populações da região Norte do Brasil - conjunto de 18 marcadores Y-STR.....	122
Figura 15 –	Eletroferograma do perfil genético de uma amostra obtido através do sequenciamento da região controle do mtDNA - <i>kit BigDye Terminator</i> .	123
Figura 16 –	Representação gráfica das proporções de ancestralidades continentais dos haplogrupos de mtDNA de populações brasileiras analisadas para a Região Controle.....	128
Figura 17 –	Representação gráfica das proporções de ancestralidades continentais dos haplogrupos de mtDNA de populações da Região Norte analisadas	

	para a Região Controle.....	129
Figura 18 –	Gráfico MDS das distâncias genéticas F_{ST} entre as populações brasileiras, incluindo a Ilha de Marajó, e populações da América do Sul - Região Controle do mtDNA.....	131
Figura 19 –	Eletoferograma do perfil genético de uma amostra controle obtido através da tipagem com o Multiplex de 46 AIM–Indels.....	138
Figura 20 –	Representação gráfica das proporções de ancestralidade nativa americana (NAM), europeia (EUR) e africana (AFR) nos 160 indivíduos da população (A) e a variabilidade entre os indivíduos genotipados (B).....	139
Figura 21 –	Representação gráfica das proporções de ancestralidade nativa americana (NAM), europeia (EUR) e africana (AFR) na população do Marajó com os marcadores do Cromossomo Y, do mtDNA, dos AIM-Indels, e do valor da média entre Cromossomo Y e mtDNA.....	140

LISTA DE TABELAS

Tabela 1 –	Critérios estabelecidos na seleção de indivíduos para o estudo da população autóctone da Ilha de Marajó.....	76
Tabela 2 –	27 loci do kit <i>Yfiler Plus</i>	83
Tabela 3 –	Condições termocíclicas para a amplificação dos multiplexes para o estudo dos Y-SNPs.....	85
Tabela 4 –	Conjuntos de <i>primers</i> das PCRs multiplexes para o estudo dos Y-SNPs.....	87
Tabela 5 –	Conjuntos de <i>primers</i> das reações SBE multiplexes para o estudo dos Y-SNPs.....	90
Tabela 6 –	Populações brasileiras utilizadas para comparação com a população da Ilha de Marajó, com base nos dados obtidos pela genotipagem dos 27 Y-STRs <i>Yfiler Plus</i>	93
Tabela 7 –	Populações da Região Norte do Brasil utilizadas para comparação com a população da Ilha de Marajó, com base nos dados obtidos pela genotipagem de 18 Y-STRs.....	93
Tabela 8 –	Condições termocíclicas para a amplificação da Região Controle do mtDNA com o <i>QIAGEN Multiplex PCR Master Mix</i>	95
Tabela 9 –	<i>Primers</i> usados na amplificação e no sequenciamento da Região Controle do mtDNA.....	96
Tabela 10 –	Condições termocíclicas para a amplificação dos fragmentos RC1 e RC2 do mtDNA com o <i>QIAGEN Multiplex PCR Master Mix</i>	96
Tabela 11 –	Populações utilizadas para comparação com a população da Ilha de Marajó, com base nos dados obtidos pelo sequenciamento da Região Controle completa	99
Tabela 12 –	Condições termocíclicas para a amplificação do multiplex AIM-Indels com o <i>QIAGEN Multiplex PCR Master Mix</i>	101
Tabela 13 –	Sistema multiplex 46 AIM-Indels.....	101
Tabela 14 –	Distribuição, por município de nascimento na Ilha de Marajó, dos avós paternos dos indivíduos analisados para o cromossomo Y.....	103
Tabela 15 –	Distribuição, por município de nascimento na Ilha de Marajó, das avós	

	maternas dos indivíduos analisados para o DNA mitocondrial.....	103
Tabela 16 –	Distribuição, por município de nascimento na Ilha de Marajó, dos quatro avós dos 160 indivíduos analisados para os AIM-Indels.....	104
Tabela 17 –	Resultado da determinação dos haplogrupos pelos <i>software Haplogroup Predictor</i> (HP) e <i>NevGen</i> e pela genotipagem com Y-SNPs, além da origem continental dos haplogrupos.....	110
Tabela 18 –	Frequências dos haplogrupos do Cromossomo Y na população da Ilha de Marajó.....	114
Tabela 19 –	Valores de Diversidade Haplotípica para populações brasileiras e populações de referência africana, europeia e nativa americana - <i>Yfiler Plus</i>	115
Tabela 20 –	Valores de Diversidade Haplotípica para populações da Região Norte -18 Y-STRs.....	116
Tabela 21 –	Matriz das distâncias genéticas baseadas no F_{ST} (abaixo da diagonal) entre todos os pares de populações comparadas com a população da Ilha de Marajó e os correspondentes valores de probabilidade de não diferenciação p (acima da diagonal) – dados <i>Yfiler Plus</i>	119
Tabela 22 –	Matriz das distâncias genéticas baseadas no F_{ST} (abaixo da diagonal) entre todos os pares de populações da Região Norte do Brasil comparadas com a população da Ilha de Marajó e os correspondentes valores de probabilidade de não diferenciação p (acima da diagonal) - conjunto de 18 marcadores Y-STR.....	120
Tabela 23 –	Frequências dos haplogrupos do mtDNA na população da Ilha de Marajó.....	126
Tabela 24 –	Valores de Diversidade Haplotípica para populações do Brasil e da América do Sul, com base nos dados obtidos pelo sequenciamento da Região Controle completa.....	129
Tabela 25 –	Matriz das distâncias genéticas baseadas no F_{ST} (abaixo da diagonal) entre todos os pares de populações comparadas com a população da Ilha de Marajó e os correspondentes valores de probabilidade de não diferenciação p (acima da diagonal) - Região Controle do mtDNA.....	132

LISTA DE ABREVIATURAS E SIGLAS

AMEL Y	gene da amelogenina
DNA	Ácido desoxirribonucléico
AIMs	<i>Ancestry Informative Markers</i>
BSA	Albumina Bovina Sérica
CAAE	Certificado de Apresentação de Apreciação Ética
CE	Eletroforese capilar
CEPH	<i>Center for the Study of Human Polymorphism</i>
CLUMPP	<i>CLUster Matching and Permutation Program</i>
Co.	<i>Company</i>
CR	Região controle do mtDNA
CRS	<i>Cambridge Reference Sequence</i>
ddNTP	Didesoxinucleotídeo trifosfato
del	Deleção
DIP	<i>Deletion-Insertion Polymorphism</i>
<i>d-loop</i>	<i>Displacement loop</i>
dNTP	Desoxinucleotídeo trifosfato
DTT	Ditiotreitol
EDTA	Ácido etilenodiaminotetracético
EDTANa2	Ácido etilenodiaminotetracético, sal dissódico
EMPOP	<i>EDNAP mtDNA population database</i>
F. A.	Frequência absoluta
F. R.	Frequência relativa
F _{ST}	Índice de Fixação
FTA	<i>Flinders Technology Associates</i>
FTDNA	<i>Family Tree DNA</i>
HD	Diversidade haplotípica
HGDP	<i>Human Genome Diversity Panel</i>
HV-I	Região hipervariável I
HV-II	Região hipervariável II
HV-III	Região hipervariável III

IBGE	Instituto Brasileiro de Geografia e Estatística
Indel	Marcador de inserção-deleção
ISFG	<i>International Society for Forensic Genetics</i>
ISOGG	<i>International Society of Genetic Genealogy</i>
LDD	Laboratório de Diagnósticos por DNA
LGA	Língua geral da Amazônia
MCMC	Markov Chain Monte Carlo
MDS	<i>Multidimensional Scaling</i>
MITOMAP	Base de dados de mtDNA <i>online</i>
MPS	<i>Massive parallel sequencing</i>
MSY	<i>Male-specific Y chromosome region</i>
mtDNA	DNA mitocondrial
n	Número
NevGen	<i>Software</i> de predição de haplogrupos do cromossomo Y
NGS	<i>Next-generation Sequencing</i>
NRY	<i>Non-recombining Region of the Y chromosome</i>
PAR1	Região pseudoautossômica 1 do cromossomo Y
PAR2	Região pseudoautossômica 2 do cromossomo Y
PCA	Análise dos Componentes Principais
PCR	<i>Polimerase Chain Reaction</i>
PD	Poder de discriminação
pH	Potencial hidrogeniônico
rCRS	<i>revised Cambridge Reference Sequence</i>
SBE	<i>Single Base Extension</i>
RFU	Unidade relativa de fluorescência
RM	<i>Rapidly mutating</i>
rRNA	<i>RNA ribossomal</i>
SAP	<i>Shrimp Alkaline Phosphatase</i>
SBE	Single Base Extension
SDS	Dodecil-sulfato de sódio
SNP	<i>Single Nucleotide Polymorfism</i>
SRY	<i>Sex-determining region Y chromosome</i>
SSC	Tampão Citrato de Sódio

STR	<i>Short Tandem Repeats</i>
TAE	Tris-acetato-EDTA
TE	TRis-EDTA
Tris	Trihidroximetil aminometano
tRNA	RNA transportador
UERJ	Universidade do Estado do Rio de Janeiro
YHRD	<i>Y Chromosome Haplotype Reference Database</i>

LISTA DE SÍMBOLOS

Yp	Braço curto do cromossomo Y
Yq	Braço longo do cromossomo Y
=	Igual
mL	Mililitro
-	Menos
°C	Grau Celsius
pb	Par de bases
kb	Mil bases
M	Concentração molar
mg	Miligrama
mM	Milimolar
mm ²	Milímetro quadrado
μL	Microlitro
ng	Nanograma
μM	Micromolar
δ	Frequência alélica diferencial
V/cm	Volts por centímetro
min	Minuto
x g	Gravidade
+	Mais
nm	Nanômetro
<	Menor
s	Segundo
cm	Centímetro
km	Quilômetro
h	Hora
ka	Mil anos
p	Nível de significância
π	Diversidade de sítios nucleotídicos
m	Metro

m2	Metro quadrado
A	Adenina
C	Citosina
G	Guanina
T	Timina
In	Índice de informatividade

SUMÁRIO

	INTRODUÇÃO	18
1	OBJETIVOS	73
1.1	Geral	73
1.2	Específicos	73
2	MATERIAL E MÉTODOS	75
2.1	Amostra Populacional	75
2.2	Extração de DNA genômico	77
2.2.1	<u>Extração de DNA - método orgânico - fenol-clorofórmio/álcool isoamílico</u>	78
2.2.2	<u>Extração de DNA - resina Chelex</u>	79
2.2.3	<u>Extração de DNA - Extração de DNA - <i>Kit QIAamp DNA Investigator</i></u>	80
2.3	Quantificação de DNA	80
2.4	Análise de marcadores do Cromossomo Y	80
2.4.1	<u>Genotipagem dos <i>loci</i> STR do cromossomo Y utilizando o <i>kit Yfiler Plus</i></u>	81
2.4.2	<u>Predição de haplogrupos - <i>software Haplogroup Predictor</i> e <i>NevGen</i></u>	82
2.4.3	<u>Genotipagem dos marcadores Y-SNP - <i>kit SNaPshot Multiplex</i></u>	84
2.4.4	<u>Análise estatística dos resultados do Cromossomo Y</u>	92
2.5	Análise de marcadores do DNA mitocondrial	94
2.5.1	<u>Genotipagem da Região Controle do mtDNA</u>	94
2.5.2	<u>Classificação dos haplogrupos de mtDNA e análise estatística</u>	97
2.6	Análise dos AIMs	98
2.6.1	<u>Genotipagem dos 46 AIM-Indels</u>	99
2.6.2	<u>Análise estatística dos resultados dos AIM-Indels</u>	100
3	RESULTADOS E DISCUSSÃO	102
3.1	Características da população	102
3.1.1	<u>Distribuição amostral em relação ao local de nascimento do ancestral da terceira geração para cada categoria de marcador</u>	102
3.2	Quantificação do DNA	104
3.3	Análise de marcadores do Cromossomo Y	104
3.3.1	<u>Haplótipos do Cromossomo Y - <i>kit Yfiler Plus</i></u>	104
3.3.2	<u>Predição de haplogrupos e genotipagem de Y-SNPs</u>	107

3.3.3	<u>Análise estatística dos resultados do Cromossomo Y – Y-STRs</u>	114
3.3.3.1	Diversidade Haplotípica	114
3.3.3.2	Distância Genética (F_{ST})	116
3.4	Análise de marcadores do mtDNA	122
3.4.1	<u>Haplótipos da Região Controle do mtDNA</u>	122
3.4.2	<u>Haplogrupos de mtDNA na população da Ilha de Marajó</u>	124
3.4.3	<u>Análise estatística – Região Controle do mtDNA</u>	127
3.4.3.1	Diversidade Haplotípica	127
3.4.3.2	Distância Genética (F_{ST})	128
3.5	Análise dos marcadores autossômicos (AIMs)	133
3.5.1	<u>Genótipos - multiplex de 46 AIM–Indels e inferência de ancestralidade</u>	133
3.5.2	<u>Comparação das proporções de ancestralidade continental – marcadores autossômicos e os marcadores de linhagem</u>	135
	CONCLUSÕES	141
	REFERÊNCIAS	143
	APÊNDICE A – Haplótipos <i>Yfiler Plus</i> de indivíduos da Ilha de Marajó.....	159
	APÊNDICE B – Haplogrupos e Haplótipos da Região Controle do mtDNA de indivíduos da Ilha de Marajó e sua origem continental.....	164
	APÊNDICE C – Genótipos dos 46 AIM-Indels de indivíduos da Ilha de Marajó e as proporções de ancestralidade continental.....	170
	APÊNDICE D – Pôster exibido no Congresso <i>Haploid Markers Conference</i> - Polônia, maio/2018	184
	APÊNDICE E – <i>Proceedings</i> do 28º Congresso - Sociedade Internacional de Genética Forense 2019 – Praga, República Tcheca.....	185
	APÊNDICE F – Artigo publicado na revista <i>Forensic Science International: Genetics</i>	187
	ANEXO A - Termo de consentimento livre e esclarecido	195
	ANEXO B - Parecer do Comitê de Ética em Pesquisa do Hospital Pedro Ernesto	197

INTRODUÇÃO

A genética populacional e sua evolução

A genética populacional ou genética de populações é uma área da ciência fundamentada pelos conceitos de genética mendeliana que estuda a distribuição dos alelos e haplótipos nas populações e como suas frequências são mantidas ou alteradas ao longo das gerações, tanto no contexto de uma espécie quanto entre espécies. Engloba o estudo das variações genéticas que ocorrem naturalmente entre os organismos, e visa compreender os processos genético-demográficos atuantes (HARTL, 2008).

Os haplótipos correspondem a uma combinação de grupos de alelos de *loci* adjacentes, podendo ser formado por um ou mais alelos, ou até por um cromossomo inteiro.

A maior mudança no estudo da evolução pela genética nos últimos 50 anos foi a introdução de ferramentas moleculares para estudar as variações nas espécies, em vez de depender de variantes visíveis. A partir daí a genética de populações em nível molecular continua em crescente evolução (CHARLESWORTH; CHARLESWORTH, 2017). Em seus estágios iniciais, estudos pioneiros basearam-se em diferenças detectáveis por eletroforese, de variantes de proteínas solúveis, codificadas por diferentes alelos de um mesmo *locus* - as aloenzimas. Esses estudos fizeram uso da descoberta, feita apenas alguns anos antes, de que os genes codificam polipeptídeos, de modo que a variação nas sequências de proteínas detectadas pela variação em sua mobilidade em um campo elétrico pode ser equiparada à variação no próprio gene. As investigações foram realizadas tanto em humanos como também em outras espécies (HARRIS, 1966; LEWONTIN; HUBBY, 1966).

No entanto, já eram percebidas as limitações da técnica de eletroforese de proteínas para medir variações genéticas, por sua incapacidade de detectar duas categorias de variantes: alterações nas moléculas de proteínas que não afetam sua mobilidade em um gel e variantes na sequência de DNA que deixam a sequência protéica inalterada - os polimorfismos silenciosos ou sinônimos. Tais limitações, assim como o estudo de polimorfismos em regiões não-codificantes, só puderam ser superadas através de estudos da variação na sequência do DNA, que inicialmente foram realizados com enzimas de restrição. O mapeamento de sítios de restrição foi o passo inicial para o desenvolvimento de parâmetros estatísticos relacionados

à diversidade genética nas sequências de DNA, como a diversidade de sítios nucleotídicos (π) (NEI; LI, 1979).

Na década de 1980, diversos estudos forneceram os primeiros *insights* sobre padrões de variação nas sequências de DNA em populações de drosófilas, assim como a correlação positiva entre o nível de variabilidade encontrado e a taxa de recombinação. A recombinação é um recurso muito importante que resulta em maior diversidade, pois em cada geração permite a reorganização da informação genética, criando novas e únicas configurações (AGUADÉ; MIYASHITA; LANGLEY, 1989; BEGUN; AQUADRO, 1992). O primeiro estudo de genética populacional com o sequenciamento do DNA foi realizado com técnicas manuais de sequenciamento de Sanger (KREITMAN, 1983; SANGER; NICKLEN; COULSON, 1977), em populações de *D. melanogaster*. Esse estudo foi pioneiro, detectando diferentes tipos de variantes genéticas e ampliando o espectro que se tinha até o momento, revelando polimorfismos de base única ou SNPs (do inglês, *Single Nucleotide Polymorphisms*) silenciosos e fora de regiões codificantes, assim como polimorfismos do tipo inserção/deleção e sequências de homopolinucleotídeos.

Com o surgimento da técnica de PCR (do inglês, *Polymerase Chain Reaction*) e do sequenciamento automático no início da década de 1980, que possibilitaram a amplificação e sequenciamento de regiões específicas do genoma, estudos em larga escala evidenciaram inúmeras variações envolvendo um grande número de genes e espécies; sendo possível estimar novos níveis de diversidade genética. Tais estudos trouxeram ainda muitos outros dados, como diferenças dos níveis de variação entre diferentes *taxa*, além de mostrarem que os SNPs superam em número todos os outros tipos de variantes no DNA (LEFFLER *et al.*, 2012; ROMIGUIER *et al.*, 2014). A essa altura, devido à necessidade de compilação e de análises em larga escala dos dados obtidos a partir da grande quantidade de sequências geradas ao longo de décadas, já se observava o surgimento de bases de dados públicas, como o GenBank, que engloba sequências de nucleotídeos de mais de 300 mil espécies (BENSON *et al.*, 2015); assim como o desenvolvimento de *software* que possibilitam o uso de diversas ferramentas estatísticas já disponíveis (CASILLAS; BARBADILLA, 2017).

Apesar da abundância de dados gerados sobre a diversidade genética, o conhecimento até esse ponto ainda se encontrava restrito a regiões do genoma já estudadas, e não à sua totalidade, o que possibilitaria uma medição mais completa e exata dos parâmetros avaliados. Sendo assim, avanços ainda maiores têm sido feitos visando à caracterização global da variação genética através do sequenciamento de genomas completos, e pode-se dizer que estamos atualmente na era da genômica populacional. Os primeiros trabalhos com

sequenciamento e análise de genomas completos surgiram durante a década passada, com o desenvolvimento das tecnologias de sequenciamento de nova geração ou NGS (do inglês, *Next Generation Sequencing*); também conhecido por MPS (do inglês, *Massive Parallel Sequencing*) (GOODWIN; MCPHERSON; MCCOMBIE, 2016; METZKER, 2010). Tais estudos, antes inviáveis economicamente pelo sequenciamento automatizado de Sanger, iniciaram-se com abordagens em larga escala de polimorfismos em múltiplos *loci* (BLACK *et al.*, 2001; LUIKART *et al.*, 2003), chegando a genomas completos de milhares de indivíduos de diversas espécies, incluindo a espécie humana (THE 1000 GENOMES PROJECT CONSORTIUM, 2010; THE 1000 GENOMES PROJECT CONSORTIUM, 2015; THE 1001 GENOMES PROJECT CONSORTIUM, 2016; BEGUN *et al.*, 2007; FAWCETT *et al.*, 2014). Esse alcance se tornou possível devido ao fato de que o MPS fornece dados genômicos populacionais a custos consideravelmente mais baixos, uma vez que permite sequenciar o genoma de grupos de indivíduos (*pool-seq*), e não somente indivíduos separados, como a técnica de Sanger (SCHLÖTTERER *et al.*, 2014). Paralelamente à quantidade crescente de dados genômicos populacionais, estimativas cada vez mais precisas das taxas de recombinação ao longo do genoma também estão sendo obtidas, possibilitando inferências mais precisas sobre sua relevância na evolução do genoma (CASILLAS; BARBADILLA, 2017).

As informações provenientes de MPS diferem consideravelmente dos dados de variação obtidos na era das aloenzimas e sequências de Sanger, principalmente no volume desses dados. Sendo assim, nesse novo patamar de quantidade de sequências geradas, a Bioinformática tornou-se essencial para atender às necessidades específicas de todas as etapas do processo, desde a aquisição de dados, verificação de qualidade destes e sua análise e representação. Essa revolução genômica das últimas duas décadas e seu impacto sobre a genética populacional vem enriquecendo essa área com muitos novos conceitos, técnicas moleculares, formatos de dados e métodos estatísticos e computacionais. Sendo assim, modelos matemáticos aplicados na interpretação dos dados também foram evoluindo, primeiramente com estudos que demonstravam que a seleção natural seria uma potente força evolutiva, levando à variação nas espécies através da observação de polimorfismos “visíveis” (ALLISON, 1964). Essa observação passa pela formulação das principais teorias sobre a evolução e a variabilidade das espécies, como o modelo da hipótese “clássica” e o modelo da hipótese “balanceada”, que compõem a teoria selecionista. A diferença fundamental entre esses modelos reside no fato de na via clássica acreditar-se que a seleção agiria sobre genes que apresentavam apenas dois tipos de alelos - alelos funcionais (“selvagens”) e alelos

mutantes, estes considerados deletérios, presentes em frequências muito baixas (MULLER, 1950); enquanto na via balanceada acredita-se que muitos genes poderiam possuir dois ou mais alelos alternativos, com frequências intermediárias mantidas nas populações, por seleção balanceada (DOBZHANSKY, 1955). Alguns anos depois, foi demonstrado que além da seleção natural, e independente desta, a deriva genética aleatória, assim como as mutações neutras, ou seja, mutações cujos efeitos não influenciam a aptidão dos indivíduos, seriam responsáveis por grande parte da variabilidade nas populações, surgindo então a teoria neutralista (KIMURA; CROW, 1964).

A deriva genética pode ser definida como uma amostragem aleatória de gametas em cada geração de uma população finita, o que resulta em uma flutuação das frequências alélicas através das gerações e diminuição de variação genética. Esse fenômeno pode ser observado, por exemplo, quando uma população tem o tamanho drasticamente reduzido por um desastre natural (efeito gargalo) ou quando um pequeno grupo se separa da população principal para fundar uma nova colônia (efeito fundador), podendo resultar em perda de alguns alelos e fixação de outros, independente do caráter adaptativo dos mesmos. Em uma população panmítica idealizada, a força da deriva genética é inversamente proporcional ao tamanho populacional (BAMSHAD *et al.*, 2004).

Passado alguns anos, propôs-se um refinamento da teoria neutralista, surgindo então a teoria “quase neutra”. Nela se considerou que a maioria das mutações é levemente deletéria ou levemente benéfica, havendo um gradiente nos coeficientes de seleção que atuam sobre as mesmas (OHTA, 1973). Posteriormente, foi desenvolvido, a partir da teoria “quase neutra”, um modelo que leva em conta a atuação balanceada tanto do coeficiente de seleção quanto da deriva genética sobre as mutações (OHTA; GILLEPIE, 1996). Esse modelo postula que mutações neutras são afetadas basicamente pela deriva genética, mas não pela seleção natural e que as mutações fortemente deletérias ou benéficas seriam mais afetadas pela seleção natural e não pela deriva. Já as mutações quase neutras seriam fixadas ou não em uma população, dependendo do balanço entre o coeficiente de seleção e o tamanho da população; quanto menor a população, maior o efeito da deriva e quanto maior a população, maior o efeito da seleção natural.

Os primeiros modelos teóricos na genética de populações aqui citados simularam a evolução ao longo do tempo, tentando entender como uma população evoluirá de um tempo passado ou presente para o futuro. Já a teoria da coalescência, surgida posteriormente, segue uma abordagem diferente, centrada na genealogia de uma amostra populacional, utilizando uma modelagem dos processos mutacionais no sentido presente → passado, ou seja, eventos

coalescentes são representados como uma genealogia dos genes (KINGMAN, 1982a, 1982b, 2000). Através do rastreamento de uma amostra atual de indivíduos de uma população, a fim de detectar os alelos de seus genes compartilhados por todos os seus membros, chega-se até uma única cópia ancestral, conhecida como o ancestral comum mais recente ou MRCA (do inglês, *Most Recent Common Ancestor*). Após alguns refinamentos, esse modelo retrospectivo simplificou consideravelmente as análises baseadas em modelos neutros, permitindo relacionar a diversidade genética observada em uma amostra com a história demográfica da população da qual foi extraída (CASILLAS; BARBADILLA, 2017).

O principal objetivo da genômica populacional é a descrição e interpretação da variação genética dentro e entre populações, mas podemos dizer que as atuais abordagens tecnológicas revolucionaram esse campo, impulsionando tanto o trabalho empírico quanto suas fundamentações teóricas. Ainda assim, as forças fundamentais da evolução permanecem como os fatores explicativos essenciais (CHARLESWORTH, 2010; LYNCH, 2007). As mutações, consideradas a fonte primária de variabilidade genética nas populações, ocorrem espontaneamente ao longo de todo o genoma. Mas por apresentarem taxas muito baixas, não respondem pelas mudanças rápidas no *genepool* de uma população e pelo nível de variabilidade observado nas populações. Sendo assim, a diversidade genética introduzida por mutações é posteriormente moldada e distribuída por outros processos, tais como deriva gênica, migrações, seleção natural, recombinação, e fluxo gênico. Em conjunto, tais processos, que envolvem a fundação e dispersão de novas populações, e favorecem os cruzamentos entre indivíduos de um mesmo grupo em detrimento a grupos separados geograficamente, levaram à distribuição de diferentes polimorfismos e polimorfismos com diferentes frequências entre populações. Tais diferenças contribuíram para a variabilidade genética que se observa atualmente e na sua distribuição ao nível mundial (AMORIM; BUDOWLE, 2016; BAMSHAD *et al.*, 2004; HOLSINGER; WEIR, 2009).

Sabe-se que, em média, dois humanos selecionados ao acaso diferem em aproximadamente três milhões de nucleotídeos, e que a maioria desses polimorfismos é neutra ou quase neutra. A distribuição desses polimorfismos, associada ao conhecimento dos padrões de recombinação e das taxas de mutação, ao longo do genoma, refletem a história demográfica de nossa espécie, permitindo o rastreamento a ancestrais que podem ter vivido em diferentes épocas e partes do mundo, visto que as informações contidas na molécula de DNA são um tipo de registro da história da espécie humana. Desta forma, muitos estudos têm sido realizados com populações atuais, trazendo inferências sobre aspectos histórico-demográficos e evolutivos dos diferentes grupos humanos, corroborando com os registros

históricos já existentes e contribuindo com novas informações. Nomeadamente, locais de origem e rotas de migração ocorridas ao longo do tempo e a compreensão de seu impacto na complexa diversidade mundial contemporânea (CAVALLI-SFORZA, 1998; HARPENDING; ROGERS, 2000; JOBLING; HURLES; TYLER-SMITH, 2003).

As migrações humanas e a origem da população moderna

As migrações consistem no deslocamento de indivíduos de uma área geográfica para outra, que, se já for habitada por indivíduos de um grupo distinto, poderá resultar em um processo de miscigenação e conseqüente formação de uma população com uma composição genética distinta (dita miscigenada), uma vez que haverá fluxo gênico entre os grupos populacionais envolvidos (JOBLING; HURLES; TYLER-SMITH, 2003).

Atualmente, a teoria mais aceita para a formação da população moderna indica que a origem dos primeiros grupos populacionais humanos se deu na África, há mais de 100 ka (BAMSHAD *et al.*, 2004; FORSTER, 2004; PENA *et al.*, 2009). Estudos realizados com o DNA mitocondrial (mtDNA) demonstram que as linhagens maternas de todos os humanos originam-se na “Eva mitocondrial”, uma mulher nascida há mais de 130 ka no sul ou leste africanos. Evidências genéticas e arqueológicas revelam que há cerca de 100 ka ocorreu inicialmente uma modesta expansão humana em toda a África, seguida de uma grande reexpansão, entre 60-80 ka atrás, na qual os humanos se dispersaram da África para colonizar outras partes do mundo. Provavelmente essa expansão foi possível por ter coincidido com uma diminuição do nível do mar, nunca ocorrida, e que abriu uma rota através do mar vermelho para o Yemen. Acredita-se que o grupo que saiu da África, nessa que ficou estabelecida como a primeira migração moderna bem-sucedida “Out of África”, era de tamanho reduzido. Tal conclusão advém do fato de apenas um tipo de linhagem do mtDNA, dentre as linhagens existentes àquela época, ser encontrada fora da África (linhagens L3), da qual todos os não-africanos descendem atualmente. Sendo assim, os primeiros europeus e asiáticos foram originados desse pequeno grupo de migrantes, que se dividiram em duas rotas principais: as rotas costeiras para a Austrália e Papua Nova Guiné, sendo as mais rápidas, em comparação com as rotas com condições climáticas mais severas no Norte, que levavam à Eurásia. Algumas linhagens de mtDNA, que se acredita terem saído da África ou surgido imediatamente após a migração, por exemplo (linhagens M), estão ausentes na população

européia atual, provavelmente devido a efeitos de deriva mais severos nessas rotas. Com a estabilização das condições climáticas da última Idade do gelo por volta de 30 ka atrás, as populações de *Homo sapiens* locais cresceram e expandiram intensamente, pela Europa e norte da Ásia, havendo o surgimento de linhagens descendentes fundadoras, conhecidas como haplogrupos, e que atualmente tem distribuição continental ou regional. A partir da Ásia mais uma etapa da expansão humana acontece, levando ao povoamento das Américas, há aproximadamente 25 ka atrás. A teoria mais aceita para o povoamento das Américas relata que o nível do mar era cerca de 120 m inferior ao que é hoje, e a América do Norte e o continente Asiático estavam conectados pela ampla ponte terrestre da Beringia, atualmente submersa. Registros antropológicos e análises genéticas indicam que, através dessa passagem, um pequeno grupo de asiáticos do norte da Sibéria migrou para o continente americano (BONATTO; SALZANO, 1997; FORSTER, 2004; RAY *et al.*, 2010). Durante o último período glacial, entretanto, alguns milênios após a chegada desses fundadores, a grande extensão dos mantos de gelo glacial comprometeu a passagem da Beringia, interrompendo as migrações da Ásia em direção à América do Norte. Embora haja um consenso sobre a origem asiática das linhagens fundadoras na colonização das Américas, existem diferentes hipóteses, baseadas em achados arqueológicos, linguísticos e dados genéticos obtidos de marcadores diversos, sobre o tempo que os fundadores levaram para povoar o continente, assim como sobre o número de ondas migratórias que ocorreram. A maioria dos estudos aponta para uma única migração (BONATTO; SALZANO, 1997; FORSTER *et al.*, 1996; MERRIWETHER; ROTHHAMMER; FERREL, 1994), enquanto outros sugerem que uma única e discreta onda de colonização é altamente inconsistente com os níveis observados de diversidade genética, chegando a sugerir três ondas, além da possibilidade de fluxo gênico recorrente entre a Ásia e a América, após a colonização inicial (GREENBERG; TURNER II; ZEGURA, 1986; RAY *et al.*, 2010; REICH *et al.*, 2012; TAMM *et al.*, 2007). Uma das hipóteses sobre o povoamento do continente, chamada de Modelo de Incubação na Beringia, considera que antes de povoarem o continente como um todo, os fundadores se estabeleceram na Beringia por aproximadamente 15 ka, isolados por uma barreira geográfica de gelo. Daí surgindo então, uma diversidade que não se observava na população fundadora, antes mesmo de espalharem-se por todo o continente americano. A outra hipótese, chamada Modelo de Colonização Direta, propõe que uma parte dos ancestrais sobreviveu na Beringia, ao ficarem isolados, pelo gelo, da maior parte dos imigrantes, que se moveu mais rapidamente ao longo do continente, do Norte para o Sul. Mais tarde, a população da Beringia se diferenciaria nos grupos linguísticos “Eskimo-Aleut” e “Na-Dene”, que ficaram restritos à América do Norte; e os que

se dispersaram em direção à América Central e América do Sul originaram a família lingüística dos ameríndios. No decorrer dos milênios seguintes, a costa atlântica do continente americano, assim como o interior foram sendo percorridos e povoados por inúmeros povos ameríndios ou indígenas, que estavam em constante migração e adaptação (BONATTO; SALZANO, 1997; FORSTER, 2004; RAY *et al.*, 2010; RIBEIRO, 1995; TAMM *et al.*, 2007).

A maioria das populações atuais da América Latina, incluindo a população brasileira, apresenta peculiaridades em relação à sua composição genética, devido ao fato de terem sido alvo de vários fluxos migratórios. Estima-se que cerca de quarenta e cinco milhões de nativos habitavam na América Latina à época do início da vinda de europeus, há cerca de cinco séculos, embora haja projeções que variam de trinta a noventa milhões de ameríndios. Porém houve um rápido decréscimo dessa população e até mesmo a extinção de grupos menores, devido a epidemias e massacres ocorridos na chegada dos invasores (SALZANO; BORTOLINI, 2002; SALZANO; SANS, 2014). As estimativas sobre o número de europeus são muito variáveis, de cerca de quatro a cinquenta milhões de pessoas, em várias ondas, sendo a última delas relativamente recente, após a Segunda Guerra Mundial. Inicialmente vieram principalmente espanhóis e portugueses, entre os séculos XV e XIX, e durante esse período, tais conquistadores europeus trouxeram de diferentes regiões da África subsaariana mais de nove milhões de escravos para trabalhar em suas lavouras, com diversas viagens, sendo cerca de quatro milhões destes apenas para o Brasil. Nos séculos mais recentes alguns países também receberam populações de origem asiática, como os imigrantes japoneses que vieram para São Paulo e Pará, no Brasil, ou as populações de origem chinesa, no Peru. Conseqüentemente, com tantas contribuições para a diversidade latino americana, pelas diferentes migrações e pelo fluxo gênico envolvendo diferentes ancestralidades, observa-se atualmente uma grande heterogeneidade da ancestralidade entre os países, assim como dentro destes. Porém, pode-se afirmar que essas populações híbridas formaram-se majoritariamente pela contribuição de três componentes genéticos continentais: nativo americano ou ameríndio, europeu e africano. Sendo assim, na América Latina como um todo, e em populações miscigenadas espalhadas pelo mundo, sabe-se que, como reflexo de sua história demográfica, os cromossomos das pessoas que integram essas populações mostram uma complexa mistura de ascendências. Mistura tal que depende tanto do número de gerações de cruzamentos como do padrão de cruzamentos que acontecem, uma vez que as pessoas pertencem a diferentes extratos socioeconômicos (RUIZ-LINARES *et al.*, 2014; SALZANO; SANS, 2014).

Marcadores genéticos em estudos forenses e populacionais

O genoma humano, isto é, a informação genética contida no interior das células e transmitida ao longo das gerações, compreende o genoma nuclear, que contém a maior parte da informação genética, e o genoma mitocondrial (mtDNA), cada um dos quais com características distintas. O genoma nuclear, tal como o nome indica, está localizado no interior do núcleo das células dos seres eucariontes. É organizado em cromossomos, sendo que cada célula é composta por 22 pares de autossomos e dois cromossomos sexuais. O cromossomo sexual X está presente em duas cópias nas mulheres (XX) e apenas uma cópia nos homens, que apresentam também um cromossomo Y (XY). Já o mtDNA encontra-se no interior das mitocôndrias, organelas citoplasmáticas responsáveis pela produção de energia, localizadas em células animais eucarióticas. Cada mitocôndria tem de centenas a milhares de cópias do mtDNA, o que representa um pequeno percentual do conteúdo genômico total de uma célula (BUTLER, 2005).

Os polimorfismos são variações na sequência de DNA entre indivíduos, ao nível de um *locus*, ou seja, o local fixo em um cromossomo onde está localizado determinado gene ou marcador genético. Sendo assim, um *locus* é dito polimórfico sempre que nele existirem pelo menos duas variantes (alelos) com frequências superiores a 1 % na população analisada (GELEHRTER; COLLINS, 1990). Os *loci* polimórficos são considerados úteis em diversas aplicações, como estudos populacionais, forenses e genealógicos, visto que possibilitam a diferenciação entre indivíduos e a identificação humana, sendo nomeados de marcadores genéticos (BUTLER, 2005).

Qualquer região do genoma, seja ela codificante ou não-codificante, pode ser rastreada para variantes genéticas, mas na prática, para estudos genéticos populacionais e forenses, marcadores não codificantes são preferíveis, uma vez que não estão diretamente sujeitos aos efeitos da seleção e, portanto, espera-se que reflitam principalmente efeitos neutros em nível populacional, como deriva, expansões, miscigenação e migração (STEIPER, 2010). A escolha de marcadores neutros no campo forense também é recomendada para evitar implicações éticas, uma vez que são menos propensos a divulgar informações associadas a doenças ou suscetibilidade genética (SCHNEIDER, 1997). No entanto, este critério vem sendo reconsiderado com o advento das tecnologias de maior desempenho, como sistemas MPS, que permitem a coleta de informações de grandes porções do genoma. Com o crescente volume de dados genéticos disponíveis, marcadores de ancestralidade, assim como de características

fenotípicas, estão cada vez mais sendo usados em investigações forenses, para obter informações sobre potenciais criminosos sem suspeitos identificados (KAYSER; DE KNIJFF, 2011; PHILLIPS *et al.*, 2014a).

Os marcadores genéticos podem apresentar diferenças entre indivíduos tanto na sequência de nucleotídeos, assim como no comprimento de segmentos de DNA. Os polimorfismos de sequência são conhecidos como SNPs, e os polimorfismos de comprimento, que incluem os marcadores de inserção-deleção bialélicos ou Indels, assim como marcadores multialélicos, como os STRs (do inglês, *Short Tandem Repeats*) e outros tipos de variações. Alguns sistemas de classificação consideram os STRs como marcadores indels multialélicos (BUTLER, 2005; WEBER *et al.*, 2002).

A seleção de marcadores a serem utilizados, no contexto forense ou da genética populacional, dependerá da demanda requerida, uma vez que diferentes cenários podem ocorrer. Em cada caso, a seleção dos marcadores genéticos apropriados está, em última análise, associado à sua variação intrapopulacional e /ou interpopulacional, assim como ao seu modo de transmissão e também a aspectos técnicos, que são muito relacionados às características das amostras biológicas disponíveis (AMORIM; BUDOWLE, 2016). Marcadores com altas taxas de mutação são muito polimórficos e úteis em genética populacional, para estudar processos recentes; e na genética forense, para identificação humana. Por outro lado, Para o estudo de eventos passados, marcadores de evolução lenta são a escolha mais adequada (SCHLOTTERER, 2000; STEIPER, 2010). Aproveitando sua abundância no genoma, bem como a facilidade de detecção, os marcadores de mutação lenta também podem ser usados em análises forenses para inferir a origem étnica de um indivíduo, por exemplo, uma vez que uma menor recorrência tende a produzir variações de frequência de alelos que refletem uma distribuição geográfica estruturada. Neste trabalho foram usados marcadores genéticos STRs e os Indels, que são polimorfismos de comprimentos; além dos SNPs, que são polimorfismos de variação de sequência.

Marcadores STR - *Short Tandem Repeats*

Dentre os principais marcadores genéticos usados nas últimas décadas estão os marcadores do tipo STR, também conhecidos como microssatélites, que estão entre os polimorfismos mais abundantes do genoma humano, ocorrendo ao longo de todos os

cromossomos a uma razão de 1 a cada 15000 pares de bases (pb), sobretudo em regiões não codificantes. Os STRs são trechos da molécula de DNA compostos de repetições consecutivas (*in tandem*), formadas por unidades de 2 a 6 pb (Figura 1). São caracterizados por variações no comprimento da sequência de DNA devidas a diferenças no número de repetições (ELLEGREN, 2004).

As diferentes variantes alélicas de uma região microssatélite são nomeadas conforme o número de unidades de repetição que formam a sequência. Desta maneira, se um STR é constituído por cinco unidades de repetição, o mesmo apresenta o alelo 5. É importante ressaltar que nem todos os alelos são formados por unidades de repetições completas, sendo nesse caso denominados alelos microvariantes. De acordo com sua estrutura, os STRs podem ser classificados em simples, compostos ou complexos. Os STRs de estrutura simples contêm um único tipo de sequência repetitiva, com sequências e comprimentos idênticos. Os compostos possuem mais de um tipo de sequência repetitiva, enquanto os complexos podem conter diversos blocos repetitivos, de comprimento variável, intermeados por sequências mais ou menos variáveis (URQUHART *et al*, 1994).

Sabe-se que STRs são gerados por mutações aleatórias e que o ganho ou a perda de unidades de repetição ocorrem devido a deslizamentos da enzima polimerase durante a replicação de DNA. As taxas médias de mutação dos STRs são de cerca de 2×10^{-3} (BRINKMANN *et al.*, 1998; SCHLOTTERER, 2000), o que faz com que sejam altamente polimórficos quando comparados a outros tipos de polimorfismos, como os Indels e os SNPs. Devido a isso apresentam alelos múltiplos, além de alta heterozigosidade. A complexidade estrutural do STR influencia sua taxa de mutação, sendo os STRs com repetições com 4 nucleotídeos considerados bastante estáveis e por isso, bastante abundantes. Por isso têm sido preferidos ao projetar os *kits* forenses disponíveis comercialmente, embora repetições maiores apresentem a vantagem de apresentar com menor frequência artefatos de análise do tipo picos *stutter*. Esses artefatos ocorrem *in vitro*, durante a amplificação do fragmento de DNA por PCR, pelo mesmo mecanismo de deslizamento da enzima polimerase. De fato, são produtos da PCR presentes em quantidades menores do que o alelo verdadeiro e diferem em tamanho do produto principal por múltiplos do comprimento da unidade de repetição. Esses artefatos de PCR às vezes podem complicar a análise de STRs, especialmente em situações de pouco DNA molde ou baixa qualidade das amostras, bem como em misturas não balanceadas de dois ou mais contribuintes, mas na maioria das situações eles podem ser facilmente identificados (AMORIM; BUDOWLE, 2016).

Em relação à metodologia de análise, os STRs podem ser analisados a partir de fragmentos de PCR relativamente curtos (100 a 400 pb), sendo possível a amplificação de vários marcadores em uma única reação de PCR multiplex. Os *primers* ou iniciadores referentes a cada região de interesse são marcados com diferentes fluoróforos e os fragmentos gerados são separados de acordo com seu tamanho, durante a eletroforese capilar. Essa simplicidade da técnica de genotipagem, se comparada com outras metodologias, juntamente com a abundância dos STRs no genoma humano e com os altos níveis de diversidade intrapopulacional esperados para esses *loci* de mutação rápida, tornaram esses marcadores amplamente utilizados em laboratórios forenses em todo o mundo. A análise de STRs tornou-se o principal método de escolha para avaliar a diversidade genética em populações ou para realizar a identificação individual, através de uma diversidade de *kits* comerciais disponíveis para tipagem nos autossomos e nos cromossomos sexuais, e bancos de dados de referência têm sido criados e atualizados, permitindo a comparação entre laboratórios (AMORIM; BUDOWLE, 2016; BUDOWLE; DAAL, 2008).

Figura 1 - Exemplos de diferentes tipos de marcadores genéticos: SNP, Indel e STR



Fonte: AMORIM; BUDOWLE, 2016.

Marcadores SNP – Single Nucleotide Polymorphisms

Os marcadores SNPs são substituições de um único nucleotídeo que ocorrem em diversas posições no genoma (Figura 1). Alguns autores incluem inserções ou deleções de base única na classe dos SNPs (BUDOWLE; DAAL, 2008), enquanto outros as consideram como polimorfismos do tipo indels (AMORIM; BUDOWLE, 2016).

Os SNPs são considerados os polimorfismos mais abundantes do genoma, representando cerca de 85 % da variação genética humana. O número de SNPs já reportados em humanos é de aproximadamente 38 milhões, o que corresponde a uma densidade de um SNP a cada 80 pb, em média (THE 1000 GENOMES PROJECT CONSORTIUM, 2012).

Teoricamente, qualquer um dos quatro nucleotídeos existentes pode ocorrer em cada posição da sequência de DNA, embora a grande maioria dos SNPs seja um marcador bialélico. Isso pode ser explicado devido à baixa taxa de mutação desses marcadores, que torna a probabilidade de duas ou mais mutações independentes ocorrerem em uma mesma posição ser muito baixa. Ainda assim, SNPs tri-alélicos e tetra-alélicos já foram relatados, e são muito úteis na detecção de misturas em análises forenses (VIGNAL *et al.*, 2002; WESTEN *et al.*, 2009).

A taxa de mutação dos SNPs é da ordem de 10^{-8} (NACHMAN; CROWELL, 2000), sendo, por isso, considerados bastante estáveis quando comparados aos STRs e muito úteis também em análises de parentesco e na identificação de indivíduos. Mas apesar de serem mais abundantes, o que também seria uma vantagem, os SNPs não são mais utilizados do que os STRs, e isso se deve em parte ao seu baixo poder de discriminação (PD) por *locus*. O fato dos SNPs, em sua grande maioria, serem bialélicos, diminui seu nível de informatividade isoladamente, sendo necessários aproximadamente 50 SNPs para atingir um valor de PD de 12 a 15 STRs (AMORIM; PEREIRA, 2005). Outra limitação para essas aplicações é que não existem bancos de dados populacionais disponíveis para SNPs como vemos para os STRs (KAYSER; DE KNIJFF, 2011).

Dependendo da localização no genoma e das diferenças nas frequências alélicas entre as populações humanas, os SNPs podem estar associados com populações específicas ou correlacionados com características físicas, como cor de olhos, de pele e de cabelo, por exemplo. Daí esses marcadores serem usados também para inferências fenotípicas e de ancestralidade (BUDOWLE; DAAL, 2008). Na área forense, análises de DNA que possam apontar a possível origem ancestral, assim como características físicas visíveis de um criminoso totalmente desconhecido na investigação, podem ser de extrema importância na solução de um caso. Sendo assim, painéis de SNPs selecionados de acordo com finalidades específicas (identificação, inferência fenotípica, inferência de ancestralidade) vêm sendo desenvolvidos e dados populacionais têm sido gerados a partir desses, em diversas populações (AMORIM; BUDOWLE, 2016; GETTINGS *et al.*, 2014).

Em relação às técnicas usadas para genotipagem de SNPs, uma vantagem é que podem ser analisados com o mesmo tipo de equipamento comumente usado para os STRs – a eletroforese capilar. Além disso, muitas das dificuldades encontradas no uso de STRs tradicionais não existem para SNPs, como a análise de amostras muito degradadas. Para esse tipo de amostra, o uso de amplicons de PCR mais curtos, de 50 pb ou menos, por exemplo, é essencial para a obtenção de um resultado bem-sucedido. Fragmentos menores, no entanto,

não são possíveis com STRs altamente polimórficos devido à sua sequência repetitiva, mas podem ser empregados com os SNPs, que têm alterações de uma única posição. Nos SNPs também não ocorrem os artefatos de análise do tipo *stutter*, comuns nos STRs (KAYSER; DE KNIJFF, 2011).

Marcadores Indel – Inserção / Deleção

Indels são polimorfismos de comprimento caracterizados pela inserção ou deleção de um ou mais nucleotídeos (Figura 1) e em sua grande maioria são bialélicos, com comprimentos variando entre 3 a 15 pb; embora já tenha se descrito indels de 1 a milhares de pares de bases. Os indels, assim como os STRs e os SNPs, estão entre os polimorfismos mais frequentes do genoma, apresentando uma frequência intermediária entre eles, de 1 a cada 7200 pb e ocorrendo nos 24 cromossomos. Aproximadamente 36 % dos indels estão localizados em regiões promotoras, assim como em introns e exons, podendo causar alguma alteração deletéria na função gênica (MILLS *et al.*, 2006; MULLANEY *et al.*, 2010; WEBER *et al.*, 2002), como por exemplo a alteração encontrada na fibrose cística (COLLINS *et al.*, 1987). Sendo assim, os indels, assim como os SNPs, são polimorfismos de potencial interesse na medicina personalizada, uma vez que podem estar associados a alterações fenotípicas (MULLANEY *et al.*, 2010). Por apresentarem baixas taxas de mutação, assim como os SNPs, os indels também podem ser úteis para inferências de ancestralidade e estudos de linhagem, dependendo de sua localização no genoma.

Embora apresentem semelhanças com os SNPs, como por exemplo, a possibilidade de serem genotipados com eficácia em amostras degradadas ou antigas, uma vez que necessitam de amplicons menores que os STRs; os indels são analisados por eletroforese capilar com base apenas no tamanho, sem a necessidade de métodos de sequenciamento direto. Ou seja, combinam as vantagens analíticas desses outros tipos de marcadores, o que tem feito com seu uso esteja aumentando em estudos populacionais e forenses (PEREIRA *et al.*, 2009; 2012a; 2012b).

Em contraste com os SNPs, que têm sido extensivamente investigados ao longo do genoma humano, os esforços para a descoberta dos indels e de outras formas de variação genética têm sido mais escassos. Embora haja indicações de uma alta densidade de indels no genoma humano, ainda há muito a ser elucidado, uma vez que um número ainda discreto

tenha sido descoberto e validado (MILLS *et al.*, 2006). Apesar de serem abundantes, o interesse neste tipo de marcador é muito recente. Um dos estudos pioneiros para a descoberta de indels ao longo do genoma humano foi conduzido por Weber e colaboradores (2002), que identificaram cerca de 2000 polimorfismos. Outros trabalhos relataram a descoberta de indels por análise de dados de sequenciamento de genes alvos relacionados a vias metabólicas, com *software* que identificam indels e SNPs (BHANGALE *et al.*, 2005; BHANGALE; STEPHENS; NICKERSON, 2006). Projetos para identificação de indels em indivíduos de diferentes origens geográficas também vem sendo realizados, com abordagens bioinformáticas, em bases de dados de seqüências originalmente geradas para a descobertas de SNPs (MILLS *et al.*, 2006), assim como em seqüenciamentos de genomas pessoais (LEVY *et al.*, 2007; WHEELER *et al.*, 2008). Comparando-se os resultados obtidos nesses sequenciamentos, observou-se que tanto o número de indels quanto o tamanho dos mesmos têm apresentado grande variação entre os indivíduos. Isso pode ser explicado pelas diferentes plataformas de sequenciamento utilizadas, assim como as abordagens analíticas. As taxas de validação dos indels por PCR e sequenciamento também variam consideravelmente, sendo necessário o estabelecimento de parâmetros como a faixa de tamanho dos indels para comparações (MULLANEY *et al.*, 2010).

Diferentes grupos de trabalho desenvolveram painéis para tipagem de indels já validados em autossomos e no cromossomo X, e estudos colaborativos com o objetivo de relatar frequências alélicas em diferentes populações em todo o mundo vêm sendo realizados. Gradativamente, dados autossômicos têm sido disponibilizados para populações europeias, africanas, asiáticas e nativas americanas (WEBER *et al.*, 2002; YANG *et al.*, 2005; PEREIRA *et al.*, 2009; SANTOS *et al.*, 2010).

O genoma humano – recombinação e propriedades de transmissão

Uma das características importantes do genoma a ser levada em consideração no uso de marcadores genéticos é a recombinação, que consiste na troca aleatória de material genético entre os cromossomos, durante a meiose. Através dela ocorre um aumento da variabilidade ao longo das gerações, e a individualização, pois, com exceção de gêmeos idênticos, a probabilidade de duas pessoas compartilharem o mesmo perfil genético é virtualmente nula, para marcadores recombinantes. Dependendo da presença ou ausência de

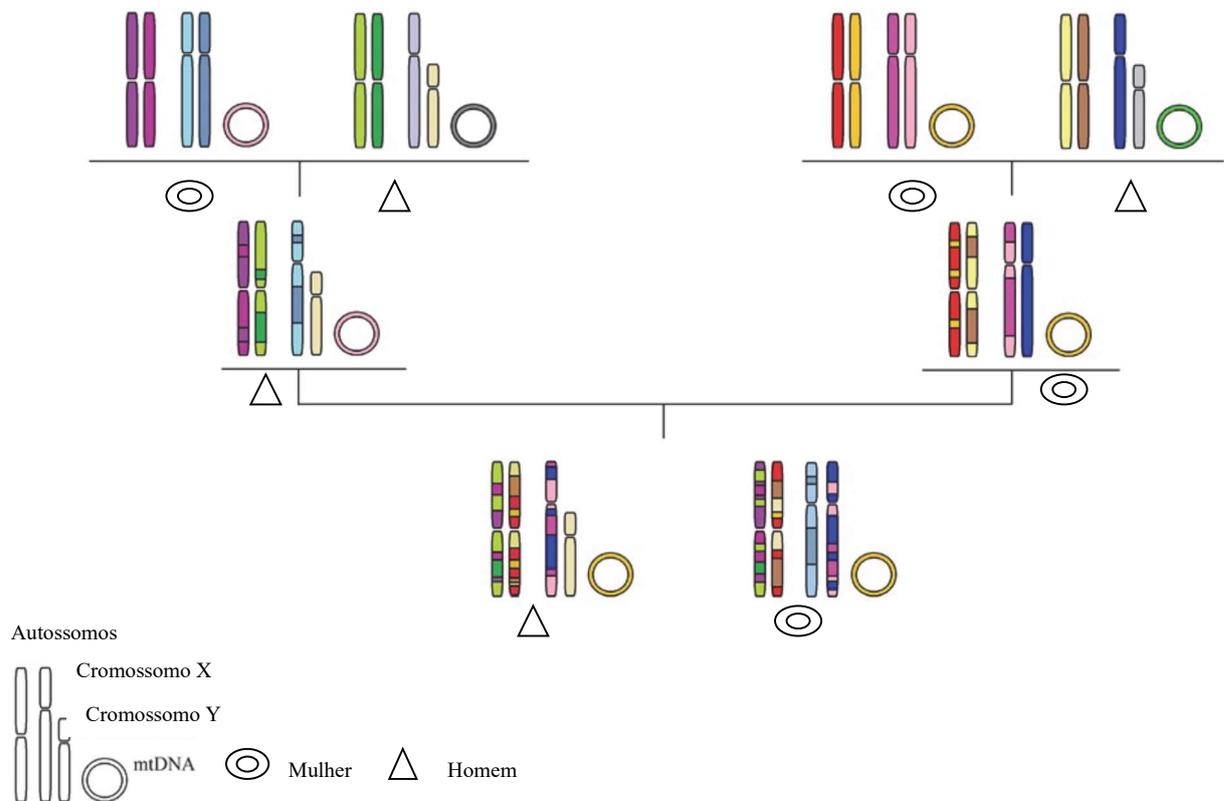
recombinação, podemos classificar as porções genômicas em dois grupos: genomas recombinantes, que permitem a identificação individual (marcadores autossômicos e do cromossomo X) e genomas não recombinantes, que permitem discriminar linhagens parentais e fornecem informações que são compartilhadas por um grupo de pessoas relacionadas (cromossomo Y - na sua maior parte - e mtDNA) (AMORIM; BUDOWLE, 2016). A Figura 2 mostra os diferentes genomas e suas propriedades de transmissão.

Genomas recombinantes: Autossomos e Cromossomo X

Devido à recombinação e ao grande número de cromossomos disponíveis para estudar, os autossomos são a primeira escolha quando se procura variação entre indivíduos. Os marcadores autossômicos ainda são, portanto, a principal fonte de informação para estudos de genética populacional e genética forense, havendo diversos *kits* comerciais bem padronizados, com alto poder de discriminação entre indivíduos (GILL *et al.*, 2012; GJERTSON *et al.*, 2007; MORLING *et al.*, 2002).

A transmissão biparental dos autossomos é importante para a sua escolha em análises de parentesco, uma vez que tanto o pai quanto a mãe contribuem com metade da informação para seu filho ou filha (Figura 2). Portanto, quaisquer indivíduos do sexo masculino ou feminino com parentesco materno ou paterno compartilharão alelos por descendência com uma probabilidade que depende do coeficiente de co-ancestralidade entre eles (PINTO; SILVA; AMORIM, 2010). Esses marcadores fornecem uma resposta satisfatória na maioria dos casos envolvendo análises de vestígios, identificação individual e na análise de parentesco. Mas em alguns casos mais complexos é necessário complementar a análise com outros tipos de marcadores que geralmente possuem diferentes características, como por exemplo, marcadores do mtDNA ou dos cromossomos X e Y (COBLE *et al.*, 2009).

Figura 2 - Diferentes tipos de genomas e suas propriedades de transmissão e de recombinação



Legenda: Os genomas recombinantes, representados pelos autossomos e pelo cromossomo X, com herança biparental. O cromossomo Y, com herança uniparental paterna e o mtDNA com transmissão uniparental materna, ambos sem recombinação.

Fonte: Adaptado de AMORIM; BUDOWLE, 2016.

Marcadores do cromossomo X e autossômicos têm o mesmo PD para perfis femininos em uma população. No entanto, porque os homens têm apenas um cromossomo X, o PD obtido para *loci* específicos deste é menor do que o obtido para *loci* igualmente diversos nos autossomos. Sendo assim, a relevância do cromossomo X se deve principalmente ao seu número diferencial entre homens e mulheres, que tem consequências nos níveis de recombinação: O único cromossomo X masculino é transmitido às filhas sem recombinação, enquanto o cromossomo X herdado da mãe contém a informação recombinada dos dois cromossomos X maternos (Figura 2). Devido a este modo particular de herança, a aplicação desses marcadores depende essencialmente do sexo dos indivíduos envolvidos.

Embora usado em um contexto mais específico, os marcadores do cromossomo X também têm alguma aplicabilidade em ciência forense, como em situações de incesto, onde os dados autossômicos não são suficientes para identificar o parentesco (SZIBOR, 2007;

GOMES *et al.*, 2012), ou em investigação de parentesco, nomeadamente em casos de paternidade onde o suposto pai não está disponível e seus parentes mais próximos devem ser analisados em vez dele. Os marcadores do cromossomo X também podem ser de interesse em testes de paternidade, quando dois supostos pais são pai e filho e o filho investigado é do sexo feminino. Nesta situação, ambos os homens não compartilharão entre si alelos do X por descendência e irão transmitir seus haplótipos X intactos para uma filha. Outra utilidade está na resolução de testes de parentesco envolvendo duas irmãs ou meias-irmãs, ajudando a considerar ou excluir a paternidade de forma mais eficiente do que marcadores autossômicos (SZIBOR, 2007).

Genomas não recombinantes: Marcadores de linhagem do Cromossomo Y e do mtDNA

Os marcadores de linhagem, também chamados de uniparentais, encontram-se localizados no genoma mitocondrial e no cromossomo Y e informam quanto à história das linhagens femininas e masculinas, respectivamente. Por essa razão são especialmente úteis no estudo de eventos populacionais que ocorrem diferencialmente no genoma de homens e mulheres (SCHAFFNER, 2004). Além disso, esses marcadores ocorrem como cópia única no genoma e por isso têm um tamanho populacional efetivo reduzido, sendo de 1/4 dos autossomos e 1/3 dos cromossomos X. Essas características, em conjunto, fazem com que tais marcadores apresentem diferenças genéticas acentuadas entre os grupos continentais humanos.

O DNA mitocondrial

O mtDNA é uma molécula circular de fita dupla constituída por duas cadeias polinucleotídicas complementares, de composição nucleotídica distinta (Figura 3). A cadeia pesada (do inglês, *Heavy*, representada por H) tem uma maior percentagem de Guaninas (G) e Adeninas (A), que contribuem para um peso molecular superior, enquanto a cadeia leve (do inglês, *Light*, representada por L) é rica em Citosinas (C) e Timinas (T) apresentando, consequentemente, um menor peso molecular. A molécula tem aproximadamente 16569 pb,

podendo esse número variar devido a inserções e / ou deleções na sequência (ANDERSON *et al.*, 1981).

Embora o mtDNA seja considerado monoclonal (o que significa que todas as cópias dentro da célula são iguais), é comum para um indivíduo apresentar heteroplasmia, ou seja, ter mais de um tipo de sequência de mtDNA, variando em tamanho ou em composição nucleotídica (BENDALL *et al.*, 1996; GILL *et al.*, 1994). Provavelmente as heteroplasmas estão presentes em todos os indivíduos, mas frequentemente em níveis tão baixos que não são detectadas pelos métodos de sequenciamento comumente usados. Geralmente, quando estão em um nível de 20 % ou mais das moléculas de mtDNA já são detectáveis pelo sequenciamento de Sanger. A química usada no sequenciamento, assim como a posição da heteroplasmia também influenciam na detecção (TULLY *et al.*, 2001).

O padrão de herança do mtDNA é uniparental por via materna. Isso porque na fecundação, apenas o núcleo do espermatozóide entra no oócito, fundindo-se diretamente com o núcleo materno, e não contribuindo com outros elementos celulares. Além disso, possíveis mitocôndrias paternas presentes são marcadas com ubiquitina e seletivamente degradadas (SUTOVSKY *et al.*, 1999). Sendo assim, uma vez que o mtDNA não recombina a cada geração, sua variabilidade deve-se apenas à ocorrência de mutações ao longo das gerações, geralmente envolvendo um único nucleotídeo (SNPs), podendo ocorrer também pequenas inserções ou deleções, especialmente em regiões homopoliméricas ou poli-C. Essas regiões localizam-se entre as posições 302 e 310 e as posições 16183 e 16194 da molécula de mtDNA.

A molécula de mtDNA é composta por duas regiões - a região codificante e a região controle (Figura 3) - classificadas de acordo com a função que desempenham na molécula. A região codificante representa cerca de 90 % do genoma mitocondrial e compreende 37 genes que se traduzem em 22 tRNAs, 2 rRNAs e 13 polipeptídios. Já a porção não codificante do mtDNA, também chamada de região controle (CR, do inglês, *Control Region*) ou *D-loop* (do inglês, *Displacement loop*), tem uma extensão aproximada de 1120 pb e contém a origem de replicação da cadeia pesada.

A taxa de mutação do genoma mitocondrial é superior à do nuclear, tendo um valor médio na ordem de 10^{-8} . A proximidade da molécula com a cadeia de transferência de elétrons, e a consequente exposição a espécies reativas de oxigênio aumentam a probabilidade de ocorrência de mutação. Além disso, a ausência de histonas protetoras faz com que essa vulnerabilidade aumente. Soma-se também a esses fatores a baixa eficiência do sistema de reparo de erros durante a replicação do mtDNA. A velocidade de mutação ao longo da

molécula é heterogênea, sendo o maior nível de variabilidade genética encontrado na região controle, que se subdivide em três segmentos hipervariáveis: HV-I, HV-II e HV-III (Figura 3). Essa alta variabilidade se explica pelo fato desses segmentos não incluírem informação necessária à codificação de nenhuma substância essencial ao funcionamento celular, sendo as mutações acumuladas nessas regiões fixadas com maior frequência, enquanto as da região codificante são alvo de seleção negativa (ANDERSON *et al.*, 1981; SOARES *et al.*, 2009).

A sequência completa do mtDNA humano foi descrita pela primeira vez, por Anderson e colaboradores, em 1981, no laboratório de Frederick Sanger, em Cambridge; motivo pelo qual ficou conhecida como Sequência de referência de Cambridge (CRS, do inglês, *Cambridge Reference Sequence*). Cada posição nucleotídica foi numerada, com início na origem de replicação da cadeia pesada (posição 1) e término na posição 16.569 (ANDERSON *et al.*, 1981). Quase duas décadas depois, o genoma mitocondrial estudado por Anderson foi reanalisado pelo grupo de Andrews e colaboradores, que identificaram 11 erros na sequência descrita em 1981. As correções foram publicadas e a sequência passou a ser denominada por Sequência de Referência de Cambridge revista (rCRS, do inglês *revised Cambridge Reference Sequence*), sendo essa considerada a sequência padrão de comparação em genética populacional e forense atualmente (ANDREWS *et al.*, 1999).

As primeiras investigações genéticas para fins forenses e populacionais basearam-se apenas em HV-I e HV-II, mas atualmente a Sociedade Internacional de Genética Forense – ISFG (do inglês, *International Society for Forensic Genetics*) recomenda que no mínimo seja utilizada a sequência da CR completa para a publicação de dados populacionais de referência (PARSON *et al.*, 2014). Em estudos que envolvem a variabilidade do genoma mitocondrial, a separação de duas linhagens maternas pode não ser possível apenas com o estudo da região controle. Na tentativa de ultrapassar esta limitação, começou-se a estudar polimorfismos SNPs localizados na região codificante do mtDNA. Na verdade, com o uso crescente de técnicas de MPS, a expectativa é que cada vez mais dados de sequenciamento da molécula completa de mtDNA, também chamada de mitogenoma, sejam publicados (KING *et al.*, 2014; MIKKELSEN *et al.*, 2014; STROBL *et al.*, 2018).

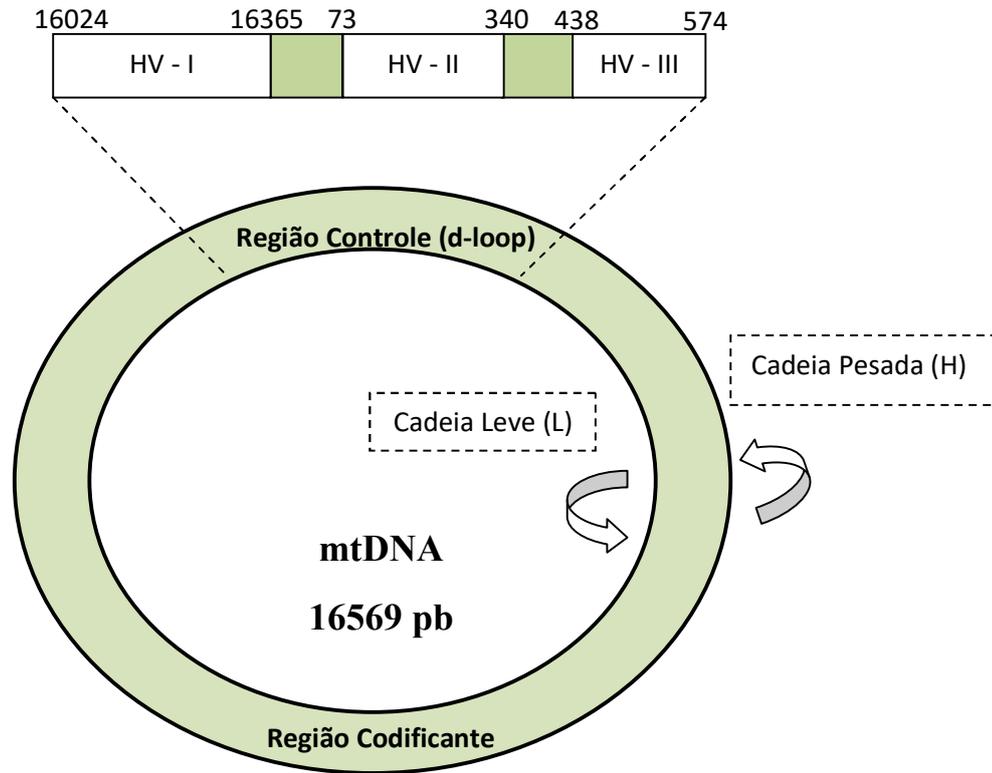
O estudo do mtDNA é amplamente utilizado como ferramenta em pelo menos três áreas distintas: a genética médica, a genética forense e a genética de populações ou antropologia molecular (KAYSER, 2007; TORRONI *et al.*, 2006; UNDERHILL ; KIVISILD, 2007). Na genética forense, o uso de marcadores do mtDNA está limitado a alguns casos de investigação forense, pelo seu reduzido PD em comparação aos marcadores autossômicos, uma vez que não permite a identificação a nível individual. O fato de serem transmitidos em

bloco, como se fossem um único marcador, pode ser uma vantagem em situações de identificação de vítimas de desastres em massa ou pessoas desaparecidas, por exemplo, quando os parentes de mesma linhagem materna estão disponíveis, ainda que sejam separados por várias gerações. Mas uma vez que indivíduos não aparentados podem ter sequências semelhantes, sempre que possível, é aconselhável que a análise de mtDNA seja complementada com dados genéticos adicionais.

Quando usado em estudos populacionais, o genoma mitocondrial apenas nos fornece a informação transmitida ao longo das gerações por via materna. No entanto, essa mesma característica, associada ao alto número de cópias, ausência de recombinação e taxa de mutação maior que o DNA nuclear, possibilita o rastreamento de linhagens maternas ao longo do tempo. Supõe-se que todos os tipos de mtDNA no *pool* genético humano podem ser rastreados até um ancestral matrilinear comum que viveu há aproximadamente 200.000 anos na África (MACAULY *et al.*, 2005), através de polimorfismos específicos, uma vez que a sequência do mtDNA evoluiu, levando à divergência das linhagens como resultado do acúmulo sequencial de mutações.

Do ponto de vista técnico, a maior vantagem da análise do mtDNA em amostras forenses está no alto número de cópias presentes em cada célula, e também na estrutura circular da molécula, que a torna mais resistente à degradação que o DNA nuclear. Por essas razões, a taxa de sucesso na extração do DNA e subsequente amplificação por PCR é muito maior para o mtDNA do que para o DNA nuclear, o que torna a análise do genoma mitocondrial particularmente útil em amostras antigas ou degradadas, e ainda em amostras muito escassas, em que a quantidade de material biológico disponível para análise é reduzida (LINCH; WHITING; HOLLAND, 2001; PAABO, 1989).

Figura 3 - Representação da estrutura circular do genoma mitocondrial



Legenda: Molécula circular do mtDNA com as cadeias pesada (H) e leve (L). A região não codificante ou região controle em destaque, mostrando as posições dos segmentos hipervariáveis HV-I, HV-II e HV-III.

Fonte: Adaptado de BUTLER, 2005.

Em relação às metodologias de análise para o mtDNA, quando se estuda apenas a região controle ou segmentos desta, comumente se usa a técnica de sequenciamento de Sanger para a detecção dos polimorfismos SNP. Nessa metodologia, o sequenciamento é realizado com *primers* direto e reverso, de forma a obter-se informação complementar das duas fitas, visando um controle de qualidade das sequências (SANGER; NICKLEN; COULSON, 1977). Quando se tem por objetivo o sequenciamento de mitogenomas, as análises realizadas com técnicas de MPS têm apresentado resultados de alta precisão, sensibilidade e eficiência, além de ser de alto rendimento e simples operação. Consistem, resumidamente, na quebra de sequências de DNA em pequenos fragmentos (aproximadamente entre 400-600 pb), o sequenciamento destes em paralelo, seguido da montagem e ordenação num único fragmento contíguo. O produto sequenciado é diretamente detectado sem que haja a necessidade de realizar eletroforese capilar. Essas técnicas têm superado algumas limitações do sequenciamento de Sanger, como a perda de informação em amostras degradadas e a detecção de heteroplasmias e misturas, por exemplo (HOLLAND; MCQUILLAN; O' HANLON, 2011; KING *et al.*, 2014; PARSON *et al.*, 2015).

Após sequenciamento do genoma mitocondrial, as sequências obtidas são descritas por comparação com a sequência de referência rCRS, sendo representadas pelo conjunto de posições diferentes em relação a ela, ficando subentendido que as posições restantes são idênticas. O conjunto de polimorfismos identificados numa sequência de mtDNA denomina-se haplótipo, o qual caracteriza uma determinada linhagem materna. A nomenclatura dos haplótipos é feita com base em diretrizes detalhadamente descritas pela comunidade científica, de forma a evitar erros de interpretação (AMORIM; FERNANDES; TAVEIRA, 2019; PARSON *et al.*, 2014; TULLY *et al.*, 2001). A existência de uma nomenclatura única permite o intercâmbio e a comparação de resultados entre laboratórios, a nível internacional, de forma adequada. A partir do agrupamento de haplótipos de mtDNA, definidos por SNPs específicos compartilhados por linhagens próximas entre si que possuem um ancestral comum, formam-se os haplogrupos, que tendem a mostrar alguma especificidade regional, que pode ser mais abrangente ou restringir-se até mesmo a grupos étnicos. Esses haplogrupos são nomeados por letras do alfabeto e podem conter sub-haplogrupos com alterações específicas (AMORIM; FERNANDES; TAVEIRA, 2019; PAKENDORF; STONEKING, 2005).

Visando detalhar as relações filogenéticas de variantes conhecidas do mtDNA, foi desenvolvida uma árvore filogenética – *PhyloTree* (VAN OVEN; KAYSER, 2009), englobando os polimorfismos identificados até aquele momento e que definem os haplogrupos e suas ramificações (Figura 4). A *PhyloTree* vem sendo atualizada à medida que novos polimorfismos em sequências de mitogenomas completos são descritos e pode ser acessada na página eletrônica <http://www.phyloTree.org>. A detecção da variabilidade da molécula completa de mtDNA garante uma maior precisão na descrição dos haplogrupos do que apenas com a região controle, uma vez que a taxa de mutação nas regiões codificantes é bastante inferior e, por isso, apresentam uma menor probabilidade de recorrência e, conseqüentemente, uma maior especificidade geográfica.

Os ramos mais profundos e antigos da *PhyloTree* incluem haplogrupos do continente africano, mais especificamente da região subsaariana. Em uma resumida descrição da distribuição dos haplogrupos, temos que: Os haplogrupos africanos são L0 a L6 (SALAS *et al.*, 2002; TORRONI *et al.*, 2006), considerados os de maior diversidade. Os haplogrupos M e N surgem na África oriental, a partir de L3, dispersando-se pela Eurásia e dando origem às linhagens "out-of-Africa" presentes em populações desta região. As linhagens europeias derivam do haplogrupo N, incluindo os haplogrupos H, I, J, K, T, U, V, W e X. Estes, em conjunto, estão presentes em 98 % da população europeia. As linhagens asiáticas derivam

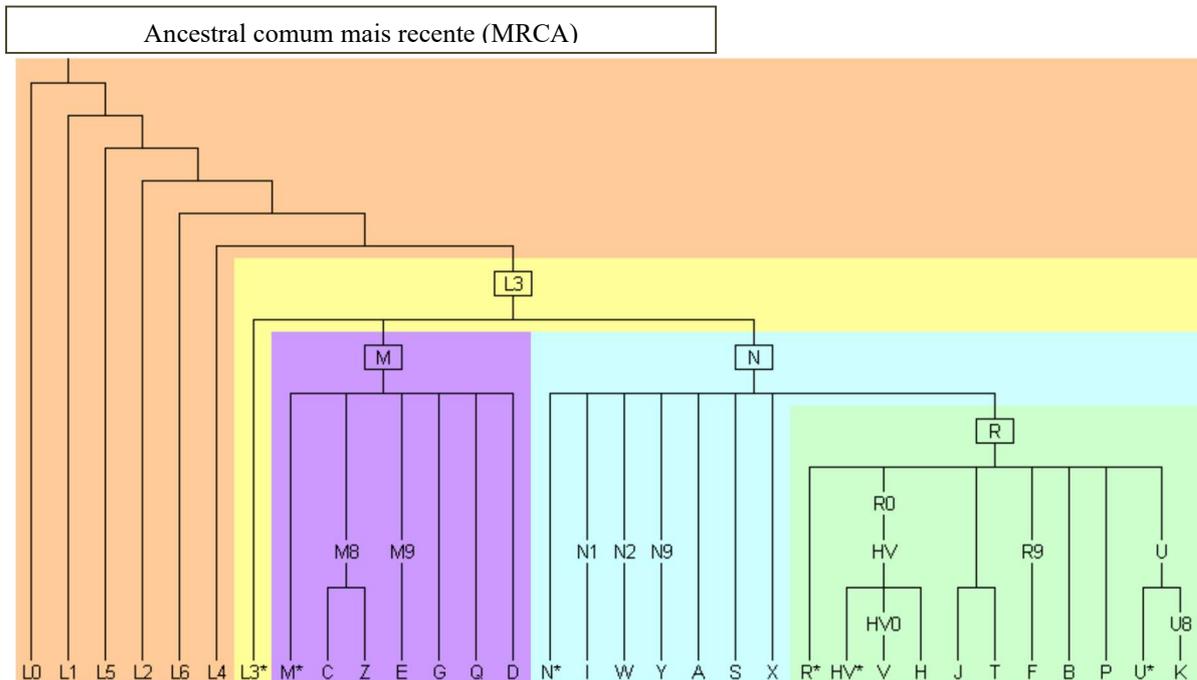
tanto do haplogrupo M como do haplogrupo N. Nas populações nativo-americanas verifica-se a presença dos haplogrupos A, B, C e D, característicos também da Ásia, enquanto os haplogrupos G, Y e Z predominam na Sibéria (MISHMAR *et al.*, 2003; QUINTANA-MURCI *et al.*, 1999).

A distribuição geográfica e a variabilidade das linhagens do mtDNA podem ser altamente informativas nas inferências de rotas de migração no passado distante. Porém, para estudar a história das populações humanas, além de analisarmos as linhagens existentes em uma população, é crucial o uso de métodos estatísticos que comparem as populações, estimando, por exemplo, distâncias genéticas entre as mesmas, a fim de compreendermos as afinidades populacionais. Além disso, o mtDNA reflete apenas a história via materna de uma população e representa apenas um *locus*, podendo não ser, por isso, tão preciso, devido a efeitos de deriva e da seleção natural sobre esse *locus*. Sendo assim, associar dados do cromossomo Y e de marcadores autossômicos em estudos populacionais pode ser muito relevante (PAKENDORF; STONEKING, 2005; WOODING *et al.*, 2004).

Bancos de dados populacionais de mtDNA, assim como bancos de dados de outros tipos de marcadores usados em genética forense, servem de base para estimativas de frequências de haplótipos de mtDNA gerados em laboratório e que sejam pertinentes a casos judiciais. Embora sejam essenciais para os cálculos estatísticos que validam as evidências apresentadas nesse contexto, os dados depositados nessas bases estão sujeitos a erros cometidos por quem os gera. Para estabelecer padrões de qualidade nas etapas de geração, análise, transferência e controle de qualidade de dados do mtDNA, de forma a atenderem os padrões requeridos em análises forenses, foi criada a base de dados EMPOP (*EDNAP mtDNA population database*), pelo Instituto de Medicina Legal de Innsbruck (PARSON; DÜR, 2007). A base de dados EMPOP é considerada a mais importante para o mtDNA, e inclui desde sequências contendo apenas a região HV-I até mitogenomas completos, sendo a grande maioria sequência da CR completa (<https://empop.online>). Essa plataforma digital está acessível para consulta pública, com dispositivos de análise disponíveis para a comunidade científica, contribuindo tanto como um banco de dados de populações de referência quanto como ferramenta de controle de qualidade para publicações contendo dados de mtDNA (AMORIM; FERNANDES; TAVEIRA, 2019).

Outro importante banco de dados *online* de mtDNA humano é o MITOMAP (<http://www.mitomap.org>), onde é possível acessar informações sobre as variantes do mtDNA associadas a dados geográficos e a doenças, além de ilustrações, distribuição de haplogrupos e suas frequências (LOTT *et al.*, 2013).

Figura 4 - Representação simplificada dos haplogrupos de mtDNA na *Phylo tree* e suas ramificações



Legenda: Os haplogrupos africanos L0-L6 são os mais antigos da árvore. O haplogrupo L3 ramifica-se nos haplogrupos euroasiáticos M e N, que por sua vez se subdividem nos haplogrupos euroasiáticos restantes e nos nativo americanos A -D. Os símbolos dos haplogrupos seguidos por um asterisco representam todas as linhagens descendentes não mostradas de um clado particular, para o qual uma letra do alfabeto não foi reservada.

Fonte: VAN OVEN; KAYSER, 2009.

O Cromossomo Y

O cromossomo Y é um dos menores cromossomos do genoma humano, cujo tamanho é de aproximadamente 58 milhões de pares de bases (MORTON, 1991). Ocorre somente em uma cópia por célula e é específico nos homens, sendo transmitido de pais para filhos (Figura 2). Nesse cromossomo estão presentes dois tipos de cromatina: eucromatina e heterocromatina, localizadas em diferentes porções. A heterocromatina tem tamanho muito variável entre indivíduos, e encontra-se no braço longo. Possui genes em sua extensão, mas em sua maior parte é caracterizada por sequências extremamente repetitivas (também conhecidas como amplicônicas) e não funcionais, que podem se apresentar de maneira consecutiva (*in tandem*) ou como sequências invertidas (palíndromo). A eucromatina está presente no restante do cromossomo e em sua extensão existem genes responsáveis por importantes funções biológicas, como por exemplo, os genes da região de determinação sexual (SRY, do inglês *Sex-determining Region*), e o gene da amelogenina (AMEL Y) (Figura

5). Aproximadamente 95 % da extensão do cromossomo Y, conhecida como região não-recombinante (NRY, do inglês, *Non-recombining Y chromosome region*) ou região masculina específica do cromossomo Y (do inglês, *Male-specific Y chromosome region* ou MSY), não está sujeita à recombinação com o cromossomo X, durante a meiose (SKALETSKY *et al.*, 2003). Apenas as regiões PAR1 e PAR2, que estão nas extremidades dos braços curto (Yp) e longo (Yq) do cromossomo, recombinam, pois são homólogas a sequências do cromossomo X e são responsáveis pelo pareamento adequado entre os dois cromossomos sexuais durante a meiose. (Figura 5) (BUTLER, 2005).

Algumas características do genoma mitocondrial são comuns ao cromossomo Y, uma vez que, assim como o mtDNA, a região NRY é de natureza haplóide e ausente de recombinação. Por isso o conjunto de alelos dos marcadores do cromossomo Y, também denominado de haplótipo, é integralmente e exclusivamente transmitido a todos os indivíduos do sexo masculino de uma mesma linhagem paterna, e a variabilidade entre eles só vai existir se ocorrerem mutações e arranjos intracromossômicos ao longo do tempo (JOBILING; HURLES; TYLER-SMITH, 2003). Comparado ao mtDNA, porém, o cromossomo Y tem maior diversidade, pois além de marcadores bialélicos SNPs e indels, existem muitos STRs disponíveis para estudos genéticos forenses e populacionais. A escolha adequada dos marcadores do cromossomo Y a serem usados em cada tipo de análise deve levar em consideração suas características singulares, como o nível de diversidade e taxa de mutação.

Figura 5 – O cromossomo Y



Legenda: Cromossomo Y e sua extensão aproximada em megabases. Em destaque estão as regiões PAR1 e PAR 2, que se recombinam com regiões homólogas do cromossomo X, além da região NRY, as porções de heterocromatina e eucromatina e os genes da região de determinação sexual (SRY) e da amelogenina AMEL Y.

Fonte: RÉGO, 2019.

Na genética forense, os marcadores do cromossomo Y encontram algumas limitações, uma vez que são aplicáveis somente em casos que envolvam o sexo masculino, e podem demonstrar vínculos genéticos apenas por via paterna. Por não serem capazes de individualização, mas sim de caracterização de linhagens, os marcadores do Y não podem excluir a paternidade de nenhum outro homem pertencente à mesma linhagem paterna do verdadeiro pai biológico, por exemplo. Sendo assim, devem ser usados em conjunto com marcadores com maior poder de discriminação entre indivíduos, como os autossomos. Apesar dessas limitações, nas análises forenses, os polimorfismos do cromossomo Y são especialmente úteis quando os perfis obtidos a partir de marcadores autossômicos não são suficientemente informativos. Um exemplo são os casos de paternidade na ausência de material biológico pertencente ao suposto pai. Nessas situações (por exemplo, quando o suposto pai já faleceu), é possível acessar o perfil genético completo de seu cromossomo Y usando qualquer parente do sexo masculino na mesma linhagem paterna e comprovar o vínculo (AMORIM; BUDOWLE, 2016; KAYSER, 2017).

Haplótipos do cromossomo Y têm sido amplamente utilizados também em investigações criminais, podendo, por exemplo, revelar se um doador de evidências deixadas em uma cena de crime é do sexo masculino, ou se existe mais de um doador do sexo masculino, no caso de evidências com múltiplas fontes. No caso de evidências onde existe mistura do DNA da vítima com o DNA do agressor, pode ser muito difícil separar os perfis genéticos de ambos usando marcadores autossômicos. Mesmo em amostras que contenham células oriundas de esperma do agressor misturadas às células da vítima, em que existe a possibilidade de uma extração diferencial, separando as frações celulares e obtendo os perfis genéticos separados, a extração diferencial tende a falhar quando a fração de esperma é muito baixa. Nestes casos, quando os marcadores autossômicos são usados, a amplificação preferencial do componente principal presente na mistura (geralmente o DNA da vítima) pode mascarar o perfil genético do estuprador. Porém, em uma mistura de DNA masculino e feminino, mesmo em baixa proporção de material masculino, a genotipagem adicional de marcadores do cromossomo Y pode fornecer um perfil específico do homem (AMORIM; BUDOWLE, 2016; KAYSER, 2017).

Os Y-STRs, por terem taxa de mutação maior que os SNPs e Indels, apresentam maior diversidade alélica; e ainda podem ser encontrados em cópia única ou em múltiplas cópias ao longo da extensão do cromossomo, gerando, assim, uma maior variedade alélica pela combinação dos *loci* existentes. Tal característica torna esses marcadores úteis em diversas aplicações nas quais o principal objetivo é a identificação de linhagens paternas, como:

investigações de parentesco e em investigações criminais, além de estudos de diversidade populacional e identificação de homens vítimas de desastres (BALLANTYNE *et al.*, 2010; KAYSER, 2017). Para um maior poder de diferenciação de indivíduos, costuma-se analisar vários *loci* Y-STR simultaneamente. Os marcadores de escolha devem estar bem caracterizados e apresentar alto nível de polimorfismo. Também é necessário conhecer a distribuição das frequências haplotípicas nas populações de interesse, assim como as taxas de mutação desses *loci*.

Apesar de seu amplo uso na ciência forense, os marcadores Y-STR podem apresentar alguns desafios nos cálculos estatísticos envolvendo suas frequências. Isso porque os dados genéticos devem ser tratados como haplótipos, para a determinação dessas frequências, tornando a construção de bancos de dados de Y-STR complexa, uma vez que todo o haplótipo do Y deve ser determinado geneticamente para cada amostra, ao invés de tipagens isoladas de marcadores, como nos autossomos. Como consequência, a maioria dos bancos de dados populacionais é representada principalmente por haplótipos muito raros ou únicos quando um grande número de *loci* está incluído. Sendo assim, para a aplicação prática desses marcadores, bancos de dados adequados devem possuir o maior número de indivíduos da população local com a tipagem do maior número possível de *loci*, a fim de driblar a ocorrência de subestrutura populacional (GUSMÃO; CARRACEDO, 2003). As frequências haplotípicas dos Y-STRs, assim como suas taxas de mutação, podem ser calculadas através de banco de dados próprio, mas distribuições de haplótipos Y-STR em populações de todo o mundo têm sido disponibilizadas por meio de publicações e, mais recentemente, por meio de bases de dados forenses de larga escala. Um dos maiores e mais utilizados bancos de dados de haplótipos de referência *online* para o cromossomo Y é o YHRD (do inglês, *Y-Chromosome Haplotype Reference Database*), cujos objetivos principais consistem na disponibilização de dados de inúmeras populações mundiais, com controle de qualidade, acerca das frequências haplotípicas e das taxas mutacionais de Y-STRs (www.yhrd.org) (KAYSER, 2017; WILLUWEIT; ROWER, 2015). O desenvolvimento desses bancos de dados é importante não apenas para a estimativa de frequência de haplótipos e subsequentes cálculos de probabilidade de correspondência em estudos forenses, mas também para a análise comparativa e populações.

A comunidade forense tem focado alguma atenção em marcadores de mutação rápida localizados no cromossomo Y, os RM-Y-STRs, cujas taxas de mutação são da ordem de 10^{-2} . Por terem taxas de mutação acima da média, estes marcadores irão fornecer uma maior probabilidade de distinguir entre indivíduos intimamente relacionados e podem complementar

análises atualmente feitas com Y-STRs convencionais (BALLANTYNE *et al.*, 2012). Alguns kits comerciais mais recentes, como o *PowerPlex Y23 System* (Promega) e *Yfiler Plus* (Thermo Fisher Scientific), já incluem alguns desses RM-Y-STRs.

Outra classe de marcadores muito utilizada na análise do cromossomo Y são os SNPs. Por serem bialélicos, oferecem menor informação a nível intrapopulacional em comparação com os marcadores Y-STR. Os marcadores Y-SNP possuem baixas taxas de mutação (na ordem de 10^{-8} por nucleotídeo/por geração) quando comparados com os STRs. A caracterização genética de populações com ambos os tipos de marcadores simultaneamente possibilita uma análise eficiente de eventos relacionados com a evolução e a história das populações humanas. Uma vez que esses marcadores apresentam diferentes taxas de mutação, são capazes de apontar aspectos populacionais a níveis micro e macro-evolutivos, evidenciando dinâmicas ocorridas, como eventos migratórios e padrões de miscigenação.

Assim como no mtDNA, devido às baixas chances de mutação recorrente, os marcadores SNP do cromossomo Y tendem a apresentar eventos moleculares únicos da evolução humana, com uma elevada especificidade geográfica e populacional, uma vez que esses genomas haplóides estão mais sujeitos aos efeitos de deriva gênica, que resultam em frequências diferenciais entre regiões geográficas. Por décadas, esses marcadores foram usados como meio de rastrear a ancestralidade biogeográfica de indivíduos e populações, inicialmente em análises genealógicas, estudos evolutivos e populacionais (JANNUZZI *et al.*, 2020; RESQUE *et al.*, 2016; UNDERHILL; KIVISILD, 2007), a fim de conhecer rotas de migrações e origens. Tal conhecimento acumulado sobre a distribuição geográfica da diversidade genética do cromossomo Y, atualmente serve como base para as aplicações forenses de inferência de ancestralidade no campo forense, assim como no mtDNA (KAYSER, 2017; UNDERHILL; KIVISILD, 2007). São úteis em situações envolvendo doadores de DNA vestigial em cenas de crimes ou de vítimas de acidentes em massa, onde a identificação não tenha sido possível com marcadores autossômicos. Também podem ser usados em casos de identificação de pessoas desaparecidas e de vítimas de desastres em que não haja outras informações sobre a origem dos restos mortais. Em todas essas situações podem apontar uma linha investigativa a ser seguida.

Por ser um marcador de evolução lenta, a probabilidade de um determinado SNP sofrer a mesma mutação em genomas independentes é muito pequena e, desta maneira, os SNPs são capazes de agrupar linhagens ao associar indivíduos por descendência. Por convenção, um conjunto específico de Y-SNPs, assim como no mtDNA, definem haplogrupos do cromossomo Y. A baixa probabilidade de recorrência dessas mutações

permite assumir que o alelo mutante ou derivado de cada marcador Y-SNP surgiu apenas uma vez na história humana, e todos os homens que possuem tal alelo descendem de um ancestral comum, no qual a mutação ocorreu pela primeira vez (MIZUNO *et al.*, 2010; UNDERHILL; KIVISILD, 2007). Os polimorfismos descritos nas árvores filogenéticas do cromossomo Y são capazes de detectar a origem de linhagens com distribuição tanto ao nível inter quanto intracontinental (CHIARONI; UNDERHILL; CAVALLI-SFORZA, 2009; CRUCIANI *et al.*, 2011).

O primeiro sistema de filogenia e nomenclatura dos haplogrupos do cromossomo Y foi publicado em 2002, por um grupo de colaboradores, e definiu 18 haplogrupos principais, nomeados pelas letras A – R (Y CHROMOSOME CONSORTIUM, 2002). Desde então, atualizações foram publicadas, incluindo diversos Y-SNPs recentemente descobertos e considerados relevantes para estabelecer as relações evolutivas das linhagens do cromossomo Y, visando aumentar a resolução da árvore filogenética. A versão mais recente da árvore é composta por 20 haplogrupos principais, nomeados pelas letras A – T (JOBILING; TYLER-SMITH, 2003; KARAFET *et al.*, 2008; VAN OVEN *et al.*, 2014). Algumas fontes de informação *online*, como as páginas eletrônicas mantidas pela Sociedade Internacional de Genealogia Genética (ISOGG; <http://www.isogg.org/tree>) e por Thomas Krahn (<http://ytree.ftdna.com>) já mostram um vislumbre da enorme topologia da árvore filogenética 'completa' usando todos os Y-SNPs conhecidos e mapeados filogeneticamente, incluindo muitos que (ainda) não foram publicados na literatura científica.

Nas duas últimas décadas, várias filogenias diferentes do cromossomo Y e nomenclaturas de Y-SNPs e haplogrupos têm sido apresentadas na literatura científica e em conferências que demonstram a presente diversidade no cromossomo Y. Essas configurações, nem sempre consensuais, podem ser atribuídas ao crescimento exponencial do número de Y-SNPs descobertos, devido principalmente ao surgimento das novas tecnologias de MPS, que suportam estudos envolvendo o sequenciamento de genomas completos a um menor custo e maior velocidade que as tecnologias convencionais. Com o volume de informação que se tem gerado a partir de tais estudos a respeito da distribuição de haplogrupos do cromossomo Y em diferentes regiões, novas variantes deverão ser descobertas, havendo uma demanda por remodelações da estrutura da árvore filogenética do cromossomo Y (LARMUSEAU *et al.*, 2015; VAN OVEN *et al.*, 2014).

A árvore filogenética e a nomenclatura sugeridas por Van Oven e colaboradores, em 2014 (VAN OVEN *et al.*, 2014), é uma proposta concisa, que não teve como objetivo usar todos os Y-SNPs conhecidos, mas sim as mutações mais estáveis que dão origem às

ramificações, em busca da melhor resolução em nível mundial. Essa versão mais atual está acessível na página eletrônica www.phylotree.org/Y, e foi baseada em um conjunto de 417 Y-SNPs. Tem sido a mais utilizada na determinação de haplogrupos, nomeando-os juntamente com as mutações que os caracterizam (Figura 6). Sendo assim, como exemplo, a notação A1-M31 refere-se ao haplogrupo A1, que pode ser identificado pela presença da mutação M31.

Resumidamente, na árvore, as duas principais divisões da base são os haplogrupos, A e B, cuja propagação é restrita à África, reforçando as evidências da origem do homem moderno. Ambos haplogrupos primários são geneticamente diversos e ramificam-se em subhaplogrupos geograficamente distintos um do outro, revelando uma possível fragmentação da população, isolamento e subsequente reexpansão no continente africano (Figura 6). O haplogrupo E também é conhecido por ser um haplogrupo africano, porém não se restringe à África, podendo ser encontrado alguns subhaplogrupos no oeste do continente Asiático e mais ao sul da Europa. Acredita-se que estava presente na população africana pela qual se deu a expansão “Out of Africa”. Os haplogrupos G, H, I, J, R, T e L são observados em populações da Europa, Oriente Médio e algumas partes da Ásia e do norte da África. Os haplogrupos C e Q podem ser encontrados no continente asiático. Porém, no continente americano, o haplogrupo C está restrito ao norte, enquanto o haplogrupo Q estabeleceu-se mais profundamente (CHIARONI; UNDERHILL; CAVALLI-SFORZA, 2009; UNDERHILL, KIVISILD, 2007).

Uma vez sabidas as mutações que definem os ramos filogenéticos do cromossomo Y, e, portanto, seus haplogrupos; é possível inferir a ancestralidade paterna de um indivíduo do sexo masculino analisando esses Y-SNPs e agrupando os resultados de maneira hierárquica, seguindo os ramos da árvore até o ancestral comum. Essa análise direcionada facilita a genotipagem, visto que não é necessário analisar todos os marcadores, mas somente os de mutações mais recentes, diminuindo o custo e o trabalho envolvido. Ao consultar a árvore (Figura 6), é possível saber, por exemplo, que todo indivíduo pertencente ao haplogrupo A1 possui as mutações M31, V168 e L1085. Porém, na prática, só é necessária a detecção da mutação final – M31.

Devido às diferenças nas taxas mutacionais entre os marcadores Y-STR e Y-SNP, um determinado haplogrupo pode ser composto por diferentes haplótipos Y-STR. Alguns haplogrupos baseados em Y-SNP com diferenças marcantes de frequências entre regiões geográficas exibem uma forte associação com a diversidade haplotípica dos Y-STRs das linhagens que os compõem, de modo que a origem geográfica indicada pelo haplogrupo também pode ser inferida a partir dos haplótipos de Y-STR. A genotipagem de marcadores

STRs é realizada de forma simples, através de PCR e eletroforese capilar somente, enquanto a determinação de polimorfismos SNPs requerem técnicas de sequenciamento ou mini-sequenciamento, que envolvem mais etapas laboratoriais. Pode ser, portanto, vantajoso; utilizar esse tipo de marcador na predição de haplogrupos, principalmente na ausência de dados de marcadores Y-SNP (JOBLING, 2001; KAYSER, 2017). Sendo assim, vários *software* têm sido desenvolvidos, com o objetivo de inferir o haplogrupo correspondente a um haplótipo de Y-STR, a fim de direcionar a escolha dos Y-SNPs a serem analisados.

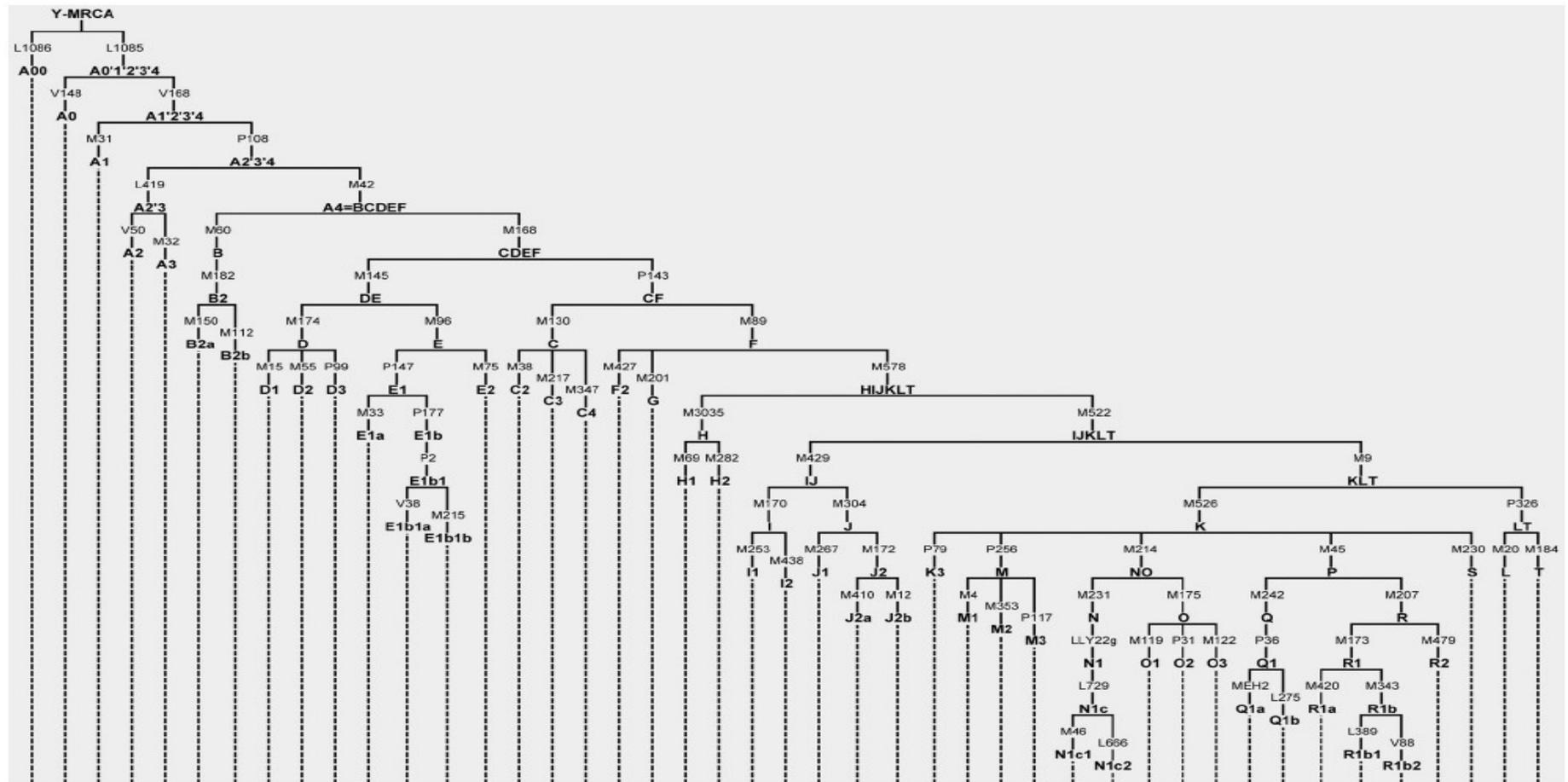
Um *software* gratuito amplamente utilizado pela comunidade científica é o *Haplogroup Predictor* (<http://www.hprg.com/hapest5/>), que se utiliza de um algoritmo implementado em um programa *online*, com abordagem Bayesiana. O *Haplogroup Predictor* indica a probabilidade de um determinado haplótipo pertencer a um haplogrupo, com base nas frequências alélicas observadas de cada marcador que compõe o haplótipo em diferentes haplogrupos. Essas frequências estão depositadas em base de dados públicas consultadas pelo programa. A abordagem Bayesiana adotada requer que, no momento da análise, seja indicada a possível região geográfica na qual os haplótipos analisados possam ter sido originados, visto que as frequências dos haplótipos podem variar conforme a região considerada. Uma opção é a que considera prioridades iguais para as regiões geográficas, devendo ser usada para indicar outras regiões que o programa não disponibiliza, como por exemplo, regiões do continente americano (ATHEY, 2006; MUZZIO *et al.*, 2011).

O *NevGen Y-DNA Haplogroup Predictor* é outro *software* gratuito que pode ser usado diretamente na internet ou instalado no computador. O *NevGen* utiliza um algoritmo baseado na abordagem Bayesiana de frequências alélicas, com o diferencial de realizar a correlação (interdependência) de valores de diferentes pares de STRs durante os cálculos de probabilidade de predição de subhaplogrupos. A base de dados consultada pelo *NevGen* é a *Family Tree DNA* (FTDNA - www.familytreedna.com) e somente considera haplótipos que estejam associados com dados de elevadas resoluções de SNPs. Além disso, o *NevGen* não necessita que a possível região de origem do haplótipo seja indicada para o cálculo da estimativa da probabilidade. A fim de evitar a atribuição equivocada de algum haplogrupo, quando o algoritmo utilizado por este programa não é capaz de estimar uma probabilidade de um determinado haplótipo ser associado a algum haplogrupo considerado pelo programa, ele indica como resultado a probabilidade deste pertencer a “Subclados não-suportados” (GENTULA; NEVSKI, 2015).

A abordagem de predição de haplogrupos utilizada por estes *software* tem como limitações principais o tamanho do banco de dados utilizado, visto que se este for de tamanho

reduzido a estimativa das frequências alélicas e a variedade de haplogrupos observados serão precárias; além da falta de representatividade de certos haplogrupos e de grupos populacionais, como por exemplo, populações dos continentes americano e asiático (MUZZIO *et al.*, 2011).

Figura 6 - Árvore filogenética de haplogrupos do cromossomo Y



Legenda: Representação da árvore filogenética do cromossomo Y, com destaque em negrito para a nomenclatura de base dos haplogrupos e os Y-SNPs que definem cada ramificação. No topo, está indicado o ancestral comum mais recente de todos os homens modernos (Y-MRCA, do inglês *Most Recent Common Patrilineal Ancestor of all modern humans*).

Fonte: VAN OVEN et.al., 2014.

Ancestralidade Biogeográfica e Marcadores Informativos de Ancestralidade

Devido ao grande aumento da compreensão da variação populacional humana, pelo detalhamento do genoma e conhecimento de sua diversidade, têm sido identificadas porções variáveis entre indivíduos, com potencial para serem usados como marcadores informativos de ancestralidade – os AIMs (do inglês, *Ancestry Informative Markers*).

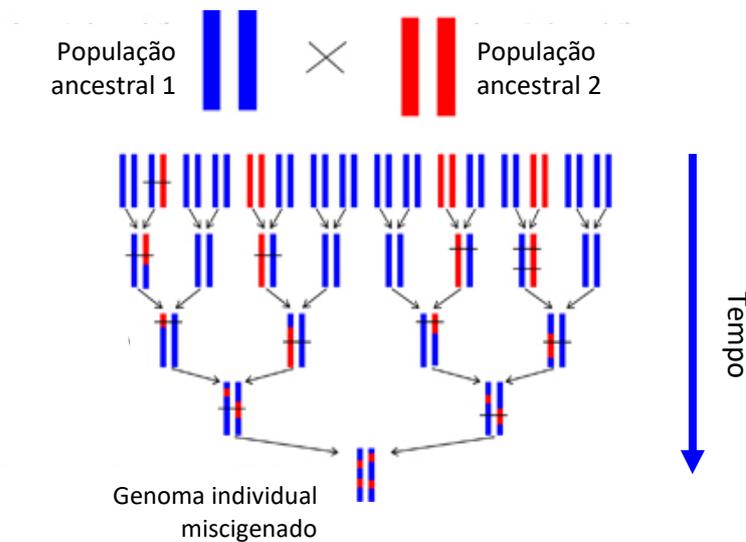
Os AIMs têm como principal característica uma baixa divergência nas frequências alélicas dentro das populações e uma alta divergência destas frequências em populações separadas geograficamente. Estes marcadores são especialmente úteis para inferir a provável origem ancestral de um indivíduo ou estimar as proporções de ancestralidade em indivíduos ou populações miscigenadas (PEREIRA *et al.*, 2012a). A miscigenação é uma forma comum de fluxo gênico e refere-se ao processo pelo qual duas ou mais populações genética e fenotipicamente diversas acasalam-se e formam uma nova população, híbrida. Como resultado desse processo, cada cromossomo de um indivíduo miscigenado se apresenta como um mosaico de segmentos cromossômicos derivados de seus ancestrais (Figura 7) (BAYE; WILKE, 2010).

Sendo assim, a ancestralidade pode ser definida como a herança genética que cada indivíduo traz de seus ancestrais diretos, resultante dos acasalamentos entre diversos membros da população em gerações anteriores, ao longo do tempo. A análise da ancestralidade biogeográfica se detém nas variações encontradas em indivíduos e que podem sinalizar a sua origem de uma região geográfica em particular. Não está restrita a avaliações genéticas, podendo ser estimada, por exemplo, através das análises biométricas de esqueletos (PHILLIPS, 2015).

Na genética forense, análises de inferência de ancestralidade biogeográfica encontram diversas aplicações, como no auxílio em investigações, podendo substituir ou colaborar com o relato de testemunhas oculares, uma vez que essas podem cometer enganos. Ou trazendo dados adicionais na ausência de bancos de dados ou coincidências com perfis genéticos depositados nestes. Nesse caso, genótipos de AIMs obtidos de evidências em uma investigação podem indicar a provável ancestralidade de seu doador. São úteis também na identificação de vítimas de desastres ou de pessoas desaparecidas. Na genética médica a confirmação da ancestralidade de doadores que fazem auto-declaração em estudos de associação caso-controle minimizam a chance de resultados espúrios por alteração da frequência de um alelo em um dos grupos do estudo, não por influência desse no fenótipo

estudado, mas por ser mais presente em indivíduos com certa origem ancestral. Dessa forma contribuem para a acurácia de bancos de dados em relação à origem de seus indivíduos. Por fim, vale mencionar a importância dessas análises também do ponto de vista antropológico, em estudos populacionais, visando elucidar rotas de migração e a origem do povoamento de uma região (MORIOT *et al.*, 2018).

Figura 7 - Representação esquemática do processo de miscigenação



Legenda: Indivíduos de populações ancestrais diferentes, representadas por cromossomos azuis e vermelhos, acasalam-se ao longo das gerações, e recebem material genético de seus ancestrais, sujeito à recombinação durante a meiose.

Fonte: Adaptado de BAYE; WILKE, 2010.

Embora os marcadores uniparentais do cromossomo Y e do mtDNA apresentem alta diferenciação em termos de origem geográfica continental e sejam informativos de ancestralidade, não são capazes de representar a ancestralidade global do indivíduo como o são os marcadores dos autossomos, por não serem recombinantes ao longo das gerações e devido ao seu padrão de herança. Esses marcadores uniparentais contam histórias complementares e independentes acerca dos ancestrais maternos e paternos dos povos. Tal característica pode resultar em interpretações equivocadas, como no caso de linhagens parentais de origem muito distante da ancestralidade total em um indivíduo pertencente a uma população miscigenada, e que apresente co-ancestralidade, o que é marcadamente comum nas populações urbanas atuais. Outra característica a ser considerada em relação ao uso de marcadores autossômicos, e que pode ser vista como vantajosa, é que necessitam de um

número de indivíduos consideravelmente menor para a elaboração de um banco de dados populacional com o fim de estimar as frequências alélicas para cálculos estatísticos. As variações haplotípicas do mtDNA e do cromossomo Y exigem bancos de dados maiores para estimativas mais precisas, e se tratando de populações provenientes de regiões pouco investigadas isso se torna uma dificuldade considerável. Daí a importância de mais estudos genéticos desses marcadores e o depósito desses dados em plataformas como o YHRD e EMPOP.

Devido a diversos fatores envolvidos em sua formação, a população humana mundial não pode ser vista de maneira uniforme; ao contrário, apresenta uma estrutura complexa. Daí a importância de se estabelecer bem os grupos populacionais, tendo uma base adequada na escolha de conjuntos de marcadores para inferência de ancestralidade, levando-se em consideração tanto o número de marcadores quanto seu nível de resolução geográfica e dados disponíveis de populações de referência. Diferentes tipos de marcadores oferecem diferentes níveis de resolução sobre a estrutura populacional. Evolutivamente, os marcadores com taxas de mutação relativamente baixas (SNPs, Inserções Alu, Indels) são os melhores *loci* para a análise da história humana ao longo do tempo em escalas maiores, ou seja, fornecem uma resolução biogeográfica em nível continental. Já os marcadores de evolução rápida, como os STRs, fornecem maior resolução para escalas de tempo mais curtas (DE KNIJFF, 2000).

Um dos pioneiros em estudos sobre a diversidade genética populacional foi Lewontin (LEWONTIN, 1972) e posteriormente diversos outros estudos vem sendo feitos, que variam nos *loci* e nas populações analisadas, visando avaliar a estrutura dos grupos populacionais humanos modernos. Os resultados encontrados são similares aos de Lewontin, que descreveu que a maior parte das variações em autossomos (aproximadamente 85 %) ocorre dentro das populações. Algumas décadas depois, outro estudo da estrutura populacional foi feito com 1054 indivíduos pertencentes a 52 populações do Projeto de Diversidade do Genoma Humano (HGDP-CEPH) (ROSENBERG *et al.*, 2002), utilizando o algoritmo de agrupamentos por similaridade genética do *software* STRUCTURE (PRITCHARD; STEPHENS; DONNELLY, 2000). As análises com o STRUCTURE para o conjunto de 377 STRs autossômicos usados identificaram 5 grupamentos continentais com similaridades genéticas, os quais foram definidos como Eurásia, África Subsaariana, Leste Asiático, América e Oceania. Admitindo-se um número de grupamentos igual a 7 no STRUCTURE, observa-se que a Eurásia se subdivide ainda em Europa, Oriente Médio e Centro-Sul Asiático. Apesar do tamanho reduzido de algumas das populações e lacunas consideráveis do ponto de vista geográfico, esse estudo com 377 STRs autossômicos ainda é considerado o mais informativo envolvendo

a população global. Poucos anos depois 2 outros estudos foram publicados, com um número maior de marcadores STR, sobre diversas populações da América e da Oceania isoladamente (WANG *et al.*, 2007; FRIEDLAENDER *et al.*, 2008). Esses estudos em conjunto forneceram as primeiras avaliações detalhadas sobre a distribuição da estrutura populacional global. Um novo estudo realizado usou 650000 SNPs para analisar o mesmo grupo de amostras HGDP-CEPH (LI *et al.*, 2008) e os resultados encontrados foram muito semelhantes. Rosenberg e colaboradores também reanalisaram as amostras com um número muito maior de STRs (933), obtendo dados concordantes com o estudo anterior (ROSENBERG *et al.*, 2005). Pode-se dizer que, apesar do detalhamento alcançado com o maior número de marcadores desses estudos, o padrão global de diversidade e da estrutura populacional humana permanece o mesmo encontrado inicialmente, com discretas, mas consistentes variações, refletindo as diferenças nas frequências alélicas. Essas tendem a aumentar sutilmente com a distância geográfica; mas pequenas discontinuidades existentes levam a separações, daí surgindo os grupamentos populacionais identificados pelo STRUCTURE. Foi revelada também, através dos estudos realizados, uma diminuição da variabilidade genética ao longo do eixo África/Ásia – Eurásia – Oceania – América, que pode ser explicada por uma série de eventos fundadores ocorridos nas migrações humanas ao longo do tempo (PHILLIPS, 2015). Os testes de ancestralidade devem levar esses fatos em consideração, uma vez que para um mesmo grupo de marcadores, o nível de heterozigosidade em populações africanas é bem maior quando comparado ao encontrado na América, por exemplo (WANG *et al.*, 2007). Pelas mesmas razões, as divergências encontradas entre populações africanas e não africanas usualmente são maiores do que entre outros pares de grupos. Sendo assim, haverá um número inferior de marcadores bem diferenciados entre populações mais próximas, como por exemplo, o par Eurásia – Leste Asiático.

Resumidamente, o que se conclui sobre as características observadas sobre a diversidade genética humana, é que indicam um padrão comum de migrações de grupos reduzidos de pessoas para novas regiões, havendo uma separação do grupo ancestral de origem seguida de rápida expansão. Esses efeitos de deriva genética tem sido uma forte força evolutiva atuante na formação da população humana contemporânea. Forças como a seleção natural, que varia regionalmente de acordo com fatores biogeográficos como clima, dieta e a presença de doenças, e também o fluxo gênico acentuado, presente em processos de miscigenação envolvendo populações diferenciadas, somam entre si na modelagem da diversidade humana. Mais recentemente, após o sequenciamento de genomas arcaicos de origem neandertal e denisovana, estudos realizados relatam uma ancestralidade neandertal

média de aproximadamente 2 % em populações modernas não africanas, e cerca de 7 % de ancestralidade denisovana em populações da Oceania. Esse fluxo gênico envolvendo o homem moderno e esses hominídeos é denominado introgressão arcaica (PHILLIPS, 2015; PICKRELL; REICH, 2014; REICH *et al.*, 2011; HUERTA-SANCHEZ *et al.*, 2014).

Inferências de ancestralidade geográfica usando marcadores genéticos dependem de três fatores essenciais, que são: um conjunto de AIMs, um banco de dados de amostras de referência genotipadas para os AIMs selecionados e um método estatístico de agrupamento por similaridade (PFAFFELHUBER *et al.*, 2020).

A escolha dos marcadores precisa ser pautada em alguns parâmetros inerentes ao marcador em si, mas também ao nível de divergência que esse marcador apresenta entre as populações a serem avaliadas. As medidas de variabilidade de um marcador entre populações podem ser expressas por diferentes sistemas baseados nas distâncias genéticas existentes, sendo esse marcador de natureza bialélica, como os SNPs ou Indels, ou multialélica, como os STRs. As distâncias genéticas, por sua vez, estão correlacionadas com as diferenças das frequências alélicas existentes nas populações para o dado marcador, sendo um dos sistemas mais utilizados o índice F_{ST} , mas também a frequência alélica diferencial (δ) e informatividade (I_n) (ROSENBERG *et al.*, 2003). Os valores dessas medidas variam de 0 a 1, e esses extremos representam, respectivamente, divergência nula e 100 % de divergência para comparações de pares de populações. Uma vez que se conheça esses valores para cada marcador frente às populações de interesse, é possível estabelecer um *ranking* de nível de diferenciação populacional para um conjunto de marcadores (SHRIVER *et al.*, 1997; FRUDAKIS *et al.*, 2003; GOLDSTEIN *et al.*, 1995; ROSENBERG *et al.*, 2003) e através de testes compactos obter inferências de ancestralidade. É fundamental, porém, verificar o balanço da divergência que os marcadores selecionados apresentam em conjunto entre os grupos populacionais envolvidos, além da disponibilidade dos dados populacionais disponíveis para esses grupos, necessários para as comparações.

Dado que a distribuição da diversidade humana levou a uma forte divergência entre as populações africanas e não africanas, seguida da divergência entre Eurásia e outras populações, e posteriormente, do Leste asiático com populações da Oceania e da América; sabe-se que a seleção de AIMs tende a encontrar muito mais *loci* informativos africanos do que para comparações de outros grupos. Já os americanos, como um grupo populacional com apenas 15 Ka de separação, têm a menor divergência em relação aos asiáticos do leste, intimamente relacionados (COLONNA *et al.*, 2011). Sendo assim, se o objetivo de um teste de ancestralidade for diferenciar a África, a Europa, o Leste Asiático e a América, é mais

difícil encontrar marcadores que distingam europeus e asiáticos do que africanos e americanos. Isso demonstra como os valores de divergência precisam de avaliação criteriosa para as comparações populacionais em um teste, atingindo um balanço adequado no conjunto de marcadores selecionado. Isso é especialmente importante quando se analisa populações miscigenadas, que apresentam co-ancestralidade (TABOADA-ECHALAR *et al.*, 2013).

Em relação aos bancos de dados populacionais disponíveis para consulta e seleção de AIMS, pode-se considerar que ainda há pouca informação sobre populações da Oceania, de nativos da América e algumas regiões da Ásia. Portanto, um objetivo válido seria caracterizar uma grande coleção de populações para um pequeno número de painéis de ancestralidade usando tamanhos de amostra razoáveis, em torno de 50 indivíduos ao menos. Porém, com o crescente uso de análises usando plataformas de sequenciamento paralelo massivo, tem sido possível disponibilizar um volume cada vez maior de informações, como os dados de SNPs do estudo com o painel HGDP-CEPH, com 650.000 *loci* (LI *et al.*, 2008) e do *1000 Genomes Project Consortium*. Os dados da Fase 3 desse projeto atingiram a marca de aproximadamente 79 milhões de variantes em 2535 indivíduos provenientes de 26 populações (THE 1000 Genomes PROJECT CONSORTIUM, 2012). A partir da disponibilidade desses dados alguns painéis de AIMS com variações no número de SNPs genotipados, que utilizam a química do tipo *Snap-Shot*, de extensão de base única, foram desenvolvidos e posteriormente otimizados (PHILLIPS *et al.*, 2007; FONDEVILA *et al.*, 2013; KERSBERGEN *et al.*, 2009; LAO *et al.*, 2006; GETTINGS *et al.*, 2014). Subsequentemente, outros painéis contendo maior número de AIMS, com maior abrangência de grupos populacionais e boa capacidade de resolução também foram desenvolvidos e aperfeiçoados. Dois conjuntos de AIMS, compatíveis com plataformas de MPS, com 128 e 55 SNPs, foram combinados (KIDD *et al.*, 2011; PHILLIPS *et al.*, 2014a), formando o *HID-Ion AmpliSeq Ancestry Panel*, otimizado para o sistema *NGS Ion PGM*, ambos da *Thermo Fisher Scientific*. Além dos dados publicados pela página eletrônica do *1000 Genomes Project Consortium*, frequências alélicas e dados genotípicos de SNPs também podem ser obtidos de outros bancos de dados de consulta pública, como SPSmart, FROGkb e Alfred (AMIGO; SALAS; PHILLIPS, 2011; RAJEEVAN *et al.*, 2012a e 2012b).

Em relação aos sistemas estatísticos para inferências de ancestralidade biogeográfica, podem ser citados: Análise Bayesiana, Análise dos Componentes Principais (PCA) e o *software* STRUCTURE, que usa a abordagem Bayesiana. Cada um desses sistemas usa dados de populações de referência e faz inferências através da comparação dos padrões de variação encontrados nessas populações e na população com ancestralidade desconhecida. Por isso, a

relevância, a qualidade e a abrangência dos dados populacionais disponíveis são fatores importantes a serem considerados na análise (PHILLIPS, 2015).

O STRUCTURE é o programa de análise populacional mais amplamente usado atualmente, podendo ser usado com marcadores do tipo SNP e STR simultaneamente (FALUSH; STEPHENS; PRITCHARD, 2003). Nesse *software*, como dito anteriormente, uma vez que a estimativa de agrupamentos ou *clusters* populacionais é feita para as populações de referência, é possível inferir a ancestralidade individual em uma nova amostra, por comparação. A cada “corrida” o STRUCTURE produz uma matriz de coeficientes de associação aos agrupamentos populacionais, permitindo comparações entre coeficientes das amostras de referência e da amostra de ancestralidade desconhecida. A análise de agrupamento normalmente não é supervisionada, portanto as amostras não são rotuladas por região e, conseqüentemente, os agrupamentos são determinados apenas por padrões de similaridade genética detectada entre as amostras. Este processo é frequentemente usado para testar a eficiência de um conjunto de marcadores para diferenciar grupos específicos, ou seja, se os *clusters* formados corresponderem bem à região geográfica de origem, o painel pode ser considerado informativo para os grupos que foram analisados. Uma variável fundamental na análise pelo STRUCTURE é a estimativa do número de grupos populacionais (K) a serem considerados, uma vez que esse precisa ser coerente com a real estrutura genética das populações analisadas. Nem sempre é simples de se determinar o melhor valor de K, e um *software* que pode auxiliar nessa escolha é o CLUMP (JAKOBSSON; ROSENBERG, 2007; KALINOWSKI, 2011).

A miscigenação é uma característica dominante das populações que habitam as margens dos continentes e uma ocorrência cada vez mais crescente, que se iniciou desde a primeira migração de pequenos grupos humanos. Ao longo de mais de 2 mil anos de história de conquistas, tráfico de escravos e crescente comércio, além das movimentações de massas e a urbanização, dos séculos mais recentes; barreiras culturais e sociais, que antes substituíam a separação geográfica entre os povos, tem sido cada vez menores. Conseqüentemente, as análises de ancestralidade genética deparam-se atualmente com uma grande variabilidade de padrões nas proporções de mistura entre os indivíduos testados. Na área forense, os investigadores criminais também têm interesse particular nas análises de co-ancestralidade porque esta abriga a possibilidade de combinações incomuns de características físicas em um determinado suspeito. Comparando-se as abordagens de inferência aqui citadas, em relação à capacidade de estimar a ancestralidade em indivíduos miscigenados, pode-se dizer que a de maior limitação é a análise Bayesiana, seguida da PCA, na qual é aconselhável usar

comparações considerando até 3 populações ancestrais por análises, e finalmente pelo STRUCTURE, mais frequentemente usado (PHILLIPS, 2015). A análise com o STRUCTURE compara os padrões de agrupamentos encontrados com os padrões de indivíduos de origem ancestral conhecida (FONDEVILA *et al.*, 2013). É importante ressaltar que limitações existem nessa abordagem, como por exemplo o fato de que as populações originais que contribuíram para a miscigenação não podem ser extrapoladas de forma 100 % eficiente a partir de amostras (KIDD *et al.*, 2011).

Além dos marcadores AIMS do tipo SNPs, que são os mais clássicos, outros tipos de marcadores também podem ser informativos de ancestralidade, como os Indels, STRs autossômicos e marcadores SNPS multialélicos. Os Indels, que são marcadores bialélicos, assim como os SNPs, analiticamente também podem ser analisados em fragmentos de PCR curtos e sistemas de multiplexes com *primers* marcados com fluoróforos. Porém são processados diretamente da PCR para a eletroforese capilar (CE), sem etapas intermediárias. Embora existam SNPs mais informativos que os Indels, estes apresentam a vantagem de distinguir melhor picos provenientes de misturas (PEREIRA *et al.*, 2009).

Diversos painéis AIM-Indels já foram desenvolvidos, entre eles um painel com 48 marcadores distribuídos em 3 mutiplexes (SANTOS *et al.*, 2010), um painel com 46 indels em 1 único multiplex (PEREIRA *et al.*, 2012a) e um painel de 21 indels também em 1 multiplex (ZAUMSEGEL; ROTHSCCHILD; SCHNEIDER, 2013). O painel de 46 AIM-Indels apresenta divergências entre África, Europa e Leste Asiático comparáveis ao painel de 34 SNPs - 34 plex (FONDEVILA *et al.*, 2013; PHILLIPS *et al.*, 2007), mas também diferencia os nativos americanos. Por essa razão é uma opção de inferência de ancestralidade genética de simples execução técnica, podendo ser realizada com um único teste de PCR – eletroforese capilar.

Em relação aos marcadores do tipo STR autossômicos, dois tipos de abordagens podem ser usadas: um painel com um grande número de marcadores ou com marcadores de alto poder de diferenciação entre as populações. Quanto à estrutura populacional global, verificou-se que no conjunto de 377 marcadores STRs usados para inferir ancestralidade, marcadores selecionados ao acaso foram mais eficientes quando comparados com marcadores SNPs escolhidos também aleatoriamente, e alguns STRs foram considerados altamente informativos em relação aos SNPs, principalmente os que tem sequência de repetição de dinucleotídeos (ROSENBERG *et al.*, 2003).

Por serem marcadores amplamente utilizados na genética forense, utilizar os dados provenientes de STRs para inferir ancestralidade pode ser um campo com grande potencial

de investigação. Alguns estudos já utilizaram, por exemplo, marcadores dos *kits* comerciais *Identifiler* (PHILLIPS *et al.*, 2011), *Identifiler Plus* (LONDIN *et al.*, 2010), *Global Filer*, *Fusion*; isoladamente ou combinados entre si ou com outros STRs adicionais e SNPs. A existência de alelos STR específicos de determinadas populações já foi relatada (PHILLIPS *et al.*, 2014b), sendo o de maior especificidade o alelo 9 do marcador D9S1120, muito frequente entre os nativos americanos (53 %) (PHILLIPS *et al.*, 2008). Um ensaio multiplex com 12 marcadores STR informativos de ancestralidade com repetições de tetranucleotídeos foi desenvolvido, apresentando bom nível de resolução, porém uma taxa maior de erro na detecção da ancestralidade africana (PHILLIPS *et al.*, 2013).

Embora o desempenho dos STRs isoladamente ainda não seja considerado promissor como AIMs, esforços crescentes vem sendo feitos, como adaptações de *software* para uso desse tipo de dado e a criação de bancos de dados com perfis e frequências genotípicas (PEREIRA *et al.*, 2011), buscando minimizar as distorções nas interpretações dos resultados.

As técnicas de NGS, além de permitirem a ampliação de painéis de AIM-SNPs e AIM-Indels, também podem auxiliar no aumento do nível de informação destes. Isso porque as diversas sequências geradas de um fragmento de centenas de bases permitem genotipar simultaneamente todas as variantes presentes no mesmo, como SNPs ou Indels embutidos em regiões de STRs ou múltiplos SNPs formando haplótipos; muitos dos quais apresentam uma distribuição informativa de ancestralidade. As combinações alélicas ou haplótipos para cada cromossomo, ou seja, sua fase, podem ser determinados, mesmo nos autossomos, o que o sequenciamento de Sanger não possibilitava anteriormente. Esses haplótipos foram classificados, de acordo com o seu tamanho em kb, em minihaplótipos (1–10 kb) e microhaplótipos (200 pb) (PAKSTIS *et al.*, 2012; KIDD *et al.*, 2014; MORIOT *et al.*, 2018). Uma avaliação cuidadosa da taxa de recombinação na região em que ocorrem é necessária para estimar seu grau de informação. Embora taxas de recombinação muito baixas ajudem a preservar os haplótipos em maiores extensões, chegando a quilobases, alguma recombinação é necessária para gerar frequências de haplótipos informativos entre as populações. Da mesma forma, fragmentos muito curtos precisam de maior atividade de recombinação para gerar novos conjuntos de alelos.

Um exemplo desses *loci* haplotípicos são os recentemente descritos marcadores compostos DIP-STR, que apresentam a vantagem de possuir tanto marcadores Indels quanto marcadores STR, combinando os benefícios desses dois tipos de marcadores. A instabilidade do marcador STR faz com que múltiplos alelos surjam, proporcionalmente ao tempo de divergência das populações, enquanto o marcador SNP, por ser mais estável, permite maior

precisão no rastreamento da linhagem de cada haplótipo. Uma outra vantagem é que o número de haplótipos possíveis formado pela combinação desses marcadores é muito maior do que o número de alelos em cada *locus* individual. Essa maior variabilidade aumenta a probabilidade de haplótipos raros, que são facilmente perdidos durante a ocorrência de eventos de deriva populacional, como o efeito fundador e o efeito gargalo (MORIOT *et al.*, 2018).

Outra espécie genômica que pode ser informativa para a ancestralidade são os SNPs multialélicos. Inicialmente foram pouco utilizados, não sendo considerados na Fase I do *1000 Genomes Project Consortium*; mas atualmente foram bem caracterizados e constituem um em cada 300 SNPs da Fase III desse projeto. Alguns desses SNPs têm grande variabilidade entre populações, e são também marcadores úteis na detecção de misturas de DNA (WESTEN *et al.*, 2009). Possuem maior variabilidade, se comparados aos SNPs bialélicos, devido ao fato de fornecerem um maior número de combinações alélicas possíveis, e a distribuição geográfica desses alelos, influenciada pelos processos de deriva, pode resultar em um alto grau de informação de ancestralidade.

A formação da população brasileira e estudos genéticos de ancestralidade

A população brasileira é considerada uma das mais heterogêneas do mundo, sendo resultado da miscigenação entre as três principais populações ancestrais que também povoaram a América Latina, nomeadamente os povos nativos da América, os europeus e os africanos.

O território brasileiro, que era até então habitado por diferentes grupos nativos (também chamados de indígenas), começou a ser ocupado pelos primeiros grupos europeus que chegaram ao Brasil, de origem portuguesa. Dados históricos relatam que no mês de Abril, em 1500, os portugueses, liderados por Pedro Álvares Cabral, chegaram ao território brasileiro, na região do litoral da Bahia; fato este que constitui um dos episódios da expansão marítima portuguesa, no início do século XV.

Nos primeiros dois séculos de colonização, vieram para o Brasil cerca de 100 mil portugueses, uma média anual de 500 imigrantes. No século seguinte, esse número aumentou: foram registrados 600 mil e uma média anual de 10 mil imigrantes portugueses. O ápice do fluxo migratório ocorreu na primeira metade do século XX, entre 1901 e 1930: a média anual ultrapassou a barreira dos 25 mil. A origem socioeconômica do português imigrante é muito

diversificada: de uma próspera elite nos primeiros séculos de colonização, passou-se a um fluxo crescente de imigrantes pobres a partir da segunda metade do século XIX (IBGE, 2000).

Ao chegarem ao Brasil, os portugueses encontraram populações indígenas que ocupavam o território de forma bem distribuída ao longo da costa litorânea e na região da bacia dos Rios Paraná e Paraguai (abrangendo os estados de São Paulo, Paraná, Mato Grosso do Sul, Minas Gerais, Goiás, Santa Catarina e a região do Distrito Federal). Segundo estudos antropológicos recentes, baseados em afinidades culturais e linguísticas, essas populações compunham dois grandes grupos bem distintos: os tupis-guaranis, que se estendiam por quase toda a costa brasileira, e os tapuias, presentes em alguns pontos do litoral. Esses englobam, por exemplo, os goitacases na foz do Rio Paraíba, os aimorés no sul da Bahia e no norte do Espírito Santo e os tremembés na faixa entre o Ceará e o Maranhão (FAUSTO, 1996). Devido à dificuldade de obtenção de dados sobre os povos indígenas que aqui habitavam na época da chegada dos portugueses, existe uma grande variação nas estimativas quanto ao seu número. Segundo dados do IBGE, estima-se que a população indígena era de, aproximadamente, 2,5 milhões de indivíduos (IBGE, 2000). Praticavam a agricultura, a pesca, a caça e atividades extrativistas, sendo suas relações econômicas basicamente de subsistência e destinada ao consumo próprio.

A chegada dos portugueses iniciou-se com o reconhecimento e a posse da nova terra, através de uma ocupação crescente, em um período colonial que se consolidou ao longo de mais de dois séculos. Essa ocupação, feita também sob o pretexto religioso de catequização dos índios pelos representantes da Igreja Católica, teve como objetivo principal a exploração dos recursos naturais, que começou pelo litoral, com as derrubadas de árvores de pau-brasil pelos próprios indígenas, em troca de objetos de pouco valor. Nesse primeiro momento ocorreu caracteristicamente uma miscigenação entre mulheres indígenas e homens portugueses, visto que a vinda de mulheres europeias ainda era pouco comum (FAUSTO, 1996). Com o objetivo de expandir a colonização da nova terra, a Coroa portuguesa inicia a instalação de engenhos de açúcar e através de incentivos sociais e econômicos promove a formação de latifúndios, concentrados sob o poder de poucos proprietários que tinham ligações com o governo. Diferente das atividades iniciais da colônia, às quais os indígenas estavam familiarizados, como o corte do pau-brasil, as atividades agrícolas às quais os colonizadores queriam impor aos nativos intensificaram conflitos já existentes. O fato é que a conquista do território brasileiro provocou uma drástica diminuição na população indígena, que apesar de ter resistido fortemente aos colonizadores, sobretudo quando se tratou de

escravizá-los, sofreu grande violência cultural, e muitas mortes decorrentes de epidemias e desses conflitos.

De acordo com os resultados do Censo Demográfico 2010, realizado pelo IBGE, 817,9 mil pessoas se declararam indígenas, representando apenas 0,4 % da população total do Brasil. Sabemos que nesse discreto contingente auto declarado indígena inclui-se também uma parte da população mestiça, resultante do contato com os europeus. Esse último censo, em contraste com censos anteriores, que adotavam apenas os quesitos de cor e raça, aprimorou a investigação desse contingente populacional em particular, introduzindo o pertencimento étnico, a língua falada no domicílio e a localização geográfica. Se forem consideradas apenas as pessoas que vivem nas terras indígenas, de uma estimativa de 2,5 milhões para o século XVI, em 1998 chegou-se a um total de apenas 302.888 índios (IBGE, 2000; IBGE, 2012).

Apesar de tantas dificuldades enfrentadas pelos indígenas, desde a época do Brasil colônia, a preservação que se tem hoje, de sua herança biológica, social e cultural, foi possível principalmente pelos deslocamentos contínuos que faziam para regiões cada vez mais distantes, sendo esse isolamento uma forma muito eficaz de resistência.

Nas décadas posteriores à chegada dos primeiros grupos de portugueses, houve a vinda tanto de outros grupos europeus como também de escravos africanos trazidos por frotas portuguesas a partir de diferentes locais da África. A introdução do trabalho escravo a partir de meados do século XVI pelos conquistadores atendia idealmente aos seus interesses, uma vez que nem havia grande oferta de trabalhadores interessados em vir para o Brasil, nem o trabalho assalariado era conveniente para os fins da colonização. Por outro lado, havia uma grande disponibilidade de terras e a necessidade de sua efetiva ocupação produtiva.

Os escravos africanos começaram a ser trazidos para o Brasil para trabalhar principalmente na economia açucareira e nas casas de famílias portuguesas. Posteriormente, trabalharam também nas minas de ouro e diamantes e nas plantações de café. Uma vez que o índio não se adaptou bem ao trabalho compulsório, essa transição da escravidão do indígena para o negro africano foi se dando gradualmente. Estima-se que entre 1550 e 1850 (quando foi criada a lei que abolia o comércio de escravos), aproximadamente, 4 milhões de escravos africanos vindos de diferentes regiões da África, em grande maioria do sexo masculino, entraram pelos portos brasileiros (IBGE, 2000). Os locais com maior influxo de escravos africanos foram Salvador, que recebeu indivíduos em sua maioria da região de Guiné, e o Rio de Janeiro, que recebeu escravos principalmente de Angola (Figura 8). Diferente dos índios, os africanos não conheciam o território, sendo mais difícil fugir, além de ter mostrado uma

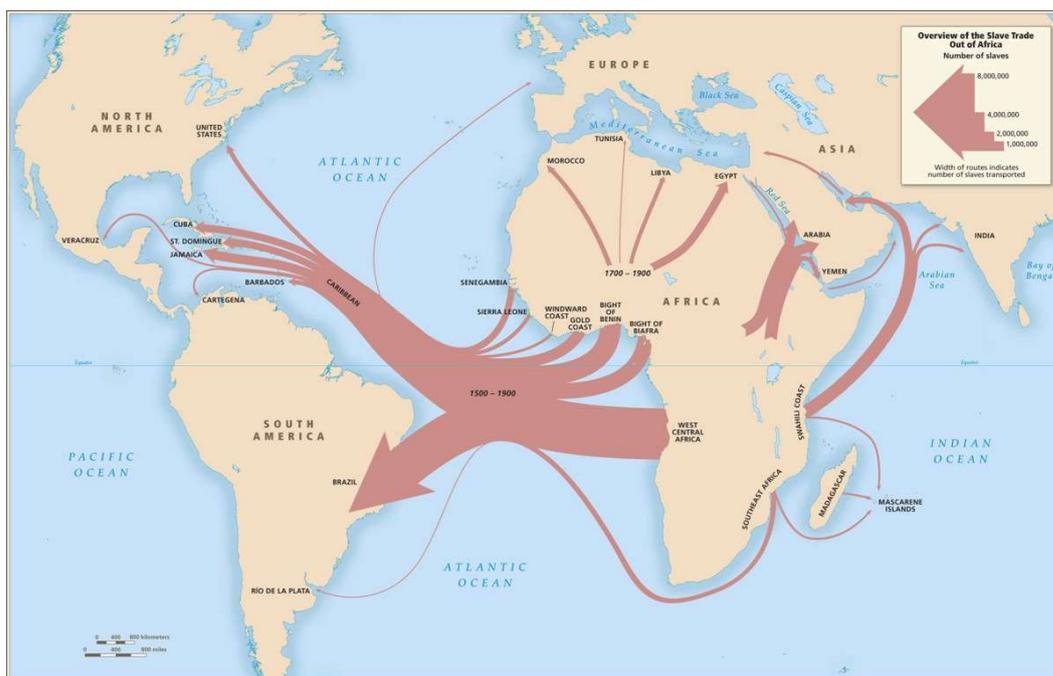
boa capacidade produtiva. Isso fazia deles uma ótima opção à escravização dos índios. Porém, fugas individuais e formações de quilombos, que eram locais de refúgio de escravos que escapavam e se organizavam, eram comuns no Brasil colonial, sendo um dos mais conhecidos o de Palmares, no Estado de Alagoas (FAUSTO, 1996).

As subseqüentes e contínuas migrações de escravos africanos e de europeus foram contribuindo para uma maior mistura na população do Brasil, que ia se formando diferencialmente. Os negros tinham alta representatividade como etnia brasileira em meados do século XIX, mas esse cenário começou a mudar devido a ondas de migrações incentivadas pelas autoridades brasileiras como forma de “branqueamento” da população, transformando o Brasil em um país predominantemente branco, principalmente nas regiões Sudeste e Sul (PENA *et al.*, 2011).

Além dos portugueses, houve invasões no nordeste brasileiro, inicialmente de franceses, que se estabeleceram no Maranhão e devido aos riscos de perda territorial atraíram para ali os portugueses, que através do Rio Amazonas, foram avançando para a Região Norte, fundando, em 1616, a cidade de Belém. Posteriormente, os holandeses chegaram ao estado de Pernambuco, em busca do açúcar, influenciando, sobretudo, na formação da população dessa região.

A colonização do norte do Brasil, diferentemente das outras regiões, ocorreu lentamente, com predomínio do trabalho compulsório indígena. Os primeiros negros chegaram à Amazônia não por intermédio dos portugueses, mas graças aos ingleses. Nas feitorias que montaram entre a costa de Macapá e a zona de estreitos, esses ingleses pretenderam realizar um empreendimento agrário de vulto, principalmente, de plantio de cana para a fabricação de açúcar e rum (FAUSTO, 1996; DE CAMPOS; DOLHNIKOFF, 2001). A chegada de escravos africanos em larga escala iniciou-se efetivamente apenas em meados do século XVIII, com estimativas de 53000 indivíduos que chegaram durante o período da escravidão (MIRANDA-NETO, 1976; PALHA *et al.*, 2011). Já o Sudeste teve sua colonização inicial através do litoral e a presença maciça de indígenas, que auxiliaram o avanço para o Centro-Sul do país, em direção ao interior. A presença de mestiços resultantes do cruzamento entre brancos e mulheres indígenas (mamelucos) era muito comum no estado de São Paulo, devido ao reduzido número de mulheres brancas.

Figura 8 - Rotas do tráfico de escravos africanos para o Brasil



Legenda: Principais rotas de influxo de escravos africanos para o Brasil. A maioria dos escravos africanos era proveniente das regiões Centro-Oeste e Sudoeste da África, principalmente das regiões da Guiné, Congo e Angola.

Fonte: <http://www.slavevoyages.org/static/images/assessment/intro-maps/01.jpg>

Na história recente do Brasil, após a abolição da escravidão em 1888, houve a vinda de imigrantes de outras regiões da Europa (Itália, Espanha e Alemanha, principalmente), e da Ásia (Síria, Líbano, China e Japão) (IBGE, 2000), que ocuparam principalmente o sul e sudeste do país; além de intensas migrações internas devido a motivações de ordem econômica.

Em conjunto, essas características da formação da população brasileira resultaram em padrões de miscigenação variáveis ao longo das cinco regiões geopolíticas nas quais o país é atualmente subdividido: Norte, Nordeste, Sudeste, Centro-Oeste e Sul. Ou seja, o território brasileiro apresenta uma distribuição heterogênea das contribuições ancestrais dos nativos americanos, europeus e africanos, uma vez que os encontros e acasalamentos entre esses ocorreram de maneira assimétrica, com europeus e africanos chegando pelo litoral e gradualmente alcançando o interior do país, para onde houve intensa migração de indígenas devido aos conflitos com os europeus (SALOUM DE NEVES MANTA *et al.*, 2013).

A heterogeneidade da população brasileira ao longo do seu território tem sido foco de interesse em muitos estudos, que usam diferentes tipos de marcadores genéticos. Em alguns deles, cada região geográfica é tratada como uma única população homogênea, enquanto outros estudos criam diferentes subdivisões: políticas (por exemplo, agrupando populações

por estado), demográficas (por exemplo, populações urbanas e rurais) ou étnicas (por exemplo, cultura, auto-declaração, ou cor da pele).

Os primeiros estudos da diversidade genética da população brasileira analisaram grupos sanguíneos e marcadores protéicos, resultando em uma análise ampla, porém superficial, da heterogeneidade dos brasileiros (SCHNEIDER; SALZANO, 1979). Os marcadores de linhagem também têm sido usados para a compreensão do complexo processo de miscigenação ocorrido no Brasil, sendo capazes de revelar os padrões de cruzamentos interétnicos ocorridos entre africanos, europeus e nativos (PENA *et al.*, 2009; RIBEIRO-DOS-SANTOS *et al.*, 2002). Estudos de ancestralidade usando marcadores autossômicos com diferentes tipos de polimorfismos vêm sendo realizados para elucidar padrões de miscigenação em populações brasileiras, uma vez que esses marcadores são informativos da ancestralidade global dos indivíduos (LINS *et al.*, 2010; PEREIRA *et al.*, 2012a; SANTOS *et al.*, 2010).

Salzano e Sans (2014), em um trabalho de revisão da literatura, reuniram dados publicados de diversos trabalhos sobre a população brasileira; sendo 13 deles realizados com marcadores de linhagens associados ou isoladamente, e 27 com marcadores autossômicos. As amostragens e o número de marcadores desses estudos foram variáveis, havendo amostras representativas das 5 regiões geopolíticas brasileiras, assim como de estados e cidades, e até mesmo grupos específicos, como comunidades afro-brasileiras. Nessa compilação de dados foi encontrado o predomínio da proporção de ancestralidade europeia, que é mais acentuada nas regiões sul e sudeste, seguida pela proporção de ancestralidade africana, em maior proporção no nordeste, e por fim o componente nativo, que é mais evidente na região Norte. A região Centro-Oeste, de acordo com estes autores, se assemelha à região Norte do país.

Considerando-se apenas a ancestralidade paterna, análises de marcadores SNP do cromossomo Y apontam na mesma direção, demonstrando que a maioria das linhagens masculinas de populações urbanas do Brasil é de origem europeia, seguida por um menor número de linhagens africanas (provenientes, principalmente, da África subsaariana) ou nativas-americanas, havendo alguma variabilidade entre as cinco regiões geopolíticas do Brasil (JANNUZZI *et al.*, 2020; PENA *et al.*, 2009; RESQUE *et al.*, 2016). Segundo o trabalho de Resque e colaboradores (2016), as regiões Centro-oeste e Sul são as que apresentam os maiores valores de ancestralidades europeias do país, enquanto que as regiões Norte e Nordeste possuem os valores mais altos de ancestralidades nativa americana e africana, respectivamente, observados em populações brasileiras.

As análises realizadas com o mtDNA isoladamente exibem uma maior ancestralidade europeia se considerado o *genepool* da população brasileira como um todo (ALVES-SILVA *et al.*, 2000). Porém, diferenças marcantes entre as regiões do Brasil são encontradas, com predomínio das linhagens europeias apenas no sul do país. Já na região Norte o predomínio é nativo, enquanto no Nordeste e maior parte do sudeste as linhagens africanas são maioria (ALVES-SILVA *et al.*, 2000; FRIEDMAN *et al.*, 2014; SIMÃO *et al.* 2018).

Com marcadores biparentais informativos de ancestralidade, análises realizadas em populações brasileiras mostram que a ancestralidade europeia é a mais prevalente em populações urbanas, com um gradiente crescente do norte para o sul do país, chegando a valores de até 79 %. As ancestralidades africana e nativa estão mais presentes nas regiões Nordeste e Norte, respectivamente (SALOUM DE NEVES MANTA *et al.*, 2013; PENA *et al.*, 2011).

Considerando-se a demografia do Brasil, além dos diferentes tipos de marcadores e critérios de amostragem utilizados nos estudos já realizados, a ancestralidade dos brasileiros em relação às contribuições europeia, africana e nativa ainda não é totalmente conhecida. Portanto, análises adicionais com AIMS, usando maiores tamanhos amostrais e que incluam populações ainda não caracterizadas, podem revelar novos aspectos da estrutura genética brasileira. No levantamento realizado por Salzano e Sans (2014), observa-se que poucas populações miscigenadas da região Norte do Brasil já foram investigadas quanto à ancestralidade genética com base em marcadores autossômicos, como Macapá e Belém.

A Ilha de Marajó

A Ilha de Marajó está localizada ao norte do estado do Pará, distante aproximadamente 90 km da capital Belém. Com cerca de três mil ilhas e ilhotas, o Marajó é o maior arquipélago flúvio-marítimo do mundo, banhado pelo Oceano Atlântico e pelos rios Amazonas e Tocantins. Sua ilha principal tem uma extensão de 50.000 m², sendo dividida administrativamente em 16 municípios: Afuá, Anajás, Bagre, Breves, Cachoeira do Arari, Chaves, Currálinho, Gurupá, Melgaço, Muaná, Ponta de Pedras, Portel, Santa Cruz do Arari, São Sebastião da Boa Vista, Salvaterra e Soure (Figura 9). A distribuição da população do Marajó apresenta pequeno predomínio da população rural com cerca de 60 %, de acordo com o IBGE, 2010, enquanto a média do país é de 16 %. Dos 16 municípios, apenas três possuem

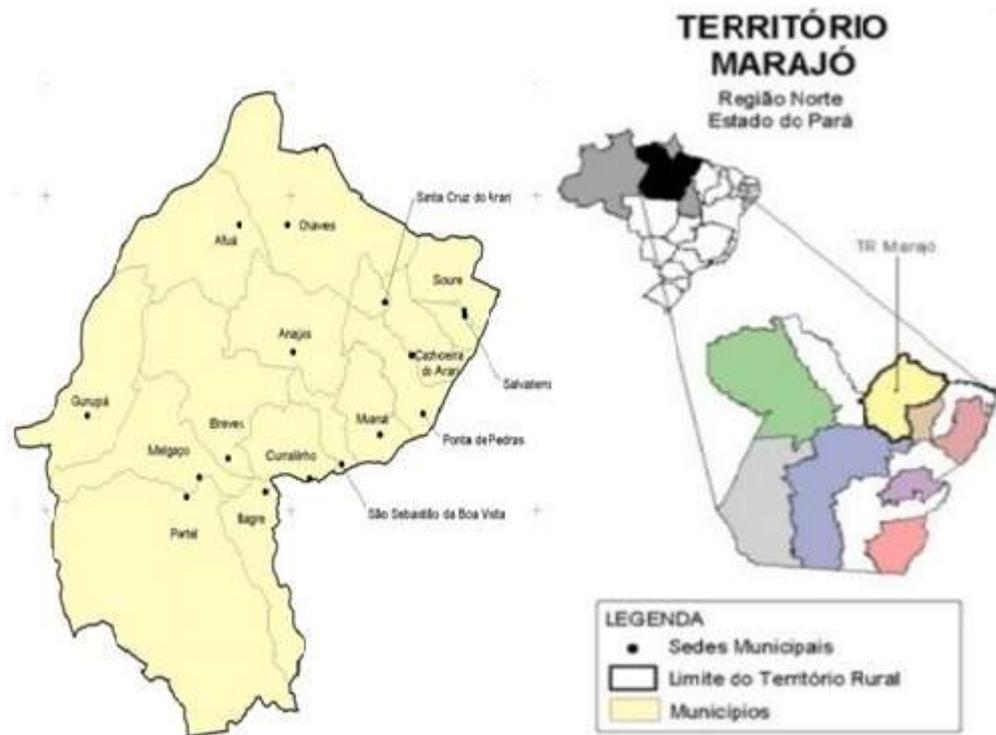
população urbana superior à rural: os municípios de Breves, Soure e Salvaterra (RELATÓRIO ANALÍTICO DO TERRITÓRIO DO MARAJÓ, 2012). Os dois últimos se destacam por serem os mais populosos, com 17000 e 22000 habitantes, respectivamente, sendo a população total da ilha de 250.000 habitantes (www.infoescola.com/geografia/ilha-de-marajo).

A teoria mais aceita sobre a origem do nome Marajó faz menção às observações de índios nativos da ilha, que a denominaram de “Mibaraió”, e que em língua tupi significa “anteparo do mar” ou “tapamar”. No período colonial era denominada Ilha Grande de Joannes, recebendo o nome atual no século XIX, à época da independência (MORIM, 2014).

O Instituto do Patrimônio Histórico e Artístico Nacional (IPHAN) identifica, através de sítios arqueológicos encontrados no Marajó, os sambaquis, indícios da mais antiga ocupação da ilha. Essas formações são enormes montanhas erguidas em baías, praias ou na foz de grandes rios por povos que habitaram o litoral do Brasil na Pré-História e constituem-se principalmente por cascas de moluscos – a própria origem tupi da palavra sambaqui significa “amontoado de conchas”. Além desses, os achados arqueológicos apontam para sociedades relativamente autônomas, organizadas em pequenas vilas familiares, vivendo de uma economia de subsistência.

Posteriormente, sociedades complexas que se caracterizaram pelo manejo de terra e de recursos hídricos espalharam-se por toda a ilha, especialmente na área de campos, junto a cabeceiras de rios e igarapés, mas ocupando também a área de floresta. Os sítios arqueológicos coloniais remontam à época do contato com os europeus. São vilas, igrejas, engenhos, fazendas, chalés, com estruturas arquitetônicas e outras evidências materiais datadas do período colonial. Esses remanescentes testemunham um longo processo histórico ocorrido na ilha, e podem auxiliar na compreensão de sua dinâmica cultural, como por exemplo, sobre as formas de contatos interétnicos entre os povos nativos, europeus e africanos (RELATÓRIO ANALÍTICO DO TERRITÓRIO DO MARAJÓ, 2012).

Figura 9 - O Território da Ilha de Marajó



Fonte: Secretaria de Desenvolvimento Territorial/ Ministério de Desenvolvimento Agrário.

Registros da ocupação humana no Arquipélago do Marajó são escassos, havendo relatos da chegada dos primeiros conquistadores, os jesuítas, em 1615, a partir da embocadura do Rio Amazonas. Nas décadas seguintes, a Coroa Portuguesa enviou observadores ao arquipélago, que encontraram ali, surpreendentemente, numerosas fazendas de gado e intensa lavoura em posse dos missionários. A partir daí copiosos relatórios sobre o arquipélago eram enviados ao rei por diversos estudiosos, entre eles o naturalista brasileiro Dr. Alexandre Rodrigues Ferreira (PEREIRA, 1956). Não existem registros escritos pelos jesuítas sobre o comportamento e o modo de vida dos seus habitantes pré-coloniais no trabalho no campo, no lar, isolada ou coletivamente. Assim, a necessidade de se realizar pesquisas sistemáticas ou mesmo o controle e a proteção para a preservação dos sítios arqueológicos existentes na região são de extrema importância (RELATÓRIO ANALÍTICO DO TERRITÓRIO DO MARAJÓ, 2012), a fim de se compreender esses aspectos etno-sociais.

Acredita-se que cerca de 30 diferentes nações indígenas habitavam o Marajó à época da colonização portuguesa. Esses povos, chamados genericamente de “Nheengaiabas” (em Tupi quer dizer "gente de fala incompreensível"), falavam línguas diferentes e localizavam-se

na parte central da ilha. Dentre as nações localizadas no território marajoara, estão: Aruãns, Sacacas, Marauanás, Caiás, Araris, Anajás, Muanás, Mapuás, Pacajás, entre outras (PACHECO, 2010). Apesar de terem grande importância para a formação da população do Marajó, ao final do século XVIII, uma grande parte da população indígena foi dizimada devido a “guerras” com os portugueses, ou havia fugido das expedições portuguesas que escravizavam os índios duramente em lavouras e cidades, tomando para si suas terras (MORIM, 2014). O genocídio dos povos marajoaras, entre eles os Anajás, foi de extrema violência, não deixando, por exemplo, informações sobre as línguas por eles faladas (glotocídio). Os colonizadores, para sanar os problemas relacionados ao multilinguismo amazônico, deram início à imposição do uso da Língua Geral Amazônica (LGA), que era baseada no tupinambá falado pelos índios (BATISTA; NOGUEIRA, 2016; FREIRE, 2003).

Nesse período a região amazônica integrou-se ao mercado mundial através da exploração mercantil e as fazendas e engenhos do século XVIII e XIX continuaram a utilizar-se largamente do trabalho de escravos trazidos da África e homens livres, estes últimos indígenas e mestiços. O Estado do Pará, em 1751, possuía 24 engenhos, classificados de “reais”, por serem completos e bem equipados. Em meados do século XIX, Marajó contava com cerca de 2000 escravos negros. Mas a decadência da produção do açúcar pela predileção da produção de aguardente, que era mais lucrativo, levou a região da Amazônia a entrar no Ciclo da Borracha, retrocedendo ao extrativismo predatório. Levas migratórias de colonos lusos se estabelecem na foz do Rio Amazonas, provenientes da África e das ilhas dos Açores. Começa assim a colonização efetiva do arquipélago, onde o gado vai representar papel relevante: os currais de influência indígena são construídos para o rebanho à medida que os fazendeiros avançam para o interior da ilha (MIRANDA-NETO, 1976).

Dentre alguns estudiosos, o paraense José Veríssimo, publicou relatos sobre as populações indígenas e mestiças da Amazônia, em “Scenas da Vida Amazônica” (VERÍSSIMO, 1899). Nele relatou suas observações sobre os processos de miscigenação que se dera na Ilha Grande de Joannes. Segundo esse crítico "o primeiro colono foi polígamo, as escravas índias faziam um harém aos soldados da conquista. Posteriormente as escravas negras também viveriam esse mesmo tipo de situação. Segundo ele, primeiramente o branco cruzou com o índio, depois o negro com este, e com aquele, e com os resultados destes sucessivos cruzamentos; daí resultou a grande mistura de “sangues” que produziu o curiboca (branco e índio), o mameluco (curiboca e branco), o mulato (branco e preto) e o cafuzo (preto e índio) e ainda outros, de entrelaçamentos destes (PEREIRA, 1956).

A resistência à escravidão mediante fugas deu origem à formação dos quilombos nas várias regiões do arquipélago. Documentos históricos mostram que no decorrer do século XVIII foram muitas as situações e movimentos de fugas da população escravizada, composta tanto por negros quanto por índios. A população de negros, indígenas e mestiços na ilha, a essa altura, correspondia a mais de 80 % da população local. Atualmente, os vaqueiros e capatazes dos latifúndios, que foram passados por herança aos atuais proprietários, descendem, em sua maioria, de antigos escravos que passaram tecnicamente à condição de agregados e dependentes após 1888 (RELATÓRIO ANALÍTICO DO TERRITÓRIO DO MARAJÓ, 2012).

Estudos de genética populacional podem contribuir significativamente para um melhor conhecimento dessa população quanto à contribuição dos diferentes grupos étnicos envolvidos em sua formação. Embora a caracterização genética da população da Ilha de Marajó ofereça interessantes possibilidades ao nível do estudo de sua origem e história, há muito conhecimento ainda por extrair, uma vez que não se encontra na literatura dados consistentes sobre suas origens, o nível de diversidade genética e subestrutura populacional.

Até a presente data apenas um trabalho foi publicado no sentido de avaliar a ancestralidade da população da Ilha de Marajó, com base em marcadores moleculares. Palha e colaboradores (2011) analisaram nove marcadores STR do cromossomo Y de 300 indivíduos de nove comunidades quilombolas da região amazônica (n=300). Dessas comunidades uma era do município de Ponta de Pedras, na Ilha de Marajó (n=58). Para o conjunto das populações quilombolas estudadas, foi observada uma elevada diversidade haplotípica (0,989), com uma maior percentagem de linhagens africanas (53,6 %), seguida pelas europeias (41,4 %) e com uma minoria de linhagens nativo americanas (5 %). Para os 58 indivíduos representantes do Marajó, a diversidade haplotípica foi menor (0,942), tendo sido observadas somente linhagens europeias (61,9 %) e africanas (38,1 %) (PALHA *et al.*, 2011).

1 OBJETIVOS

1.1 Objetivo geral

No intuito de contribuir para o conhecimento da estruturação gênica e da ancestralidade da população brasileira, o presente trabalho tem como objetivo contribuir para o conhecimento acerca da ancestralidade genética da população da Ilha de Marajó, no estado do Pará, situado à região Norte, através do estudo com marcadores de DNA de origem uniparental (cromossomo Y e mtDNA) e também com marcadores autossômicos informativos de ancestralidade (AIMs).

Além disso, pretende-se elaborar uma base de dados da população da Ilha de Marajó, para esses marcadores, que poderá ser utilizada em posteriores estudos de genética populacional, bem como em análises forenses e em estudos clínicos que tenham essa população como referência.

Pretende-se, ainda, comparar os dados obtidos para esses marcadores genéticos com dados obtidos para outras populações brasileiras, a fim de aprofundar o conhecimento sobre a estruturação gênica do Brasil.

1.2 Objetivos específicos

De maneira a alcançar o objetivo geral, estipularam-se os seguintes objetivos específicos:

- a) Caracterizar as linhagens masculinas com marcadores genéticos do tipo Y-STR e Y-SNP, em indivíduos não aparentados nativos da Ilha de Marajó;
- b) Estimar as frequências dos haplótipos de Y-STR na amostra populacional e contribuir com novos dados para o banco de dados haplotípicos YHRD;
- c) Estimar as frequências de haplogrupos do cromossomo Y, relacionando-os à sua origem geográfica – europeia, africana ou nativa americana através da análise de marcadores SNPs;

- d) Caracterizar as linhagens maternas através do sequenciamento da região controle completa do DNA mitocondrial, em indivíduos não aparentados nativos da Ilha de Marajó;
- e) Estimar as frequências dos haplótipos de mtDNA na amostra populacional e contribuir com a elaboração de um banco de dados de haplótipos do DNA mitocondrial da população da Ilha de Marajó;
- f) Estimar as frequências de haplogrupos do DNA mitocondrial, relacionando-os à sua origem geográfica – europeia, africana ou nativa americana;
- g) Comparar os dados obtidos de marcadores Y-STR e mtDNA na população da Ilha de Marajó com aqueles de outras populações brasileiras já estudadas e de populações mundiais, através da análise das distâncias genéticas entre pares de populações;
- h) Caracterizar, com marcadores autossômicos informativos de ancestralidade (AIMs) do tipo inserção/deleção (Indel), indivíduos não aparentados nativos da Ilha de Marajó, utilizando um multiplex contendo 46 *loci*;
- i) Criar uma base de dados para os AIM-Indels e inferir a ancestralidade dos indivíduos analisados, determinando as contribuições relativas dos principais grupos étnicos ancestrais formadores da população brasileira – nativo americanos, africanos e europeus;
- j) Comparar os dados obtidos de para os AIM-Indels na população da Ilha de Marajó com aqueles de outras populações brasileiras já estudadas;
- k) Comparar os resultados obtidos para as ancestralidades paterna (cromossomo Y), materna (mtDNA) e global (Indels autossômicos) com os dados históricos dos padrões de miscigenação das populações do Brasil.

2 MATERIAIS E MÉTODOS

Este estudo foi realizado no Laboratório de Diagnósticos por DNA (LDD), do Departamento de Ecologia, Instituto de Biologia Roberto Alcântara Gomes, Universidade do Estado do Rio de Janeiro (UERJ). A coleta das amostras biológicas foi realizada em colaboração com o pesquisador Dr. Luiz Marcelo de Lima Pinheiro, do Laboratório de Biologia na Saúde da Universidade Federal do Pará – Campus Soure, nos municípios de Soure e Salvaterra.

2.1 Amostra populacional

Ao longo deste trabalho, foi analisada uma amostra populacional de indivíduos pertencentes a ambos os sexos, não aparentados, residentes na Ilha de Marajó, no estado do Pará, a fim de contribuir para a caracterização genética da mesma.

Amostras biológicas de 251 participantes foram obtidas através da coleta de manchas de sangue periférico em papel Whatman FTA (Sigma-Aldrich Co.), sendo armazenadas à temperatura ambiente até a extração do DNA. Todos os participantes consentiram na utilização de suas amostras biológicas e o uso de seu material genético, de forma não identificada, em projetos de pesquisa, ao assinar um Termo de consentimento livre e esclarecido (Anexo A) no ato da coleta. O emprego destas amostras em estudos de Genética populacional foi aprovado pelo Comitê de Ética da UERJ, parecer CAAE: 0067.0.228.000-09 (Anexo B). As coletas aconteceram nos municípios de Soure e Salvaterra.

Foram coletados dados sobre o local de nascimento dos doadores, assim como de seus pais e avós maternos e paternos. Essas informações nos serviram para atender aos critérios estabelecidos para seleção da amostra nas análises a serem realizadas (Tabela 1), visando um aprofundamento do conhecimento sobre a história da ancestralidade genética da população da Ilha de Marajó, a qual será referida como “população autóctone”. Por essa razão procurou-se buscar uma ancestralidade de no mínimo três gerações, de forma a eliminar a influência de migrações recentes para a Ilha de Marajó nos resultados a serem obtidos para os marcadores de ancestralidade. Por outro lado, todas as amostras coletadas serão utilizadas para a

caracterização do perfil genético da população atual residente na Ilha de Marajó. Esta informação possibilitará a criação de bancos de dados relevantes para fins forenses ou para estudos clínicos.

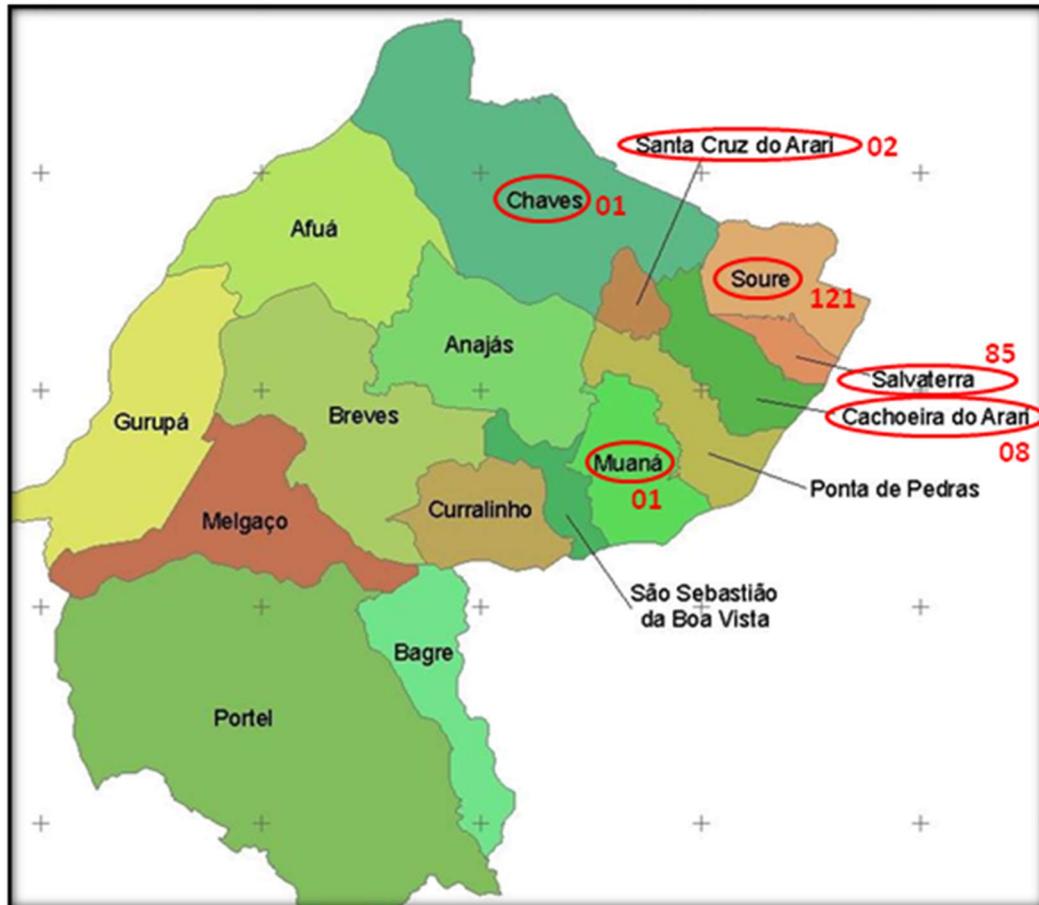
Tabela 1 - Critérios estabelecidos na seleção de indivíduos para o estudo da população autóctone da Ilha de Marajó

Categoria do marcador	Critério	Número de indivíduos selecionados / número total
Cromossomo Y	Indivíduos do sexo masculino com avô paterno nascido na Ilha de Marajó	97/ 251
DNA mitochondrial	Indivíduos de ambos os sexos com avó materna nascido na Ilha de Marajó	241/ 251
AIMs	Indivíduos de ambos os sexos com avós maternos e paternos nascidos na Ilha de Marajó	185/ 251

Legenda: Tipo de marcador a ser estudado, com o critério de seleção de no mínimo 3 gerações e o número de indivíduos que tiveram amostras coletadas e atenderam a esses critérios.

Do total de 251 amostras denominadas de população autóctone, apenas 218 indivíduos nasceram na Ilha de Marajó e 33 nasceram em outros lugares (outros municípios do Estado do Pará - 30, Amapá - 1, Amazonas - 1, Guiana Francesa - 1), apesar de residirem e terem avós naturais do Marajó, atendendo assim aos critérios apresentados na Tabela 1. Os 218 indivíduos nativos da Ilha de Marajó têm como local de nascimento os municípios indicados na Figura 10, todos pertencentes à parte leste da ilha.

Figura 10 - A Ilha de Marajó e seus 16 municípios, com destaque para os municípios de origem de 218 indivíduos da população autóctone



Legenda: Municípios de nascimento de 218 indivíduos autóctones participantes do estudo de ancestralidade da Ilha de Marajó e o número de amostras coletadas de cada município.

Fonte: <http://sit.mda.gov.br/download.php?ac=obterDadosBas&m=1501808>

2.2 Extração de DNA genômico

A extração do DNA das amostras foi realizada no LDD, na UERJ, através de três metodologias distintas: a extração por resina Chelex 100 (BioRad), o método orgânico, com o uso de fenol-clorofórmio/álcool isoamílico; e a extração com o *Kit QIAamp DNA Investigator* (Qiagen).

Embora as eficácias desses métodos de extração possam ser comparáveis, resultando na amplificação do DNA extraído, verificamos que preparações de DNA de amostras extraídas pela resina quelante Chelex apresentaram menor durabilidade, mostrando-se

instáveis em ampliações posteriores. Por essa razão, posteriormente foram realizadas extrações pelo método orgânico, assim como pelo *Kit QIAamp DNA Investigator* (Qiagen).

2.2.1 Extração de DNA - método orgânico - fenol-clorofórmio / álcool isoamílico

Um fragmento de aproximadamente 5 mm² de papel Whatman FTA (Sigma-Aldrich Co.), contendo mancha de sangue, foi picotado com o uso de um bisturi e depositado em microtubos de 1,5 mL. Foram adicionados 1 mL de tampão SSC 1x (NaCl 0,15 M e C₆H₅Na₃O₇.2 H₂O 0,015 M, pH 7) a cada tubo, seguido de agitação em vortex (VX38, Warmnest) e de centrifugação por 5 min a 12000 x g (Centrífuga Centrimicro, modelo 242 – Fanem). O sobrenadante foi descartado com o auxílio de uma pipeta, restando aproximadamente 30 µL no tubo. Foram acrescentados em cada tubo 400 µL de tampão de extração (tris 0,01 M; EDTANa₂ 0,01 M; NaCl 0,1 M; SDS 20 %, pH 8), 25 µL de proteinase K 10 mg/mL e 20 µL de DTT 1 M. Em seguida a amostra foi incubada a 37 °C por 18 a 24 horas em um termobloco. Após a incubação, a extração pelo fenol-clorofórmio e precipitação com etanol seguiram protocolo descrito por Sambrook e colaboradores (SAMBROOK; FRITCSH; MANIATIS, 1989), com adaptações. Foram adicionados 200 µL de fenol/clorofórmio/álcool isoamílico na proporção 25:24:1 (Invitrogen), a preparação foi homogeneizada por agitação em vórtex e centrifugada por 3 min a 12000 x g, havendo a separação da solução em duas fases. A fase aquosa superior foi transferida para um novo tubo de 1,5 mL e foram adicionados 200 µL de clorofórmio/álcool isoamílico (24:1), seguido de homogeneização por agitação em vórtex e centrifugação por 3 min a 12000 x g, separando, novamente, a preparação em duas fases. A fase aquosa (superior) foi transferida para um novo tubo de 1,5 mL e foram adicionados 1 mL de etanol absoluto recém retirado do *freezer* (-20 °C) e 35 µL de acetato de sódio 3 M, seguindo-se agitação no vórtex e incubação a -20 °C por, pelo menos, 16 horas. Em seguida, a amostra foi centrifugada à temperatura de 4 °C por 15 min a 10.000 x g (Centrífuga 5418R, Eppendorf). O sobrenadante foi cuidadosamente descartado e foi adicionado 1mL de etanol 70 % , sendo realizada nova centrifugação por 5 min a 12.000 x g. O sobrenadante foi mais uma vez descartado cuidadosamente e as amostras foram deixadas à temperatura ambiente por 18 a 24 horas para evaporação do etanol residual.

Por fim, foram acrescentados 40 μL de TE (tris-HCl 5 mM, EDTA 0,1 mM pH 8), em cada tudo. As amostras foram armazenadas em *freezer* a $-20\text{ }^{\circ}\text{C}$.

2.2.2 Extração de DNA - resina Chelex

Chelex 100 (Bio-Rad) é uma resina que atua como quelante para íons metálicos polivalentes. Durante o processo de extração as células são inicialmente lisadas com água ultrapura. A alcalinidade da solução e a fervura também levam à ruptura das células, permitindo que os grupos quelantes se liguem aos componentes celulares, liberando a molécula do DNA. A extração com Chelex é um bom método, pois é rápido, não requer múltiplas transferências de tubos e não utiliza solventes orgânicos tóxicos como o método com fenol-clorofórmio; porém não consegue remover totalmente inibidores (como a hemoglobina) que podem ser prejudiciais em etapas posteriores (PHILLIPS; MCCALLUM; WELCH, 2012).

A extração de DNA a partir das amostras de sangue colhidas foi realizada de acordo com a técnica desenvolvida por Walsh e colaboradores (WALSH; METZGER; HIGUCHI, 1991). Um fragmento de aproximadamente 5 mm^2 de papel Whatman FTA (Sigma-Aldrich Co.), contendo mancha de sangue, foi picotado com o uso de um bisturi e depositado em microtubos de 1,5 mL. Foi adicionado 1 mL de água ultrapura e as amostras foram deixadas à temperatura ambiente por 30 min. Após este período foram centrifugadas a 12000 x g por 5 min (centrífuga Centrimicro, modelo 242 – Fanem). O sobrenadante foi descartado cuidadosamente com o auxílio de uma pipeta e foram adicionados 200 μL de Chelex 100 a 5 % sob agitação manual, utilizando uma ponteira com a extremidade cortada. Posteriormente, esta preparação foi agitada em vórtex e submetida a duas incubações sucessivas: a primeira a $56\text{ }^{\circ}\text{C}$, por 20 min e a segunda a $100\text{ }^{\circ}\text{C}$, por 8 min. Após as incubações, as amostras foram novamente centrifugadas a 12000 x g , por 5 min e armazenadas em *freezer* a $-20\text{ }^{\circ}\text{C}$. As centrifugações foram realizadas na centrífuga Centrimicro, modelo 242 – Fanem.

2.2.3 Extração de DNA - Kit QIAamp DNA Investigator

Uma parte das amostras utilizadas teve o DNA extraído com o *kit QIAamp DNA Investigator*, que envolve quatro etapas principais: ruptura das membranas celulares usando uma combinação de atividade enzimática e lise mecânica (aquecimento e agitação); ligação do DNA à membrana de sílica da coluna; lavagem de contaminantes através da membrana usando tampões e a eluição do DNA. Esse método de extração com membrana contendo sílica é considerado bastante eficaz, sendo capaz de remover inibidores da amostra enquanto mantém um alto rendimento de DNA de boa qualidade (PHILLIPS; MCCALLUM; WELCH, 2012). Foram seguidas as instruções do protocolo do *kit* para extração do DNA em amostras de sangue em papel FTA e o volume de eluição utilizado foi de 30 μL , em água ultrapura.

2.3 **Quantificação de DNA**

As amostras extraídas com fenol-clorofórmio foram quantificadas utilizando o *kit* comercial Qubit dsDNA HS (Invitrogen), seguindo as instruções do fabricante. A faixa de detecção deste *kit* varia de 0,2-100 ng/ μL . Para calibrar o instrumento, dois padrões, *Standard # 1* (0 ng/ μL) e *Standard # 2* (10 ng/ μL), foram preparados e medidos no instrumento. As amostras a serem quantificadas foram preparadas por combinação de 195 μL de solução reagente e 5 μL de DNA e posteriormente foi feita a conversão do valor de concentração obtido na leitura, levando-se em consideração essa diluição.

2.4 **Análise de marcadores do Cromossomo Y**

A análise molecular de marcadores do cromossomo Y foi realizada em 97 amostras de indivíduos não aparentados pertencentes ao sexo masculino, residentes e que possuem avôs paternos nascidos na Ilha de Marajó.

A determinação dos haplótipos do cromossomo Y foi realizada mediante o emprego dos 25 marcadores STR incluídos no *kit* comercial *Yfiler Plus*, que correspondem a 27 *loci*,

uma vez que os marcadores DYS385 e DYF387S1 correspondem a 2 *loci* cada (*Applied Biosystems*).

A determinação dos haplogrupos presentes nessa amostra foi realizada mediante o emprego de 47 Y-SNPs selecionados de modo a permitirem discriminar os principais haplogrupos que se espera encontrar em populações miscigenadas da América do Sul, como os de origem nativo-americana, europeia e africana. Essa seleção foi feita a partir dos haplótipos Y-STR, através da predição do haplogrupo mais provável, com os *software online Haplogroup Predictor* (<http://www.hprg.com/hapest5/>, acessado em setembro de 2019) e *NevGen* (<http://www.NevGen.org>, acesso em outubro de 2019). Com base nesse resultado, foram genotipados os Y-SNPs relevantes para a sua confirmação, por reações do tipo SBE (do inglês, *Single Base Extension*) ou extensão de base única, técnica também conhecida por minisequenciamento.

2.4.1 Genotipagem dos *loci* STR do cromossomo Y utilizando o kit *Yfiler Plus*

Para a análise de *loci* Y-STR, o *kit* comercial *Yfiler Plus* (*Applied Biosystems*) foi utilizado, seguindo a metodologia indicada pelo fabricante. Este *kit* consiste em um sistema multiplex de 6 fluoróforos incorporados nos *primers* por meio de ligantes não-nucleotídicos, que permitem uma amplificação e separação eficientes de 27 *loci* Y-STR em uma única reação de PCR. São utilizados cinco fluoróforos nas amostras (6-FAM, VIC, NED, TAZ e SID), que emitem fluorescências nas cores azul, verde, amarelo, vermelho e roxo, respectivamente, permitindo a análise de diferentes marcadores através da distinção de alelos de *loci* diferentes com o mesmo tamanho. O sexto fluoróforo, que consiste no LIZ, de cor laranja, é utilizado no padrão interno de tamanho, o *GeneScan 600 LIZ v2.0* (*Thermo Fisher Scientific*). Na Tabela 2 estão indicados os 27 *loci* STR, os fluoróforos utilizados para cada marcador e os alelos incluídos no padrão alélico que compõe o *kit*.

Na reação foram utilizados aproximadamente 1-2 η g de DNA extraído pelo método orgânico ou 0,5 μ L (DNA extraído por Chelex ou *Kit QIAamp DNA Investigator*), 2,0 μ L de tampão *Yfiler Plus Master Mix*, 1,0 μ L de *Yfiler Plus Primer Set* e água ultrapura autoclavada para completar o volume final de 5,0 μ L de reação. Juntamente com as alíquotas de DNA dos indivíduos das amostras populacionais estudadas, também foi amplificado o DNA controle 007 como controle positivo das reações. Como controle negativo foi preparada uma reação

sem o acréscimo de DNA, sendo o volume do mesmo substituído por água ultrapura autoclavada.

O preparo da PCR foi realizado em uma cabine esterilizada por luz ultravioleta de comprimento de onda de 254 nm e as reações foram realizadas em termociclador automático modelo *Veriti 96-Well Thermal Cycler (Applied Biosystems)*, segundo condições termocíclicas indicadas pelo fabricante do *kit*.

O volume de 1 µL dos produtos da amplificação foi aplicado em placa a ser colocada no analisador genético ABI 3500 (*Applied Biosystems*), juntamente com uma mistura contendo 8,8 µL de Formamida Hi-Di (*Applied Biosystems*) e 0,2 µL de padrão interno de tamanho *GeneScan 600 LIZ dye Size Standard v2.0 (Applied Biosystems)*. Em um dos poços, a essa mistura foi acrescentado 1 µL de *Yfiler Plus Allelic Ladder*. Em seguida, a placa foi levada ao sequenciador para a realização da eletroforese em capilares de 50 cm de comprimento preenchidos com o polímero POP-7 (*Applied Biosystems*).

Por fim, os resultados obtidos são representados graficamente por um eletroferograma no qual se observam picos resultantes do processo de emissão, detecção e tradução do sinal luminoso emitido pelos fluoróforos, cuja intensidade é medida em RFUs, através do programa *GeneMapper v. 4.1 (Thermo Fisher Scientific)*, que também faz a nomeação dos alelos ao comparar os alelos detectados com os alelos do padrão alélico *Yfiler Plus Allelic Ladder*.

2.4.2 Predição de haplogrupos - *software Haplogroup Predictor e NevGen*

A fim de evitar genotipagens desnecessárias, a escolha de marcadores Y-SNP utilizados na genotipagem das amostras estudadas foi direcionada por meio dos *software online Haplogroup Predictor e NevGen*, que indicam os prováveis haplogrupos de uma determinada amostra com base na análise de seus haplótipos Y-STR.

Tabela 2 - 27 loci do kit *Yfiler Plus*

<i>Locus</i>	Alelos incluídos no padrão alélico do kit <i>Yfiler Plus</i>	Fluoróforos	DNA controle 007
DYS576	10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25	6-FAM	19
DYS389I	9, 10, 11, 12, 13, 14, 15, 16, 17		13
DYS635	15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30		24
DYS389II	24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35		29
DYS627	11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27		21
DYS460	7, 8, 9, 10, 11, 12, 13, 14	VIC	11
DYS458	11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24		17
DYS19	9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19		15
YGATAH4	8, 9, 10, 11, 12, 13, 14, 15		13
DYS448	14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24		19
DYS391	5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16		11
DYS456	10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24	NED	15
DYS390	17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29		24
DYS438	6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16		12
DYS392	4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20		13
DYS518	32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49		37
DYS570	10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26	TAZ	17
DYS437	10, 11, 12, 13, 14, 15, 16, 17, 18		15
DYS385a/b	6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28		11, 14
DYS449	22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40		30
DYS393	7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18	SID	13
DYS439	6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17		12
DYS481	17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32		22
DYF387S1 a/b	30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44		35, 37
DYS533	7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17		13

Legenda: *Loci Yfiler Plus*, seus respectivos fluoróforos e alelos observados no padrão alélico e no DNA controle 007

Sendo assim, por exemplo, ao colocar no programa o haplótipo de uma determinada amostra e este ter indicado com uma probabilidade elevada (considerada maior que 90 % no presente estudo) que a amostra pode pertencer ao haplogrupo R, a genotipagem desta amostra em particular foi realizada pelo multiplex com *primers* específicos para Y-SNPs

determinantes do haplogrupo R (Multiplex R, segundo a Figura 11). Por outro lado, se os programas não puderam determinar com elevada probabilidade o haplogrupo de uma amostra ou se discordarem entre si; para esta amostra em questão foi feita a genotipagem preliminar com multiplexes que são compostos por *primers* específicos para Y-SNPs de ramos mais basais da árvore filogenética (Multiplex 1, segundo a Figura 11), a fim de direcionar uma segunda genotipagem com os Y-SNPs determinantes de determinados haplogrupos.

2.4.3 Genotipagem dos marcadores Y-SNP - kit *SNaPshot Multiplex*

Tomando como base os resultados obtidos na predição dos haplogrupos e considerando o conhecimento histórico sobre a origem da população brasileira, assim como os dados obtidos através de estudos genéticos que corroboram para esses relatos (JANNUZZI *et al.*, 2020; OLIVEIRA *et al.*, 2014; PENA; SANTOS; TARAZONA-SANTOS, 2020; RESQUE *et al.*, 2016), foi possível direcionar os marcadores Y-SNP mais interessantes para serem utilizados neste estudo. Foram selecionados 47 marcadores Y-SNP para a análise da ancestralidade das linhagens masculinas do Marajó. Tais marcadores foram agrupados de maneira hierárquica em 5 multiplexes, que caracterizam os principais haplogrupos europeus, africanos e ameríndios relacionados às linhagens masculinas brasileiras (Figura 11). Sendo assim, a genotipagem de marcadores Y-SNP neste estudo foi realizada por PCR multiplex seguida pelo sequenciamento de base única, também conhecido como SBE.

As sequências gênicas que contêm os Y-SNPs de interesse foram amplificadas pela técnica de PCR através da realização dos 5 multiplexes citados anteriormente (Figura 11). Foram utilizados aproximadamente 1-2 ng de DNA extraído pelo método orgânico ou 0,5-2 µL (DNA extraído por Chelex ou *Kit QIAamp DNA Investigator*), além de 2,5 µL do tampão *QIAGEN Multiplex PCR Master Mix 2x*, proveniente do *QIAGEN Multiplex PCR kit*, 0,5 µL de mix dos *primers* na concentração de 2 µM e água ultrapura autoclavada para completar o volume final de 5,0 µL de reação. Para cada uma das preparações foram feitos um controle negativo (reação sem o acréscimo de DNA) e um controle positivo utilizando DNA de um indivíduo previamente genotipado cujo haplogrupo possuía o estado ancestral (sem mutações) para os Y-SNPs analisados.

Para a amplificação do DNA, o preparo da PCR foi realizado em uma cabine esterilizada por luz ultravioleta de comprimento de onda de 254 nm e as reações foram

realizadas em termociclador automático modelo *Veriti 96-Well Thermal Cycler (Applied Biosystems)*. As condições termocíclicas foram descritas por Gomes e colaboradores (2010) e são apresentadas na Tabela 3.

Tabela 3 - Condições termocíclicas para a amplificação dos multiplexes para o estudo dos Y-SNPs

Etapa	Ciclos	Temperatura (° C)	Tempo
Desnaturação inicial	1	95	15 min
Desnaturação	35	94	30 s
Anelamento		60	1 min 30 s
Extensão		72	1 min
Extensão final	1	72	10 min

Legenda: Etapas utilizadas na reação de amplificação, com número de ciclos, suas respectivas temperaturas e durações.

Os conjuntos de iniciadores utilizados para cada SNP, assim como os tamanhos de fragmentos obtidos em pares de base e a reação de PCR multiplex a qual pertencem estão descritos na Tabela 4.

Para a avaliação inicial dos produtos de amplificação, foi realizada eletroforese em gel de poliacrilamida (29:1 acrilamida:bis-acrilamida), a uma concentração de 10 %. As amostras foram preparadas adicionando-se 2 µL do produto de amplificação a 1 µL de *Safer Dye* (Kasvi) e posteriormente aplicadas no gel e em uma cuba vertical, a 5,5 V/cm, por 1 h 30 min, em tampão de corrida TBE 1x (tris 89 mM, ácido bórico 89 mM, EDTA 2 mM, pH 8). Também foi aplicado no gel o marcador de tamanho molecular *GeneRuler 50 bp DNA Ladder (Thermo Fisher Scientific)*, composto de uma mistura de fragmentos de DNA com comprimentos entre 50 a 1000 pb. Após a eletroforese, o gel foi visualizado em um transiluminador de luz azul (modelo K33-333, Kasvi), sendo possível a identificação dos fragmentos de interesse por comparação com o marcador de peso molecular.

A etapa seguinte foi a realização da reação multiplex SBE – de extensão de base única, utilizando-se o *kit SNaPshot Multiplex (Applied Biosystems)*. Porém, antes da reação SBE, os produtos amplificados foram submetidos à purificação com enzima USB ExoSAP-IT (*Thermo Fisher Scientific*), com o objetivo de remover oligonucleotídeos e dNTPs não incorporados durante a reação de amplificação (1 µL de produto de PCR e 0,5 µL de enzima USB ExoSAP-IT). A purificação foi realizada em termociclador automático modelo *Veriti 96-Well*

Thermal Cycler (Applied Biosystems), a 37° C por 30 min e, em seguida, a 80 °C por 15 min, para inativação da enzima.

Para o preparo das reações SBE foram adicionados diretamente aos produtos de purificação 1 µL de *SNaPshot Multiplex Ready Reaction Mix (Thermo Fisher Scientific)* e 2,5 µL de mix de *primers* SBE (Tabela 5). O marcador 12f2, do Multiplex GIJ, não foi incluído nessa tabela porque seu polimorfismo é detectado pelo tamanho do fragmento gerado na PCR, uma vez que se trata de uma inserção do tipo Alu. As reações SBE foram então submetidas às condições termocíclicas de 25 ciclos a 96 °C por 10 s, 50 °C por 5 s e 60 °C por 30 s em termociclador *Veriti 96-Well Thermal Cycler (Applied Biosystems)*, segundo as instruções do fabricante. Após a reação de SBE, esta foi purificada adicionando-se 1 µL de enzima *SAP (Thermo Fisher Scientific)*, a fim de remover ddNTPs não incorporados. A seguir, essa preparação foi incubada a 37 °C por 1 h e a 85 °C por 15 min. O volume de 1 µL dos produtos da purificação foram aplicados em placa a ser colocada no analisador genético ABI 3500 (*Applied Biosystems*) para a eletroforese em capilar de 50 cm preenchido com o polímero POP-7 (*Applied Biosystems*), juntamente com uma mistura contendo 8,8 µL de Formamida Hi-Di (*Applied Biosystems*) e 0,2 µL de padrão interno de tamanho *GeneScan 120 LIZ dye Size Standard v2.0 (Applied Biosystems)*.

Os resultados obtidos do minisequenciamento foram representados graficamente por um eletroferograma no qual se observam picos resultantes do processo de emissão, detecção e tradução do sinal luminoso emitido pelos fluoróforos, cuja intensidade é medida em RFUs, através do programa *GeneMapper v. 4.1 (Thermo Fisher Scientific)*. Uma vez que a cada nucleotídeo que possa ser incorporado na reação SBE está associado um fluorófo específico, informado pelo fabricante do *kit SNaPshot Multiplex (Applied Biosystems)*, é possível detectar, pela região de emissão de energia do fluoróforo, representada por cor específica informada pelo fabricante do *kit*, qual é a variante alélica para cada SNP selecionado.

Tabela 4 - Conjuntos de *primers* das PCRs multiplexes para o estudo dos Y-SNPs (continua)

Y-SNP	Amplicon (pb)	<i>Primer</i> direto (5'→3')	<i>Primer</i> reverso (5'→3')	Referência	PCR Multiplex
M9	340	GCAGCATATAAAACTTTTCAGG	AAAACCTAACTTTGCTCAAGC	BRION <i>et al.</i> , 2005	MULTIPLEX 1 EUROPEU (BRION <i>et al.</i> , 2005)
M173	172	GCACAGTACTCACTTTAGGTTTGC	GCAGTTTTCCCAGATCCTGA	BRION <i>et al.</i> , 2005	
SRY1532	167	TCCTTAGCAACCATTAATCTGG	AAATAGCAAAAACCTGACACAAGGC	BRION <i>et al.</i> , 2005	
M213	145	GGCCATATAAAAACGCAGCA	* <i>Primer</i> SBE reverso	BRION <i>et al.</i> , 2005	
P25	121	GGACCATCACCTGGGTAAAGT	AGTGCTTGTCCAAGGCAGTA	BRION <i>et al.</i> , 2005	
Tat	112	GACTCTGAGTGTAGACTTGTGA	GAAGGTGCCGTAAAAGTGTGAA	BRION <i>et al.</i> , 2005	
M22	106	GCTGATAGTCCTGGTTTCCCTA	TGAGCATGCCTACAGCAGAC	BRION <i>et al.</i> , 2005	
M70	81	TCATAGCCCACTATACTTTGGAC	CTGAGGGCTGGACTATAGGG	BRION <i>et al.</i> , 2005	
92R7	55	TGCATGAACACAAAAGACGTA	GCATTGTTAAATATGACCAGC	BRION <i>et al.</i> , 2005	
M62	309	ACTAAAACACCATTAGAAACAAAGG	CTGAGCAACATAGTGACCCC	BRION <i>et al.</i> , 2005	
M267	256	CGTTGTCCCTGTGTTTCCAT	CTGTTGCCCAGGCTAGTGTC	NOGUEIRO <i>et al.</i> , 2010	
M172	187	TCCTCATTACCTGCCTCTC	TCCATGTTGGTTTGGAACAG	BRION <i>et al.</i> , 2005	
P58	180	ACAGGAGGCCATAATGCAAC	GAGCCTCACACCTTCTCTG	SIMÃO <i>et al.</i> , 2021	
M170	158	TGCAGCTCTTATTAAGTTATGTTTCA	CCAATTACTTTCAACATTTAAGACC	BRION <i>et al.</i> , 2005	
M201	144	TCAAATTGTGACACTGCAATAGTT	CATCCAACACTAAGTACCTATTACGAA	BRION <i>et al.</i> , 2005	
12f2	90	CACTGACTGATCAAAAATGCTTACAGAT	GGATCCCTTCTTACACCTTATACA	BRION <i>et al.</i> , 2005	
P2	180	GCTCCAGCCATCTTTTCCTTA	CTTCTCTCATGAGGGTTTTGGA	GOMES <i>et al.</i> , 2010	MULTIPLEX E (GOMES <i>et al.</i> , 2010)
M293	230	AAAGAGATTGATCGGTGCATA	GCTGGCTAATACTTCCACAGAG	GOMES <i>et al.</i> , 2010	
M154	130	TACTCACACAAACCAAGAAGAAACA	AACCATTGTGTTACATGGCCTA	GOMES <i>et al.</i> , 2010	
M81	203	GCACTATCATACTCAGCTACACATCTC	AACCATTGTGTTACATGGCCTA	GOMES <i>et al.</i> , 2010	
M85	283	TGGCATCCAATACTAGCTGATAAAC	AATGCTCACGCTTGTGTTCT	GOMES <i>et al.</i> , 2010	
M78	235	GGATGGCTGTATGGGTTTCT	ATAGTGTTCCCTTACCTTTTCTT	GOMES <i>et al.</i> , 2010; BRION <i>et al.</i> , 2005	
M35	198	GCATGGTCCCTTTCTATGGAT	GAGAATGAATAGGCATGGGTTT	BRION <i>et al.</i> , 2005	
M96	88	GTGATGTGTAACCTGGAAAACAGG	GGACCATATATTTGCCATAGGTT	BRION <i>et al.</i> , 2005	
V6	102	GATGGCACAGTGTTTCGACAG	CTTCTCTCCAAATGCCTGCT	GOMES <i>et al.</i> , 2010	
M2	162	AAGTCCAGACCCAGGAAGGT	ACAGCTCCCCCTTTATCCTC	GOMES <i>et al.</i> , 2010	
M123	213	TGCTCTCAGGGGAAAATCTG	AGCAAAGTTGAGGTTGCACA	GOMES <i>et al.</i> , 2010	
M191	122	AAAAATGGAGTGTTTATCAGAGCTT	CCCAGACACACCAAAAATATCTC	GOMES <i>et al.</i> , 2010	
M33	190	CACAACCTTCATTGGCTACGG	GTTGAAGCCCCCAAGAGAGAC	GOMES <i>et al.</i> , 2010	

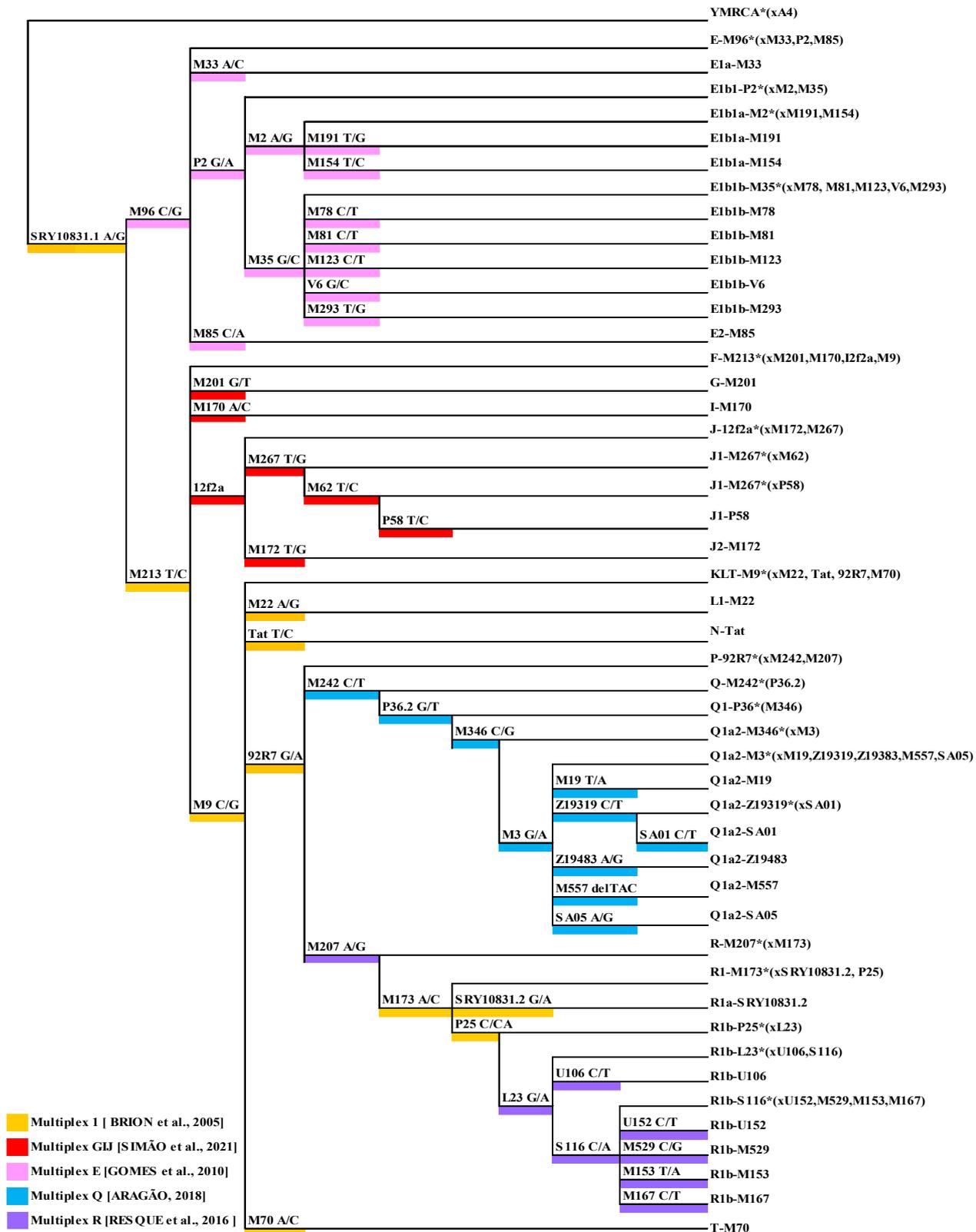
Tabela 4 - Conjuntos de *primers* das PCRs multiplexes para o estudo dos Y-SNPs (conclusão).

Y-SNP	Amplicon(pb)	<i>Primer</i> direto (5'→3')	<i>Primer</i> reverso (5'→3')	Referência	PCR Multiplex
SA01 M557	370	AAGATCCCACCACTGCACTC	CTCTGGCCCCTAACAAACCT	ARAGÃO, 2018	MULTIPLEX Q (ARAGÃO, 2018)
M3	304	CATTAAAGCCGGTCACAGGT	CTGCCAGGGCTTTCAAATAG	NOGUERA <i>et al.</i> , 2014	
P36.2	299	GAGGAGGGGGAGAGAGAAAA	TTCAAACAGCCCACCAGATA	NOGUERA <i>et al.</i> , 2014	
M19	277	TCACCAGAGTTTCAAATAG	ACAGACACAAAGGGCCAACCT	NOGUERA <i>et al.</i> , 2014	
M346	247	GGCCTGAAAATGTGGAAAGA	AGCCTGAAAATGTGGAAAGA	NOGUERA <i>et al.</i> , 2014	
SA05	236	GAACCAAAGCACAGCACTCA	ATGCTCATGGCCTACACCTC	ARAGÃO, 2018	
Z19483	211	CCATGTAGGAGGAGGCAAAA	CATCACAAAAGCCAAAAGCA	ARAGÃO, 2018	
Z19319	163	TTTGCTGAAGTTGCCTGTCA	AGTTCCAGTCAGGGCAATCA	ARAGÃO, 2018	
M242	155	TTGTGCAAAAAGGTGACCAA	TTTCGCTTTAAGGGCTTTCA	NOGUERA <i>et al.</i> , 2014	
M153	76	TTCTCAGACACCAATGGTCCTA	TCTGACTTGGAAAGGGGAAA	RESQUE <i>et al.</i> , 2016	MULTIPLEX R (RESQUE <i>et al.</i> , 2016)
M167	131	GAGGCTGGGCCAAGTTAAG	CTTCTCGGAACCACTACCA	RESQUE <i>et al.</i> , 2016	
U152	163	GAAACATTCCACGCTTGAGG	AGCCTCTTTTTGGCTTCCAT	RESQUE <i>et al.</i> , 2016	
U106	192	TCCTGAATAGCAAATCCCAAAG	AATGGCAGAGGTAGGAGGAAAT	RESQUE <i>et al.</i> , 2016	
M529	228	GCCCCAAAACAACAGAATA	GGAAGCATTGAGAAGCAGGT	RESQUE <i>et al.</i> , 2016	
L23	229	ACACAGTGAAACCCCGTCTC	AAGATTGTGGGGACAAAGGA	RESQUE <i>et al.</i> , 2016	
S116	241	TCAGTCAGGGCAAATCTGAA	GGTGGAGTTGGGGCTAAAGT	RESQUE <i>et al.</i> , 2016	
M207	322	CGTTACAACCTATGGGGCAA	TCCTCTCTGAAATGCCGAAT	RESQUE <i>et al.</i> , 2016	

Legenda: Nesta tabela, estão listados os marcadores Y-SNP analisados neste estudo, o tamanho dos produtos de amplificação gerados em pares de bases (pb), as seqüências dos *primers* direto e reverso no sentido 5'→3', os artigos de referência para cada conjunto de *primer* e as reações de PCR multiplex às quais os marcadores foram agrupados e amplificados.

Nota: Para o marcador M213, utilizou-se o *primer* SBE reverso na reação de PCR (*cagctgtgaaagtctgacaaTCAGAACTTAAAACATCTCGTTAC), pois o *primer* reverso da PCR confeccionado tinha um problema na seqüência.

Figura 11 - Marcadores SNP do cromossomo Y usados neste trabalho, em destaque na árvore filogenética do cromossomo Y



Legenda: Conjunto de marcadores Y-SNP utilizados nas amostras deste estudo. As cores indicam cada multiplex e os respectivos marcadores amplificados e ao fim de cada ramo da árvore se encontra o haplogrupo determinado por esses marcadores.

Fonte: A autora.

Tabela 5 - Conjuntos de *primers* das reações SBE multiplexes para o estudo dos Y-SNPs (continua)

Y-SNP	Produto de SBE (pb)	Sequência (5'→3')	Referência	Multiplex SBE
M9 SBE	48	GTGCCACGTCGTGAAAGTCTGACAAGAAACGGCCTAAGATGGTTGAAT	BRION <i>et al.</i> , 2005	MULTIPLEX 1 EUROPEU (BRION <i>et al.</i> , 2005)
M173 SBE	34	GTCTGACAACCTACAATTCAAGGGCATTTAGAAC	BRION <i>et al.</i> , 2005	
SRY1532	30	TCTGACAATTGTATCTGACTTTTTTCACACAGT	BRION <i>et al.</i> , 2005	
M213 SBE	45	CACGTCGTGAAAGTCTGACAATCAGAACTTAAAACATCTCGTTAC	BRION <i>et al.</i> , 2005	
P25 SBE	26	CAATCTGCCTGAAACCTGCCTG	BRION <i>et al.</i> , 2005	
Tat SBE	42	TCGTGAAAGTCTGACAACCTGAAATATTAATTAACAAC	BRION <i>et al.</i> , 2005	
M22 SBE	18	GCCATTCCTGGTGGCTCT	BRION <i>et al.</i> , 2005	
M70 SBE	34	TCTGACAATAGGGATTCTGTTGTGGTAGTCTTAG	BRION <i>et al.</i> , 2005	
92R7 SBE	28	AAGCATGAACACAAAAGACGTAGAAG	BRION <i>et al.</i> , 2005	
M62 SBE	27	CTGACAACAATGTTTGTGGCCATGGA	BRION <i>et al.</i> , 2005	MULTIPLEX GIJ (SIMÃO <i>et al.</i> , 2021)
M267 SBE	37	TGAAAGTCTGACAACCTCCACACAAAATACTGAAMGT	NOGUEIRO <i>et al.</i> , 2010	
M172 SBE	18	AAACCCATTTTGATGCTT	BRION <i>et al.</i> , 2005	
P58 SBE	40	CACGTCGTGAAAGTCTGACAATGACATTTGTGTGCTTTGC	SIMÃO <i>et al.</i> , 2021	
M170 SBE	22	ACACAACCCACACTGAAAAAAA	BRION <i>et al.</i> , 2005	
M201 SBE	34	CCCCCCCCCCCCCGATCTAATAATCCAGTATCAACTGAGG	BRION <i>et al.</i> , 2005	
P2 SBE	16	GCCCCTAGGAGGAGAA	GOMES <i>et al.</i> , 2010	MULTIPLEX E (GOMES <i>et al.</i> , 2010)
M293 SBE	21	AAAGAGATTGATCGGTGCATA	GOMES <i>et al.</i> , 2010	
M154 SBE	26	AAACATGGCCTATAATATTCAGTACA	GOMES <i>et al.</i> , 2010	
M81 SBE	27	CCCCCTAAATTTTGTCTTTTTTGAA	BRION <i>et al.</i> , 2005	
M85 SBE	30	CTTGTGTTCTATTAAGTGTAGTTTTGTTAG	GOMES <i>et al.</i> , 2010	
M78 SBE	34	CCCCCCCCCACAACCTTAACAAAGATACTTCTTTC	BRION <i>et al.</i> , 2005	
M35 SBE	36	CCCCCCCCCCCCCCCCCAGTCTCTGCCTGTGTC	BRION <i>et al.</i> , 2005	
M96 SBE	40	CCCCCCCCCGTAACTTGAAAACAGGTCTCTCATAATA	BRION <i>et al.</i> , 2005	
V6 SBE	42	GCCACGTCGTGAAAGTCTGACAATGCTGTGATTCTGATGTG	GOMES <i>et al.</i> , 2010	
M2 SBE	45	GTGCCACGTCGTGAAAGTCTGACAATTTATCCTCCACAGATCTCA	GOMES <i>et al.</i> , 2010	
M123 SBE	48	TAGGTGCCACGTCGTGAAAGTCTGACAATCTAGGTATTCAGGCGATG	GOMES <i>et al.</i> , 2010	
M191 SBE	51	GGTGCCACGTCGTGAAAGTCTGACAACATTTTTTTCTTTACAACCTTGACTA	GOMES <i>et al.</i> , 2010	
M33 SBE	54	GTGCCACGTCGTGAAAGTCTGACAACAGTTACAAAAGTATAATATGTCTGAGAT	GOMES <i>et al.</i> , 2010	

Tabela 5 - Conjuntos de *primers* das reações SBE multiplexes para o estudo dos Y-SNPs (conclusão)

Y-SNP	Produto de SBE (pb)	Sequência (5'→3')	Referência	Multiplex SBE
SA01 SBE	38	gtcgtgaaagtctgacaaTTTGTCAAGTGTAGAGTGG	ARAGÃO, 2018	MULTIPLEX Q (ARAGÃO, 2018)
M557 SBE	45	tgccacgtcgtgaaagtctgacaaGAACAGGGTTGCAAACGGTA	ARAGÃO, 2018	
M3 SBE	17	TCACCTCTGGGACTGA	NOGUERA <i>et al.</i> , 2014	
P36.2 SBE	35	gtcgtgaaagtctgacaaCATCTATCTATCCATTATTCTCTCT	NOGUERA <i>et al.</i> , 2014	
M19 SBE	28	tgacaaGTAGAGACATCTGAAACCCAC	NOGUERA <i>et al.</i> , 2014	
M346 SBE	28	ctgacaaCAGCCAAGAGGACAGTAAGA	NOGUERA <i>et al.</i> , 2014	
SA05 SBE	48	aggtgccacgtcgtgaaagtctgacaaATGTTTCTAGGGTGAGCCTGT	ARAGÃO, 2018	
Z19483 SBE	42	acgtcgtgaaagtctgacaaATAAGCTGTCTGGCTATTTCA	ARAGÃO, 2018	
Z19319 SBE	40	tcgtgaaagtctgacaaCCATCATCTCAACCTAAAATCC	ARAGÃO, 2018	
M242 SBE	23	aaAAAAAGGTGACCAAGGTGCT	NOGUERA <i>et al.</i> ,	
M153 SBE	23	AAAGCTCAAAGGGTATGTGAACA	RESQUE <i>et al.</i> , 2016	MULTIPLEX R (RESQUE <i>et al.</i> , 2016)
M167 SBE	16	AAGCCCCACAGGGTGC	RESQUE <i>et al.</i> , 2016	
U152 SBE	32	CAAGGATAAGAAAAATGAGTATTGTGAAAATA	RESQUE <i>et al.</i> , 2016	
U106 SBE	28	TCTGACAATAGCAAATCCCAAAGCTCCA	RESQUE <i>et al.</i> , 2016	
M529 SBE	23	AATAACAACCGCTCTCTCAGACA	RESQUE <i>et al.</i> , 2016	
L23 SBE	18	GCGACAGAGCGAGACTCT	RESQUE <i>et al.</i> , 2016	
S116 SBE	35	GAAAGTCTGACAAGAGTTGGGGCTAAAGTGAAAG	RESQUE <i>et al.</i> , 2016	
M207 SBE	28	AACAAATGTAAGTCAAGCAAGAAATTTA	RESQUE <i>et al.</i> , 2016	

Legenda: Nesta tabela, estão listados os *primers* SBE utilizados neste estudo, suas sequências no sentido 5'→3', os artigos de referência para cada conjunto de *primer*, o tamanho dos produtos de minissequenciamento gerados e as reações de SBE multiplex às quais os marcadores foram agrupados e amplificados.

Nota: Para o marcador M22, o *primer* SBE utilizado foi modificado em duas bases ([CCdel] GCCATTCTGGTGGCTCT) com relação ao *primer* referência, descrito por Brion *et al.*, 2005.

2.4.4 Análise estatística dos resultados do Cromossomo Y

Com base nos dados obtidos pela genotipagem dos 27 loci Y-STRs com *Yfiler Plus*, foram realizadas estimativas de parâmetros de diversidade genética intrapopulacional, como as frequências haplotípicas, determinação do número de haplótipos compartilhados e cálculo da diversidade haplotípica (HD), utilizando o *software* Arlequin versão 3.5 (EXCOFFIER; LISCHER, 2010).

O mesmo *software* foi utilizado nos cálculos de distâncias genéticas (F_{ST}) e probabilidades de não-diferenciação entre a amostra estudada e outras 12 populações miscigenadas com dados previamente publicados para *Yfiler Plus*, representativas das regiões geopolíticas brasileiras (Tabela 6). Além dessas, foram comparadas populações de referência de origens africana [Kenya (n=62) - Bantu_Luhya, outro - YA004206] (IACOVACCI *et al.*, 2017), europeia [Madri, Espanha (n=126) - YA003147] (MARTÍN *et al.*, 2004) e nativa Americana [Oxapampa, Peru (n=58) - Ashaninka - YA004112] (TINEO *et al.*, 2015) com dados disponíveis no banco de dados YHRD.

Como não existem dados de *Yfiler Plus* disponíveis para outras populações da região Norte do Brasil, nova análise foi realizada, calculando os valores de diversidade para um conjunto parcial de 18 marcadores para os quais existem dados publicados por Palha e colaboradores (2012) e Purps e colaboradores (2014). Este conjunto incluiu os seguintes Y-STR: DYS19, DYS385, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS570, DYS576, DYS635 e YGATAH4. As distâncias genéticas (F_{ST}) e probabilidades de não-diferenciação entre a amostra estudada e essas populações (Tabela 7) para esse conjunto de marcadores também foram calculadas com o *software* Arlequin v. 3.5.

Nos cálculos das distâncias genéticas foram excluídos os marcadores DYS385 e DYS389I, e o número de repetições do marcador DYS398I foi subtraído de DYS398II, visto que esses marcadores apresentam mais de um alelo em um haplótipo. Além disso, alelos nulos, microvariantes, deleções, duplicações e triplicações existentes foram codificados como dados em falta e simbolizados com “?”.

Para visualização das distâncias genéticas pareadas obtidas, gráficos de escalonamento multidimensional (MDS) foram construídos através do programa *Statistica Software* v. 14.0.0.15 (TIBCO Software Inc.).

As proporções de haplogrupos europeus, africanos e Ameríndios na amostra analisada foram calculadas por contagem direta no Excel (*Microsoft Corporation*).

Tabela 6 - Populações brasileiras utilizadas para comparação com a população da Ilha de Marajó, com base nos dados obtidos pela genotipagem dos 27 Y-STRs - *Yfiler Plus*

População/ Região	Número de indivíduos	Etnia	Referência
Marajó (N)	97	Miscigenada	Este trabalho
Maranhão (NE)	296	Miscigenada	JANNUZZI <i>et al.</i> , 2020
Piauí (NE)	42	Miscigenada	JANNUZZI <i>et al.</i> , 2020
Ceará (NE)	38	Miscigenada	JANNUZZI <i>et al.</i> , 2020
Rio Grande do Norte (NE)	19	Miscigenada	JANNUZZI <i>et al.</i> , 2020
Pernambuco (NE)	141	Miscigenada	JANNUZZI <i>et al.</i> , 2020
Mato Grosso (CO)	100	Miscigenada	JANNUZZI <i>et al.</i> , 2020
Goiás (CO)	204	Miscigenada	JANNUZZI <i>et al.</i> , 2020
Rio de Janeiro (SE)	258	Miscigenada	JANNUZZI <i>et al.</i> , 2018
São Paulo (SE)	169	Miscigenada	JANNUZZI <i>et al.</i> , 2020
Espírito Santo (SE)	409	Miscigenada	JANNUZZI <i>et al.</i> , 2020; STANGE <i>et al.</i> , 2019
Espírito Santo-pomeranos (SE)	79	Miscigenada	STANGE <i>et al.</i> , 2019
Rio Grande do Sul (S)	211	Miscigenada	JANNUZZI <i>et al.</i> , 2020

Legenda: N - Região Norte, NE - Região Nordeste, CO - Região Centro-Oeste, SE - Região Sudeste, S - Região Sul.

Tabela 7 - Populações da Região Norte do Brasil utilizadas para comparação com a população da Ilha de Marajó, com base nos dados obtidos pela genotipagem de 18 Y-STRs

População	Número de indivíduos	Etnia	Referência
Marajó	97	Miscigenada	Este trabalho
Belém	400	Miscigenada	PALHA <i>et al.</i> , 2012
Santarém	72	Miscigenada	PALHA <i>et al.</i> , 2012
Macapá	65	Miscigenada	PALHA <i>et al.</i> , 2012
Rio Branco	31	Miscigenada	PALHA <i>et al.</i> , 2012
Porto Velho	135	Miscigenada	PALHA <i>et al.</i> , 2012
Boa Vista	67	Miscigenada	PALHA <i>et al.</i> , 2012
Palmas	30	Miscigenada	PALHA <i>et al.</i> , 2012
Manaus	74	Miscigenada	PALHA <i>et al.</i> , 2012
São Gabriel da Cachoeira	61	Nativa Americana	PURPS <i>et al.</i> , 2014

2.5 Análise de marcadores do DNA mitocondrial

A determinação dos haplótipos de mtDNA foi realizada através de sequenciamento da Região Controle (RC - entre as posições nucleotídicas 16024 e 576) para 101 indivíduos com avós maternas nascidas na Ilha de Marajó. A determinação dos haplogrupos aos quais pertencem as linhagens maternas dessa população foi feita com base nos haplótipos obtidos por comparação das sequências com a rCRS, através do *software SeqScape v2.7*. Na classificação dos haplótipos seguiram-se as diretrizes propostas pela ISFG (PARSON *et al.*, 2014).

2.5.1 Genotipagem da Região Controle do mtDNA

Para a análise da RC completa do mtDNA foi realizado o sequenciamento pelo método de Sanger (SANGER; NICKLEN; COULSON, 1977), com o *kit BigDye Terminator V3.1 Cycle Sequencing (Applied Biosystems)*.

Foram utilizados os *primers* L15997 e H639 a uma concentração de 0,2 μ M, com aproximadamente 1-5 ng de DNA extraído pelo método orgânico e 2,5 μ L de *QIAGEN Multiplex PCR Master Mix (QIAGEN)*, num volume final de de 5,0 μ L a ser completado com água ultrapura autoclavada. No caso de amostras de DNA extraído por Chelex ou *Kit QIAamp DNA Investigator*, o volume de amostra utilizado no PCR foi de 0,5 a 2 μ L. Juntamente com as alíquotas de DNA das amostras, foi preparada uma reação em que o volume de DNA foi substituído por água ultrapura autoclavada, como controle negativo e uma reação com amostra de DNA previamente utilizada e que tinha bom rendimento de amplificação com os *primers* utilizados.

Para a amplificação do DNA, o preparo da PCR foi realizado em uma cabine esterilizada por luz ultravioleta de comprimento de onda de 254 nm e as reações foram realizadas em termociclador automático modelo *Veriti 96-Well Thermal Cycler (Applied Biosystems)*. As condições termocíclicas são apresentadas na Tabela 8 e as sequências dos *primers* na Tabela 9.

Algumas amostras de DNA falharam em amplificar a RC com esse par de *primers*, que gera um fragmento de 1211 pb. Por existir a possibilidade dessas amostras estarem

parcialmente degradadas, usamos um protocolo alternativo para a obtenção da sequência completa, que amplifica fragmentos menores. Nesse caso a amplificação foi realizada em 2 fragmentos, denominados RC1 e RC2, definidos pelos pares de *primers* L15967 - H20 (622 pb) e L16475 - H639 (733 pb), respectivamente (Tabela 9).

As condições de amplificação para esse protocolo foram com os *primers* a uma concentração de 0,5 μ M, BSA 0,16 μ g/ μ L, com aproximadamente 2-4 ng de DNA extraído pelo método orgânico ou pelo *Kit QIAamp DNA Investigator* (QIAGEN) e 5,0 μ L de *QIAGEN Multiplex PCR Master Mix* (QIAGEN), num volume final de de 10,0 μ L, a ser completado com água ultrapura autoclavada. No caso de amostras de DNA extraído por Chelex ou não quantificadas, o volume de amostra utilizado no PCR foi de 1 a 2 μ L. As condições termocíclicas são apresentadas na Tabela 10.

Tabela 8 - Condições termocíclicas para a amplificação da Região Controle do mtDNA com o *QIAGEN Multiplex PCR Master Mix*

Etapa	Ciclos	Temperatura (°C)	Tempo
Desnaturação inicial	1	95	10 min
Desnaturação	35	94	30 s
Anelamento		60	1 min e 30 s
Extensão		72	1 min
Extensão final	1	72	10 min

Legenda: Etapas utilizada na reação de amplificação da região controle do mtDNA, com número de ciclos, suas respectivas temperaturas e durações.

Para a avaliação inicial dos produtos de amplificação, foi realizada eletroforese em gel de agarose a uma concentração de 1,2 % em tampão de corrida TAE 1x (tris 40 mM, ác. acético 20 mM, EDTA 1 mM). As amostras foram preparadas adicionando-se 2 μ L do produto de amplificação a 1 μ L de *Safer Dye* (Kasvi) e posteriormente aplicadas no gel e em uma cuba horizontal, a 5,5 V/cm, por 30 min. Também foi aplicado no gel o marcador de tamanho molecular *GeneRuler 50 bp DNA Ladder* (*Thermo Fisher Scientific*), composto de uma mistura de fragmentos de DNA com comprimentos entre 50 a 1000 pb.

Após a eletroforese, o gel foi visualizado em um transiluminador de luz azul (modelo K33-333, Kasvi), sendo possível a identificação dos fragmentos de interesse por comparação com o marcador de peso molecular. De acordo com a intensidade da banda obtida em uma avaliação visual, foi possível estipular a quantidade de produto de PCR na reação de sequenciamento. No caso do protocolo alternativo (RC1 e RC2), apenas as amostras com os 2 fragmentos amplificados foram sequenciadas.

Os produtos de PCR foram purificados a partir da adição de 2 μ L do reagente *illustra ExoProStar* (*GE Healthcare*) a 5 μ L de cada amostra amplificada, deixando, em seguida, a

mistura reagir no termociclador *Veriti 96-Well Thermal Cycler (Applied Biosystems)* durante 30 min, sendo a 37 °C nos primeiros 15 min e 80 °C nos 15 min finais. Essa etapa, que retira *primers* e nucleotídeos não incorporados na amplificação, antecede a etapa da reação de sequenciamento.

Tabela 9 - *Primers* usados na amplificação e no sequenciamento da Região Controle do mtDNA

<i>Primer</i>	Sentido	Sequência	Referência
L15967	Direto	5'- GTC TTT AAC TCC ACC ATT AGC ACC- 3'	NGUIDI, 2021
L15997	Direto	5'- CAC CAT TAG CAC CCA AAG CT - 3'	WILSON <i>et al.</i> , 1995
L16475	Direto	5'- TAGCTAAAGTGAAGTGTATCC- 3'	NGUIDI, 2021
L16268	Direto	5'- CAC TAG GAT ACC AAC AAA CC - 3'	GABRIEL <i>et al.</i> , 2001
L16555	Direto	5'- CCC ACA CGT TCC CCT TAA AT - 3'	SIMÃO <i>et al.</i> , 2018
L314	Direto	5'-CCGCTTCTGGCCACAGCACT-3'	SIMÃO <i>et al.</i> , 2018
H016	Reverso	5'- CCC GTG AGT GGT TAA TAG GGT - 3'	EDSON <i>et al.</i> , 2004
H020	Reverso	5'- AGCTCCCGTGAGTGGTTAATA- 3'	NGUIDI, 2021
H159	Reverso	5'- AAATAATAGGATGAGGCAGGAATC - 3'	GABRIEL <i>et al.</i> , 2001
H388	Reverso	5'- GTTTAAGTGCTGTGGCCAGAAG-3'	SIMÃO <i>et al.</i> , 2018
H484	Reverso	5'-TGAGATTAGTAGTATGGGAG-3'	EDSON <i>et al.</i> , 2004
H639	Reverso	5'- GGG TGA TGT GAG CCC GTC TA - 3'	SIMÃO <i>et al.</i> , 2018

Legenda: Nome dos *primers* diretos e reversos e suas respectivas sequências

Tabela 10 - Condições termocíclicas para a amplificação dos fragmentos RC1 e RC2 do mtDNA com o *QIAGEN Multiplex PCR Master Mix*

Etapa	Ciclos	Temperatura (°C)	Tempo
Desnaturação inicial	1	95	11 min
Desnaturação	36	95	1 min
Anelamento		60-RC1; 55-RC2	1 min
Extensão		72	1 min
Extensão final	1	72	10 min

Legenda: Etapas utilizada na reação de amplificação dos fragmentos RC1 e RC2 da região controle do mtDNA, com número de ciclos, suas respectivas temperaturas e durações.

O sequenciamento foi efetuado através do *kit BigDye Terminator V3.1 Cycle Sequencing (Applied Biosystems)*, segundo instruções do fabricante. Em casos que não foi atingida dupla cobertura no sequenciamento com os *primers* usados na PCR, os fragmentos foram submetidos a uma ou mais reações com diferentes *primers*, no sentido direto ou reverso, que possibilitaram a cobertura de toda a região controle. Ou seja, dependendo do estado de cada amostra, em relação à qualidade das sequências, com posições duvidosas, ou com presença de heteroplasmias de comprimento e/ou de posição, foram feitos sequenciamentos com outros *primers*, em ambos os sentidos, até que não houvesse dúvida em

qualquer das posições (PARSON; BANDELT, 2007). Nas amostras em que foi observada heteroplasmia de comprimento associada a inserções AC entre as posições 513 e 525 e/ou pela transição na posição 460, o sequenciamento foi também obtido com os *primers* L314, H338 e H484. Os *primers* usados no sequenciamento estão listados na Tabela 9.

As reações de sequenciamento foram realizadas em termociclador automático modelo *Veriti 96-Well Thermal Cycler (Applied Biosystems)*. Todos os *primers* foram usados a uma concentração de 0,25 µM e a quantidade do produto de PCR purificado a ser utilizado na reação variou de 0,5 a 2,0 µL, dependendo da intensidade da banda correspondente no gel de agarose.

Após a reação de sequenciamento, o produto foi purificado em colunas contendo resina Sephadex G-50 Medium (*GE Healthcare Life Sciences*), que faz a separação das moléculas de ácidos nucleicos de tamanhos diferentes por exclusão. O produto da reação de sequenciamento foi cuidadosamente, adicionado sobre cada coluna, e os tubos foram centrifugados a 3879 x g por 3 min. Ao produto purificado, coletado em microtubo após eluição da coluna, foram adicionados 10 µL de Formamida Hi-Di (*Applied Biosystems*), agente desnaturante. A solução resultante foi aplicada em placa para a eletroforese capilar no Analisador Genético ABI 3500 (*Applied Biosystems*), que realiza a separação dos fragmentos por eletroforese em capilares de 50 cm de comprimento preenchidos com polímero POP-7 (*Applied Biosystems*) e detecta os nucleotídeos incorporados, marcados com um fluoróforo específico.

A determinação dos haplótipos da região controle do mtDNA foi realizada através da comparação das sequências obtidas com a rCRS (*revised Cambridge Reference Sequence*), através do *software SeqScape v2.7 (Applied Biosystems)*. Na determinação dos haplótipos seguiu-se as diretrizes propostas pela ISFG (PARSON *et al.*, 2014).

2.5.2 Classificação dos haplogrupos de mtDNA e análise estatística

Os haplogrupos foram classificados com auxílio com a ajuda da base de dados EMPOP database v4/R13 (PARSON; DÜR, 2007) e confirmados manualmente na *Phylotree - mtDNA tree Build 17*, atualizada em Fevereiro de 2016 (VAN OVEN; KAYSER, 2009). Os haplótipos serão submetidas à base de dados EMPOP, para efeitos de controle de qualidade.

As proporções de haplogrupos europeus, africanos e ameríndios na amostra analisada foram calculadas por contagem direta no Excel (*Microsoft Corporation*) e comparadas com outras populações brasileiras anteriormente analisadas incluídas na Tabela 11 e com populações da Região Norte do Brasil (NOGUEIRA *et al.*, 2017).

Através da consulta de dados da literatura, baseados nas sequências da RC de algumas populações brasileiras e da América do Sul (Tabela 11), foram obtidos valores referentes ao número de haplótipos na população e sua diversidade haplotípica (HD). Para outras populações, das quais tais dados não estavam disponíveis, esses valores foram calculados juntamente com a população da Ilha de Marajó, através dos haplótipos publicados, utilizando o *software* Arlequin versão 3.5 (EXCOFFIER; LISCHER, 2010). Essas populações foram: Colômbia, Equador, Paraguai, Peru e São Paulo.

O mesmo *software* foi utilizado nos cálculos de distâncias genéticas (F_{ST}) e probabilidades de não-diferenciação entre a amostra estudada e as outras populações com dados previamente publicados para a Região Controle completa (Tabela 11).

O alinhamento das sequências para as análises no *software* Arlequin foi realizado no *software* Haplosearch (FREGEL; DELGADO, 2011) com posterior correção de alinhamento da EMPOP (PARSON; DÜR, 2007). Uma vez que as sequências homopoliméricas que comumente ocorrem nas posições 16183-16194, 302-310 e 568-573 do mtDNA, assim como as repetições diméricas entre 513-525, rotineiramente não são consideradas nas interpretações de perfis na casuística forense e em bases de dados, tais posições foram excluídas das sequências analisadas neste trabalho para fins de comparação de diversidade intra- e interpopulacional (PARSON *et al.*, 2014).

Para auxiliar na visualização das distâncias genéticas pareadas obtidas foi construído o gráfico de escalonamento multidimensional (MDS) em duas dimensões através do programa *Statistica Software* v. 14.0.0.15 (TIBCO Software Inc.).

2.6 Análise dos AIMs

O sistema utilizado neste trabalho para inferência de ancestralidade reúne 46 AIM-Indels que apresentam diferenças nas frequências alélicas entre os grupos ancestrais africano, europeu, nativo americano e asiático, sendo altamente divergentes entre pelo menos dois destes grupos (PEREIRA *et al.*, 2012a). Após a genotipagem das amostras foi feita a análise

de inferência de ancestralidade com o *software* STRUCTURE (PRITCHARD; STEPHENS; DONNELLY, 2000).

Tabela 11 - Populações utilizadas para comparação com a população da Ilha de Marajó, com base nos dados obtidos pelo sequenciamento da Região Controle completa

População	Número de indivíduos	Etnia	Referência
Marajó	101	Miscigenada	Este trabalho
Argentina, Norte	98	Miscigenada	BOBBILO <i>et al.</i> , 2009
Argentina, Sul	47	Miscigenada	BOBBILO <i>et al.</i> , 2009
Argentina, Centro	193	Miscigenada	BOBBILO <i>et al.</i> , 2009
Colômbia, Medellín	94	Miscigenada	AUTON <i>et al.</i> , 2015
Distrito Federal	306	Miscigenada	FREITAS <i>et al.</i> , 2019
Equador	156	Miscigenada	BAETA <i>et al.</i> , 2014; BRANDINI <i>et al.</i> , 2018
Espírito Santo, Centro	54	Miscigenada	DOS REIS <i>et al.</i> (2019)
Espírito Santo, Metropolitano	81	Miscigenada	DOS REIS <i>et al.</i> (2019)
Espírito Santo, Norte	17	Miscigenada	DOS REIS <i>et al.</i> (2019)
Espírito Santo, Sul	62	Miscigenada	DOS REIS <i>et al.</i> (2019)
Paraguai	537	Miscigenada	SIMÃO <i>et al.</i> , 2021
Paraná	122	Miscigenada	POLETTI <i>et al.</i> , 2018
Peru, Lima	83	Miscigenada	AUTON <i>et al.</i> , 2015
Rio de Janeiro	205	Miscigenada	SIMÃO <i>et al.</i> , 2018
Santa Catarina	80	Miscigenada	PALENCIA <i>et al.</i> , 2010
São Paulo	142	Miscigenada	PRIETO <i>et al.</i> , 2011
Venezuela, Caracas	101	Miscigenada	CASTRO DE GUERRA <i>et al.</i> , 2012

2.6.1 Genotipagem dos 46 AIM-Indels

A genotipagem dos marcadores AIM-Indels autossômicos foi realizada em 160 amostras de indivíduos não aparentados, residentes e que possuem avós e avôs maternos e paternos nascidos na Ilha de Marajó.

Para a análise dos 46 marcadores foi usado um conjunto de 46 pares de *primers* em um único sistema multiplex descrito por Pereira e colaboradores (2012a). O preparo da PCR foi realizado em uma cabine esterilizada por luz ultravioleta de comprimento de onda de 254 nm e as reações foram realizadas em termociclador automático modelo *Veriti 96-Well Thermal Cycler (Applied Biosystems)*. As condições termocíclicas são apresentadas na Tabela 12. A Tabela 13 contém a informação dos 46 marcadores em relação à sequência polimórfica e ao fluoróforo incorporados ao *primer* por meio de ligantes não-nucleotídicos. Os quatro fluoróforos utilizados (6-FAM, VIC, NED, PET) emitem fluorescência em diferentes comprimentos de onda, permitindo, assim, que os fragmentos dos 46 AIM-Indels

amplificados na mesma reação de PCR sejam separados e detectados eficientemente na eletroforese capilar.

A reação de amplificação dos marcadores AIM-Indels foi adaptada do protocolo de Pereira e colaboradores para um volume de 5 μL . Na PCR, os *primers* foram utilizados a uma concentração de 1 μM , com aproximadamente 1-2 ηg de DNA extraído pelo método orgânico ou 0,5 - 2 μL , no caso de DNA extraído por Chelex ou *Kit QIAamp DNA Investigator*. Além disso, 2,5 μL de *QIAGEN Multiplex PCR Master Mix (QIAGEN)* e água ultrapura autoclavada para completar o volume final de reação de 5,0 μL . Juntamente com as alíquotas de DNA das amostras, foi preparada uma reação em que o volume de DNA foi substituído por água ultrapura autoclavada, como controle negativo da reação e uma reação com amostra de DNA previamente utilizada e que tinha bom rendimento de amplificação para esse sistema multiplex.

O volume de 1 μL dos produtos da purificação foram aplicados em placa a ser colocada no analisador genético ABI 3500 (*Applied Biosystems*) para a eletroforese em capilar de 50 cm preenchido com o polímero POP-7 (*Applied Biosystems*), juntamente com uma mistura contendo 8,8 μL de Formamida Hi-Di (*Applied Biosystems*) e 0,2 μL de padrão interno de tamanho *GeneScan 500 LIZ dye Size Standard v2.0 (Applied Biosystems)*.

Na análise dos produtos de amplificação e nomeação dos alelos usou-se o *software GeneMapper v. 4.1 (Thermo Fisher Scientific)*.

2.6.2 Análise estatística dos resultados dos AIM-Indels

Para estimar as proporções de cada contribuição continental foi utilizado o *software STRUCTURE v2.3.4 (PRITCHARD; STEPHENS; DONNELLY, 2000)*. Foi realizada uma análise supervisionada, utilizando informações prévias sobre a origem geográfica das amostras de referência da África, da Europa e dos nativos americanos, do painel de diversidade humana HGDP-CEPH (<http://www.hagsc.org/hgdp/>) [PEREIRA *et al.*, 2012a]. As corridas de *STRUCTURE* compreenderam repetições de 100000 etapas seguidas por 100.000 iterações MCMC. Foi considerada uma contribuição tri-híbrida de ameríndios, europeus e africanos ($K = 3$), por serem sabidamente as populações ancestrais formadoras da população brasileira.

Foi utilizado o modelo *Admixture (Use population Information to test for migrants)*. As frequências de alelos foram correlacionadas e atualizadas utilizando apenas indivíduos com o parâmetro *POPFLAG = 1* (neste caso, as amostras de HGDP-CEPH utilizadas como referência – africanas, europeias e ameríndias).

Tabela 12 - Condições termocíclicas para a amplificação do multiplex AIM-Indels com o *QIAGEN Multiplex PCR Master Mix*

Etapa	Ciclos	Temperatura (°C)	Tempo
Desnaturação inicial	1	95	15 min
Desnaturação	28	94	30 s
Anelamento		60	1 min e 30 s
Extensão		72	45 s
Extensão final	1	72	10 min

Legenda: Etapas da reação de amplificação dos AIM-Indels, com ciclagem - temperaturas e durações.

Tabela 13 - Sistema multiplex 46 AIM-Indels

Marcador	Polimorfismo Indel / Fluoróforo	Marcador	Polimorfismo Indel/ Fluoróforo
MID-1470	-/GTTAC; 6-FAM	MID-2431	-/ATTG; VIC
MID-777	-/GAA; 6-FAM	MID-2264	-/AAGT; VIC
MID-196	-/CAT; 6-FAM	MID-2256	-/CAT; NED
MID-881	-/ACTT; 6-FAM	MID-128	-/ATT; NED
MID-3122	-/ATCT; 6-FAM	MID-15	-/AAATACACAC; NED
MID-548	-/CT; 6-FAM	MID-2241	-/GTCCAATA; NED
MID-659	-/CT; 6-FAM	MID-419	-/AATGGCA; NED
MID-2011	-/CTAGA; 6-FAM	MID-943	-/TGAT; NED
MID-2929	-/TA; 6-FAM	MID-159	-/CCCCA; NED
MID-593	-/TT; 6-FAM	MID-2005	-/AACAAT; NED
MID-798	-/GGGAAA; 6-FAM	MID-250	-/CA; NED
MID-1193	-/AT; 6-FAM	MID-1802	-/GGA; NED
MID-1871	-/TT; 6-FAM	MID-1607	-/TG; NED
MID-17	-/TAAC; 6-FAM	MID-1734	-/CCAT; PET
MID-2538	-/AACA; 6-FAM	MID-406	-/AG; PET
MID-1644	-/GT; 6-FAM	MID-1386	-/AAACTATTCATTTTTTACCCT; PET
MID-3854	-/TCTA; VIC	MID-1726	-/CAAGAACTATAAT/CACTATCTATTAT; PET
MID-2275	-/TCAGCAG; VIC	MID-3626	-/AATATAATTTCTCCA; PET
MID-94	-/AAC; VIC	MID-360	-/AA; PET
MID-3072	-/GCCCCCA; VIC	MID-1603	-/TTGT; PET
MID-772	-/TAG; VIC	MID-2719	-/AACT; PET
MID-2313	-/ATTATAACT; VIC		
MID-397	-/TTCT; VIC		
MID-1636	-/AA; VIC		
MID-51	-/TTTAT; VIC		

Legenda: O código MID para cada marcador, a sequência Indel polimórfica e o fluoróforo utilizado em cada par de *primer*. O código MID indica a nomenclatura dos marcadores utilizada na base de dados *Marshfield Dialelic Insertion/Deletion Polymorphisms*.

Fonte: Tabela adaptada de PEREIRA *et al.*, 2012a.

3 RESULTADOS E DISCUSSÃO

Os dados obtidos nesta tese foram divulgados de forma parcial ao longo do período do doutorado em congressos internacionais sob a forma de pôster e *proceedings* (Apêndices D e E). Os resultados referentes à genotipagem com os marcadores do cromossomo Y foram publicados parcialmente na revista *Forensic Science International: Genetics* (<https://doi.org/10.1016/j.fsigen.2019.102163>) [Apêndice F].

3.1 Características da população

As amostras foram coletadas de indivíduos com avós nascidos na Ilha de Marajó. Esse critério foi estabelecido para que os dados obtidos estejam relacionados ao menos a três gerações passadas, diminuindo efeitos devido a migrações sobre essas amostras.

3.1.1 Distribuição amostral em relação ao local de nascimento do ancestral da terceira geração para cada categoria de marcador

A distribuição, em relação aos municípios de nascimento na Ilha de Marajó, para os ancestrais de terceira geração considerados no critério de seleção da Tabela 1 (avó materna para mtDNA, avô paterno para o cromossomo Y e todos os 4 avós para os AIM-Indels), dos indivíduos cujas amostras foram analisadas, está representada nas Tabelas 14, 15 e 16.

Observa-se, pelo número de indivíduos indicado nas tabelas 15 e 16, que nem todas as amostras selecionadas pelos critérios da Tabela 1 para a ancestralidade materna com o mtDNA e para a ancestralidade global, com os marcadores autossômicos AIM-Indels, foram analisadas. Essas amostras não tiveram seu DNA amplificado satisfatoriamente para esses marcadores; provavelmente por apresentarem algum grau de degradação, resultante das condições de manuseio, transporte e armazenamento desde o momento da coleta até a realização da reação de PCR.

Tabela 14 - Distribuição, por município de nascimento na Ilha de Marajó, dos avôs paternos dos indivíduos analisados para o cromossomo Y

Município	N amostral – Y (n = 97)	Contribuição relativa
Salvaterra	45	46,40 %
Soure	42	43,30 %
Chaves	4	4,12 %
Cachoeira do Arari	4	4,12 %
Muaná	1	1,03 %
Breves	1	1,03 %

Legenda: Os municípios representados no estudo do cromossomo Y por número de indivíduos e por contribuição relativa.

Tabela 15 - Distribuição, por município de nascimento na Ilha de Marajó, das avós maternas dos indivíduos analisados para o DNA mitocondrial

Município	N amostral – mtDNA (n = 101)	Contribuição relativa
Salvaterra	46	45,55 %
Soure	43	42,57 %
Cachoeira do Arari	6	5,94 %
Chaves	3	2,97 %
Santa Cruz do Arari	1	0,99 %
Indeterminado	2	1,98 %

Legenda: Os municípios representados no estudo do mtDNA por número de indivíduos e por contribuição relativa.

Pode-se observar, através das distribuições obtidas com essa população, que a região da Ilha mais representada nesse estudo é o leste da Ilha de Marajó, onde se situam os Municípios de Soure e de Salvaterra, que são vicinais. Tal fato pode ser explicado pelo fato das coletas terem sido realizadas nesses dois municípios apenas.

Tabela 16 - Distribuição, por município de nascimento na Ilha de Marajó, dos quatro avós dos 160 indivíduos analisados para os AIM-Indels

Município	N amostral – AIMS (n = 640)	Contribuição relativa
Soure	316	49,4 %
Salvaterra	259	40,5 %
Cachoeira do Arari	30	4,7 %
Chaves	16	2,5%
Santa Cruz do Arari	8	1,2 %
Ponta de Pedras	4	0,6 %
Muaná	4	0,6 %
Breves	2	0,3 %
Indeterminado	1	0,2 %

Legenda: Os municípios representados no estudo dos AIM-Indels por número de avós dos indivíduos e por contribuição relativa. O número de 640 foi obtido multiplicando 160 (indivíduos) por 4 (avós).

Nota: O termo “indeterminado” significa que esse pode ser qualquer município da Ilha de Marajó.

3.2 Quantificação do DNA

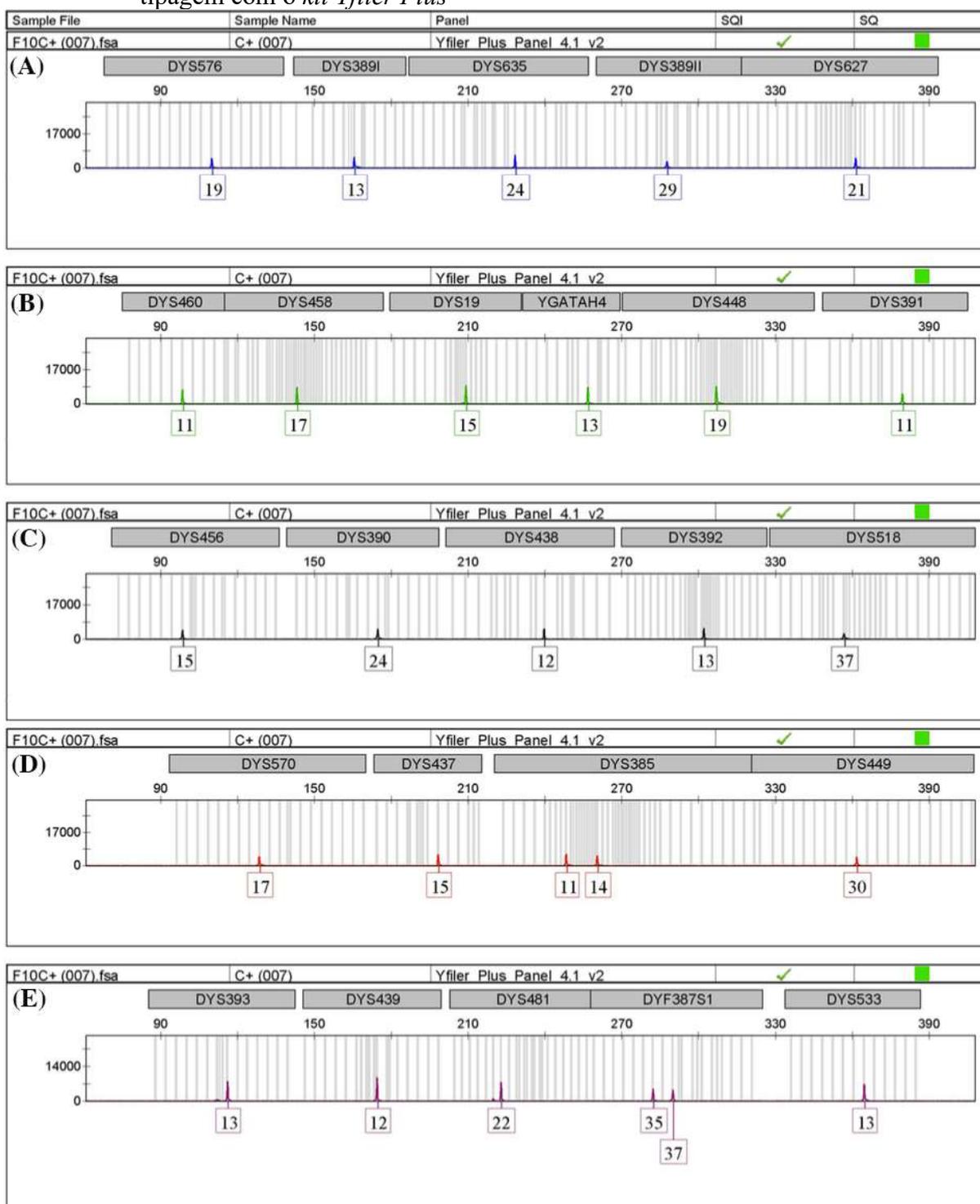
Dentre as amostras cujo DNA foi obtido pelo método orgânico, foi observada uma concentração média de $1,13 \pm 0,809$ ng/ μ L.

3.3 Análise de marcadores do Cromossomo Y

3.3.1 Haplótipos do Cromossomo Y - kit Yfiler Plus

Foram analisadas 97 amostras de indivíduos não aparentados do sexo masculino conforme descrito no item 2.4.1 de Materiais e Métodos. A Figura 12 apresenta um eletroferograma representativo de um resultado da tipagem do controle positivo 007 com o *kit Yfiler Plus*. Os pares de *primers* deste *kit* são marcados com um dos 5 fluoróforos: 6-FAM, VIC, NED, TAZ e SID, que emitem fluorescências em diferentes comprimentos de onda, traduzidos pelo *software* de análise nas cores azul, verde, amarelo, vermelho e roxo, respectivamente.

Figura 12 - Eletroferograma do perfil genético da amostra controle (007) obtido através da tipagem com o kit *Yfiler Plus*



Legenda: O kit *Yfiler Plus*, que reúne 27 loci Y-STR (A) *primers* marcados com 6-FAM; (B) *primers* marcados com VIC; (C) *primers* marcados com NED; (D) *primers* marcados com TAZ; (E) *primers* marcados com SID. O eixo X representa o tamanho dos fragmentos em pb, o eixo Y a intensidade de fluorescência (RFUs) e os picos nomeados pelos algarismos, que identificam o alelo, representam o número de repetições *in tandem* do polimorfismo STR no *locus*.

Fonte: Adaptado do *software GeneMapper v. 4.1*.

Conforme pode ser observado no eletroferograma, dos 27 marcadores amplificados, somente os marcadores DYS385 e DYF387S1 são multialélicos. O marcador DYS385 é caracterizado por apresentar 2 alelos, correspondendo à amplificação de dois *loci* com o mesmo par de *primers*. Os alelos dos 2 *loci* podem coincidir em tamanho, visualizando-se apenas um pico no eletroferograma. Já para o marcador DYF387S1 o número de cópias é muito variável, correspondendo à amplificação de vários *loci* (originados por duplicação) com o mesmo par de *primers*. Os alelos de vários *loci* podem coincidir em tamanho, visualizando-se um número menor de picos no eletroferograma do que o número de duplicações existentes para esse marcador.

O conjunto de alelos numa amostra, tipados para cada um dos marcadores deste *kit*, é denominado haplótipo. Os haplótipos encontrados nas amostras da Ilha de Marajó estão listados na Tabela do Apêndice A, assim como suas respectivas frequências. Como é possível observar, verifica-se 85 haplótipos distintos nos 97 indivíduos analisados. Desses, 78 são haplótipos únicos, quatro haplótipos são compartilhados por dois indivíduos (MRJ-059 e MRJ-264; MRJ-076 e MRJ-188; MRJ-198 e MRJ-199; MRJ-234 e MRJ-238;), um haplótipo é compartilhado por três indivíduos (MRJ-249, MRJ-250 e MRJ-251) e 2 haplótipos são compartilhados por quatro indivíduos (MRJ-025, MRJ-219, MRJ-220 e MRJ-221; MRJ-255, MRJ-304, MRJ-306 e MRJ-307).

Foi observada a ocorrência de dois alelos microvariantes: O alelo 14.2 no *locus* DYS627 (amostra MRJ-208) e o alelo 20.2 no *locus* DYS458 (amostra MRJ-202). Não foram observadas duplicações em nenhum dos *loci* tipados nessa população, exceto para os *loci* DYS385 e DYF387S1, que habitualmente possuem mais de uma cópia. A amostra MRJ-202 apresentou alelo nulo para o *locus* DYF387S1; resultado confirmado por duas análises realizadas com amplificações de DNA provenientes de extrações independentes.

Os dados de 76 haplótipos Y-STR da população analisada neste trabalho foram enviadas ao banco de dados YHRD, sob o número de acesso YA004365-1 e foram incluídos no estudo intitulado “*Male lineages in Brazilian populations and performance of haplogroup prediction tools*”, publicado por Jannuzzi e colaboradores (JANNUZZI *et al.*, 2020).

Constatou-se que todos os alelos encontrados para essa população já haviam sido depositados no YHRD, inclusive os alelos microvariantes 14.2 do *locus* DYS627 e 20.2 do *locus* DYS458. Em relação ao *locus* DYS458, estudos anteriores observaram que as microvariantes DYS458.2, como a encontrada nesse estudo, são prevalentes em relação a microvariantes DYS458.1 (MYRES *et al.*, 2009; IACOVACCI *et al.*, 2017).

Conforme citado anteriormente, uma amostra apresentou alelo nulo para o marcador DYF387S1, e essa ausência de amplificação tem 80 ocorrências no YHRD, num total de 100932 haplótipos *Yfiler Plus* presentes nesta base de dados. A ocorrência de alelos nulos provavelmente resulta de alguma mutação na sequência de anelamento do *primer* usado na reação, o que dificulta sua amplificação.

3.3.2 Predição de haplogrupos e genotipagem de Y-SNPs

Os haplogrupos mais prováveis para cada uma das amostras foi indicado pelos *software Haplogroup Predictor* e *NevGen*, para as 97 amostras analisadas com o *kit Yfiler Plus*. Essas amostras também foram genotipadas para Y-SNPs através de 5 conjuntos multiplexes assinalados na Figura 11, tomando por base o resultado da predição dos haplogrupos, segundo metodologia descrita no item 2.4.3 de Materiais e Métodos. Os haplogrupos indicados pelos *software* e os haplogrupos determinados pelos Y-SNPs e sua origem continental estão apresentados na Tabela 17.

Em relação aos resultados obtidos na predição dos haplogrupos, observa-se que ambos os *software* são informativos quanto ao haplogrupos mais provável na maioria dos casos, havendo para esse grupo de amostras uma taxa de erro de 9,27 % (amostras MRJ-008, MRJ-148, MRJ-225, MRJ-255, MRJ-273, MRJ-283, MRJ-304, MRJ-306 e MRJ-307) para o *Haplogroup Predictor* e de apenas 1,03 % (amostra MRJ-148) para o *NevGen*. O *software NevGen*, apresentou, portanto, uma menor taxa de erro, como já relatado anteriormente em populações miscigenadas da América do Sul (JANNUZZI *et al.*, 2020).

Foram encontrados 20 haplogrupos aos quais as linhagens das amostras analisadas pertencem (Tabela 18). Calculando-se as proporções de ancestralidade para esses haplogrupos, considerando sua origem continental, encontra-se maioria dos haplogrupos europeus (71 %), seguidos de africanos (17 %) e nativos (12 %). Para esse cálculo não foi considerada a origem geográfica de 3 amostras pertencentes ao haplogrupos E definido pela mutação M78, portanto o número total considerado foi para 94 amostras. Apesar de esse haplogrupos ter sua origem no leste da África, também é muito frequente na Península Ibérica, e sua inserção no Brasil pode ter acontecido tanto pela vinda dos europeus quanto dos escravos africanos, durante o período da colonização (CRUCIANI *et al.*, 2007).

O haplogrupo mais frequente foi o R-S116*(xU152,M529,M153,M167), que é um subhaplogrupo do haplogrupo principal R, definido pela mutação M269. Ele é presente nas porções oeste e sudoeste da Europa com uma frequência de, aproximadamente, 50 % na Península Ibérica e mais de 80 % no País Basco (RESQUE *et al.*, 2016; VALVERDE *et al.*, 2015). A elevada frequência deste haplogrupo (40,2 % neste estudo) já foi relatada em outras populações brasileiras analisadas (JANNUZZI *et al.*, 2020; RESQUE *et al.*, 2016), que relataram, por exemplo, 32,5 % de ocorrência para o Brasil e 34,2 % considerando a Região Norte. Esses resultados corroboram os dados históricos acerca da colonização do Brasil pelos europeus, que recebeu principalmente indivíduos portugueses provenientes da região da Península Ibérica. Em seguida, os haplogrupos europeus mais observados nesta amostra foram I-M70 e G-M20. O haplogrupo I-M170, que representa, em média, 18 % das linhagens paternas da Europa, já tendo sido relatado com frequências superiores na Escandinávia e sul da Europa (ROOTSI *et al.*, 2004). Já o haplogrupo G-M201 é o mais comum na região do Cáucaso, que engloba o leste da Europa e oeste da Ásia, chegando a atingir 70 % de frequência, que decrescem na direção leste, com menores incidências no Oriente Médio e no Sul da Europa (BERGER *et al.*, 2013).

As linhagens de origem africana mais frequentes na amostra do Marajó pertencem ao haplogrupo E-M2*(xM154,M191), seguido de E-M85 e E-M191. O haplogrupo E-M2* é representativo da África subsaariana, característico de populações Niger-Congo (GOMES *et al.*, 2010). Em outros estudos com populações brasileiras como Amazonas, Maranhão e São Paulo (CARVALHO *et al.*, 2010; JANNUZZI *et al.*, 2020), este também foi o haplogrupo subsaariano mais frequente. Sua maior prevalência ocorre no Nordeste e no Sudeste, segundo Resque e colaboradores (2016). Mas também foram observados E-M33, E-M85, E-M191, E-M154, E-M35 (DE AZEVEDO *et al.*, 2009; FIGUEIREDO *et al.*, 2016; JANNUZZI *et al.*, 2020; RESQUE *et al.*, 2016).

Relativamente ao componente nativo americano, observou-se que o haplogrupo mais frequente foi o Q-M3, mas também foram observados o Q-SA05 em 3 indivíduos e o Q-M346 em um indivíduo. O haplogrupo Q-M3* surgiu na Ásia central e migrou para as Américas. Em indivíduos nativos Sul Americanos sua prevalência é próxima a 100 %, mas outros subhaplogrupos de Q-M3 já foram descritos (GEPPERT *et al.*, 2015). Em estudos anteriormente realizados em populações brasileiras, o haplogrupo de origem nativa Americana mais frequente também foi o Q-M3, mas também se observou o Q-M346 (FIGUEIREDO *et al.*, 2016; RESQUE *et al.*, 2016).

Os resultados obtidos nas proporções de ancestralidades continentais do cromossomo Y na população da Ilha de Marajó, em comparação com dados do trabalho de Resque e colaboradores (2016), que abrange as cinco regiões geopolíticas do Brasil, têm em comum a coexistência de linhagens com diferentes origens filogeográficas e o predomínio da contribuição europeia, seguida pela africana, exceto pela Região Norte do país, que apresenta um componente nativo ligeiramente maior. Porém, a ancestralidade nativa de 12 % do Marajó é maior do que a encontrada para a Região Norte (aproximadamente 8 %) e do valor de 9,52 %, descrito para Manaus, por Carvalho e colaboradores (2010). Já em populações fora da região Norte, tais linhagens têm sido observadas com frequências bem menores (DE AZEVEDO *et al.*, 2009; OLIVEIRA *et al.*, 2014). Provavelmente a maior presença de linhagens masculinas de origem nativa americana em populações da Região Norte do Brasil se deve a diferenças nos processos históricos de colonização de cada região brasileira, considerando-se os aspectos demográficos existentes.

Tabela 17 - Resultado da determinação dos haplogrupos pelos *software Haplogroup Predictor (HP)* e *NevGen* e pela genotipagem com Y-SNPs, além da origem continental dos haplogrupos (continua)

Amostra	HP	% HP	<i>NevGen</i>	% NG	Haplogrupo - Y-SNPs	Origem continental
MRJ_008	E1b1b/G2a/I2a	56,2 % / 23,5 % / 7,1 %	J2a1 PF5191>> PF5177	11,43 %	J2-M172	EUR
MRJ_012	Q	100 %	Q M346>> Z780	53,10 %	Q1a2-M3*(xM19,Z19319,Z19383,M557,SA05)	NAM
MRJ_018	I2a1	100 %	I2a1a Sardinian M26>PF4088	96,53 %	I-M170	EUR
MRJ_023	G2a	100 %	G2a1 Z6552 > L293 >> FGC1160	96,35 %	G-M201	EUR
MRJ_025	Q	99,90 %	Q M346>> Z780	36,91 %	Q1a2-M3*(xM19,Z19319,Z19383,M557,SA05)	NAM
MRJ_036	R1b	100 %	R1b	100 %	R1b-S116*(xU152,M529,M153,M167)	EUR
MRJ_043	R1b	100 %	R1b	99,90 %	R1b-S116*(xU152,M529,M153,M167)	EUR
MRJ_044	R1b	100 %	R1b	45,74 %	R1b-S116*(xU152,M529,M153,M167)	EUR
MRJ_045	R1b	100 %	R1b	100 %	R1b-S116*(xU152,M529,M153,M167)	EUR
MRJ_047	I2a(xI2a1)	99,40 %	I2c2 Y16419	21,78 %	I-M170	EUR
MRJ_050	R1b	100 %	R1b	100 %	R1b-U152	EUR
MRJ_053	R1b	100 %	R1b	99,99 %	R1b-S116*(xU152,M529,M153,M167)	EUR
MRJ_056	I1	100 %	I1 Z58>Z59> Z60>Z140> FGC22406	26,33 %	I-M170	EUR
MRJ_059	R1b	100 %	R1b	99,99 %	R1b-S116*(xU152,M529,M153,M167)	EUR
MRJ_060	E1b1b	100 %	E1b1b V22	80,70 %	E1b1b-M78	EUR/AFR
MRJ_068	Q	100 %	Q M346>> L330	83,82 %	Q1a2-M3*(xM19,Z19319,Z19383,M557,SA05)	NAM
MRJ_071	I1	100 %	I1 Z63>>L849> S2077> S2097	40,33 %	I-M170	EUR
MRJ_072	R1b	100 %	R1b	99,93 %	R1b-S116*(xU152,M529,M153,M167)	EUR
MRJ_076	R1b	100 %	R1b	99,99 %	R1b-S116*(xU152,M529,M153,M167)	EUR
MRJ_082	R1b	100 %	R1b	99,65 %	R1b-S116*(xU152,M529,M153,M167)	EUR
MRJ_083	R1b	100 %	R1b	100 %	R1b-S116*(xU152,M529,M153,M167)	EUR
MRJ_091	E1b1a	100 %	E1b1a V38>> M4231	100 %	E1b1a-M2*(xM154,M191)	AFR
MRJ_104	G2a	100 %	G2a2 >> M278	100 %	G-M201	EUR

Tabela 17 - Resultado da determinação dos haplogrupos pelos *software Haplogroup Predictor* (HP) e *NevGen* pela genotipagem com Y-SNPs, além da origem continental dos haplogrupos (continuação)

Amostra	HP	% HP	<i>NevGen</i>	% NG	Haplogrupo - Y-SNPs	Origem continental
MRJ 105	E1b1a	100 %	E1b1a V38>> L485	10,98 %	E1b1a-M2*(xM154,M191)	AFR
MRJ 106	I1	100 %	I1 Z59>> CTS7362>> BY266> L1302	32,89 %	I-M170	EUR
MRJ 120	G2a	100 %	G2a2b2a1c CTS342	49,27 %	G-M201	EUR
MRJ 122	I1	100 %	I1 L22> CTS11603	0,68 %	I-M170	EUR
MRJ 126	E1b1a	100 %	E1b1a V38>> L485	78,77 %	E1b1a-M191	AFR
MRJ_129	G2a	100 %	G2a2 > L497>> Z725>> CTS4803 S2808> S23438	26,22 %	G-M201	EUR
MRJ 132	R1b	100 %	R1b	99,94 %	R1b-L23*(xU106,S116)	EUR
MRJ 134	R1b	100 %	R1b	98,70 %	R1b-S116*(xU152,M529,M153,M167)	EUR
MRJ 136	E1b1a	100 %	E1b1a V38>> L485	83,86 %	E1b1a-M2*(xM154,M191)	AFR
MRJ 137	G2a	100 %	G2a2b2a1c CTS342> Z724>> Z26414	33,86 %	G-M201	EUR
MRJ 140	G2a	100 %	G2a2b2a1c CTS342	69,55 %	G-M201	EUR
MRJ 144	R1b	100 %	R1b	100 %	R1b-U106	EUR
MRJ 147	E1b1a	100 %	E1b1a V38>> L485	78,77 %	E1b1a-M191	AFR
MRJ 148	E1b1b/E1b1a	95,1 % / 4,9 %	negativo para todos	0	E2-M85	AFR
MRJ 164	I1	100 %	I1 DF29> Y18697> Y7282	24,67 %	I-M170	EUR
MRJ 169	E1b1a	100 %	E1b1a V38>> M4231	100 %	E1b1a-M2*(xM154,M191)	AFR
MRJ 177	E1b1a	100 %	E1b1a V38>> M4231	66,94 %	E1b1a-M2*(xM154,M191)	AFR
MRJ 179	E1b1b	100 %	E1b1b V22	54,56 %	E1b1b-M78	EUR/AFR
MRJ 180	R1a	100 %	R1a	100 %	R1a-SRY100813.2	EUR
MRJ 183	R1b	100 %	R1b	97,91 %	R1-M173*(xSRY10831.2,P25)	EUR
MRJ 185	R1b	100 %	R1b	100 %	R1b-S116*(xU152,M529,M153,M167)	EUR
MRJ 188	R1b	100 %	R1b	99,90 %	R1b-S116*(xU152,M529,M153,M167)	EUR
MRJ 191	R1b	100 %	R1b	99,93 %	R1b-S116*(xU152,M529,M153,M167)	EUR
MRJ 194	R1b	100 %	R1b	100 %	R1b-S116*(xU152,M529,M153,M167)	EUR
MRJ 195	Q	100 %	Q M346>> M3> M902	97,07 %	Q1a2-SA05	NAM

Tabela 17 - Resultado da determinação dos haplogrupos pelos *software Haplogroup Predictor* (HP) e *NevGen* e pela genotipagem com Y-SNPs, além da origem continental dos haplogrupos (continuação)

Amostra	HP	% HP	<i>NevGen</i>	% NG	Haplogrupo - Y-SNPs	Origem continental
MRJ_197	R1b	100%	R1b	100%	R1b-S116*(xU152,M529,M153,M167)	EUR
MRJ_198	R1b	100%	R1b	100%	R1b-S116*(xU152,M529,M153,M167)	EUR
MRJ_199	R1b	100%	R1b	100%	R1b-S116*(xU152,M529,M153,M167)	EUR
MRJ_200	E1b1a	100%	E1b1a V38>> L485	98,32%	E1b1a-M191	AFR
MRJ_201	E1b1a	100%	E1b1a V38>> M4231	92,78%	E1b1a-M2*(xM154,M191)	AFR
MRJ_202	J1	100%	J1a >> P58>> FGC11> FGC3723	2,44%	J1-P58	EUR
MRJ_208	I1	100%	I1 Z63>>L849> S2077> PR683> L1237	24,94%	I-M170	EUR
MRJ_213	Q	100%	Q M346>> L330	100%	Q1a2-M346*(xM3)	NAM
MRJ_219	Q	99,90%	Q M346>> Z780	36,91%	Q1a2-M3*(xM19,Z19319,Z19383,M557,SA05)	NAM
MRJ_220	Q	99,90%	Q M346>> Z780	36,91%	Q1a2-M3*(xM19,Z19319,Z19383,M557,SA05)	NAM
MRJ_221	Q	99,90%	Q M346>> Z780	36,91%	Q1a2-M3*(xM19,Z19319,Z19383,M557,SA05)	NAM
MRJ_225	H / E1b1a	95,3% / 4,7%	T >> CTS11451> PF7455	98,84%	T-M70	EUR
MRJ_226	R1b	100%	R1b	99,99%	R1b-M529	EUR
MRJ_234	R1b	100%	R1b	100%	R1b-S116*(xU152,M529,M153,M167)	EUR
MRJ_236	E1b1b	100%	E1b1b > V12	80%	E1b1b-M78	EUR/AFR
MRJ_237	R1b	100%	R1b	100%	R1b-S116*(xU152,M529,M153,M167)	EUR
MRJ_238	R1b	100%	R1b	100%	R1b-S116*(xU152,M529,M153,M167)	EUR
MRJ_246	Q / J1	95,5% / 4,5%	Q M346>> M3> M902	72,85%	Q1a2-SA05	NAM
MRJ_247	R1b	100%	R1b	99,96%	R1b-S116*(xU152,M529,M153,M167)	EUR
MRJ_249	R1b	100%	R1b	100%	R1b-S116*(xU152,M529,M153,M167)	EUR
MRJ_250	R1b	100%	R1b	100%	R1b-S116*(xU152,M529,M153,M167)	EUR
MRJ_251	R1b	100%	R1b	100%	R1b-S116*(xU152,M529,M153,M167)	EUR
MRJ_252	Q / J1	96,6% / 3,4%	Q M346>> M3> M902	8,90%	Q1a2-M3*(xM19,Z19319,Z19383,M557,SA05)	NAM

Tabela 17 - Resultado da determinação dos haplogrupos pelos *software Haplogroup Predictor* (HP) e *NevGen* e pela genotipagem com Y-SNPs, além da origem continental dos haplogrupos (conclusão)

Amostra	HP	% HP	<i>NevGen</i>	% NG	Haplogrupo - Y-SNPs	Origem continental
MRJ 254	R1b	100%	R1b	99,99%	R1b-S116*(xU152,M529,M153,M167)	EUR
MRJ 255	E1b1b / E1b1a	97,7% / 2,3%	E2 M75	9,09%	E2-M85	AFR
MRJ 256	R1b	100%	R1b	99,97%	R1b-S116*(xU152,M529,M153,M167)	EUR
MRJ 261	E1b1a/ E1b1b	95,6%/ 4,4%	E1b1a V38>> M4231	76,12%	E1b1a-M2*(xM154,M191)	AFR
MRJ 262	R1b	100 %	R1b	100 %	R1b-M529	EUR
MRJ 264	R1b	100 %	R1b	99,99 %	R1b-S116*(xU152,M529,M153,M167)	EUR
MRJ 270	R1b	100 %	R1b	99,82 %	R1b-S116*(xU152,M529,M153,M167)	EUR
MRJ 271	E1b1b	100 %	E1b1b V257>M81	100 %	E1b1b-M81	EUR
MRJ 273	H	99,80 %	J2a2 PF5008 >> PF5040	64,38 %	J2-M172	EUR
MRJ 275	R1b	100 %	R1b	98,84 %	R1b-S116*(xU152,M529,M153,M167)	EUR
MRJ 278	R1b	100 %	R1b	98,25 %	R1b-S116*(xU152,M529,M153,M167)	EUR
MRJ 279	R1b	100 %	R1b	100 %	R1b-S116*(xU152,M529,M153,M167)	EUR
MRJ 282	R1b	100 %	R1b	91,85 %	R1b-S116*(xU152,M529,M153,M167)	EUR
MRJ 283	H/ E1b1b	98,3 % / 1,7 %	J2a1 Z387>L70> CTS3601	48,70 %	J2-M172	EUR
MRJ 288	R1b	100 %	R1b	91,61 %	R1b-L23*(xU106,S116)	EUR
MRJ 290	E1b1a	100 %	E1b1a V38>> L485	89,98 %	E1b1a-M191	AFR
MRJ 292	R1b	100 %	R1b	99,99 %	R1b-S116*(xU152,M529,M153,M167)	EUR
MRJ 293	Q	100 %	Q M346>> M3> M902	98,19 %	Q1a2-SA05	NAM
MRJ 295	R1b	100 %	R1b	99,88 %	R1b-S116*(xU152,M529,M153,M167)	EUR
MRJ 299	R1b	100 %	R1b	100 %	R1b-S116*(xU152,M529,M153,M167)	EUR
MRJ 304	E1b1b	100 %	E2 M75	9,09 %	E2-M85	AFR
MRJ 305	R1b	100 %	R1b	100 %	R1b-S116*(xU152,M529,M153,M167)	EUR
MRJ 306	E1b1b	100 %	E2 M75	9,09 %	E2-M85	AFR
MRJ 307	E1b1b	100 %	E2 M75	9,09 %	E2-M85	AFR
MRJ 308	R1b	100 %	R1b	100 %	R1b-S116*(xU152,M529,M153,M167)	EUR
MRJ 310	R1b	100 %	R1b	100 %	R1b-S116*(xU152,M529,M153,M167)	EUR

Legenda: Haplogrupos previstos e as probabilidades calculadas para o *software Haplogroup Predictor* (% HP) e para o *software NevGen* (% NG).

Tabela 18 - Frequências dos haplogrupos do Cromossomo Y na população da Ilha de Marajó

Haplogrupo	Origem	F.A.	F.R.
E1b1a-M2*(xM154,M191)	Africana	7	0,072165
E2-M85	Africana	5	0,051546
E1b1a-M191	Africana	4	0,041237
Q1a2-M3*(xM19,Z19319,Z19383,M557,SA05)	Nativa Americana	7	0,072165
Q1a2-SA05	Nativa Americana	3	0,030928
Q1a2-M346*(xM3)	Nativa Americana	1	0,010309
R1b-S116*(xU152,M529,M153,M167)	Europeia	39	0,402062
I-M170	Europeia	8	0,082474
G-M201	Europeia	6	0,061856
J2-M172	Europeia	3	0,030928
R1b-M529	Europeia	2	0,020619
R1b-L23*(xU106,S116)	Europeia	2	0,020619
E1b1b-M81	Europeia	1	0,010309
T-M70	Europeia	1	0,010309
R1b-U106	Europeia	1	0,010309
R1b-U152	Europeia	1	0,010309
R1-M173*(xSRY10831.2,P25)	Europeia	1	0,010309
J1-P58	Europeia	1	0,010309
R1a-SRY100813.2	Europeia	1	0,010309
E1b1b-M78	Indefinida	3	0,030928

Legenda: Haplogrupos observados na amostra populacional da Ilha de Marajó com suas respectivas frequências absolutas (F.A.) e relativas (F.R.) e as regiões em que são frequentemente encontrados. *A mutação M81 se originou no norte da África, porém, atualmente, é muito frequente na região da Península Ibérica. Sendo esta aprovável fonte que trouxe a mutação M81 para o Brasil.

3.3.3 Análise estatística dos resultados do Cromossomo Y – Y-STRs

3.3.3.1 Diversidade Haplotípica

A diversidade haplotípica representa a probabilidade de dois indivíduos escolhidos ao acaso numa mesma população possuírem haplótipos distintos, relacionando-se com o poder de discriminação de um determinado conjunto de marcadores numa população.

Com base nos dados obtidos pela genotipagem dos Y-STRs com *Yfiler Plus*, foram realizadas estimativas da diversidade haplotípica por meio do *software* Arlequin versão 3.5. Para a amostra com 97 haplótipos do cromossomo Y da Ilha de Marajó foi observado um índice de diversidade de $0,9959 \pm 0,0025$. Esse valor de diversidade é inferior ao valor encontrado quando comparado a outras populações brasileiras miscigenadas (Tabela 19). Isso

pode ser explicado pela presença dos haplótipos compartilhados nessa população (quatro haplótipos são compartilhados por dois indivíduos, um haplótipo é compartilhado por três indivíduos e 2 haplótipos são compartilhados por quatro indivíduos).

Tabela 19 - Valores de Diversidade Haplótica para populações brasileiras e populações de referência africana, europeia e nativa americana - *Yfiler Plus*

. População/ Região	Nº de indivíduos/ Nº de haplótipos	Etnia	Diversidade Haplótica
Marajó (N)	97/85	Miscigenada	0,9959 ± 0,0025
Maranhão (NE)	296/288	Miscigenada	0,9998 ± 0,0003
Piauí (NE)	42/42	Miscigenada	1,0000 ± 0,0052
Ceará (NE)	38/38	Miscigenada	1,0000 ± 0,0060
Rio Grande do Norte (NE)	19/19	Miscigenada	1,0000 ± 0,0171
Pernambuco (NE)	141/141	Miscigenada	1,0000 ± 0,0008
Mato Grosso (CO)	100/99	Miscigenada	0,9998 ± 0,0015
Goiás (CO)	204/197	Miscigenada	0,9996 ± 0,0005
Rio de Janeiro (SE)	258/257	Miscigenada	0,99997 ± 0,0003
São Paulo (SE)	169/166	Miscigenada	0,9998 ± 0,0007
Espírito Santo (SE)	409/403	Miscigenada	0,9999 ± 0,0002
Espírito Santo-pomeranos (SE)	79/77	Miscigenada	0,9994 ± 0,0022
Rio Grande do Sul (S)	211/203	Miscigenada	0,9996 ± 0,0005
Kenya	62/54	Africana	0,9952 ± 0,0040
Espanha	126/126	Europeia	1,0000 ± 0,0010
Peru	58/44	Nativa Americana	0,9843 ± 0,0082

Legenda: N - Região Norte, NE - Região Nordeste, CO - Região Centro-Oeste, SE - Região Sudeste, S - Região Sul

Com dados disponíveis para outras populações da região Norte do Brasil, publicados por Palha e colaboradores (2012) e Purps e colaboradores (2014), a análise comparativa de diversidade haplotípica foi realizada para um subconjunto de 18 Y-STRs. A população de Marajó apresentou o menor valor de diversidade haplotípica ($0,9953 \pm 0,0026$) em comparação às populações de Belém, Boa Vista, Macapá, Manaus, Palmas, Porto Velho, Rio Branco, Santarém e São Gabriel da Cachoeira (Tabela 20). Os marcadores utilizados nessa análise foram: DYS19, DYS385, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS570, DYS576, DYS635 e YGATAH4.

Tabela 20 - Valores de Diversidade Haplotípica para populações da Região Norte -18 Y-STRs

População	Nº de indivíduos/ Nº de haplótipos	Etnia	Diversidade haplotípica
Marajó	97/83	Miscigenada	0,9953 ± 0,0026
Belém	400/399	Miscigenada	0,9999 ± 0,0002
Boa Vista	67/67	Miscigenada	1,0000 ± 0,0026
Macapá	65/65	Miscigenada	1,0000 ± 0,0027
Manaus	74/74	Miscigenada	1,0000 ± 0,0022
Palmas	30/30	Miscigenada	1,0000 ± 0,0086
Porto Velho	135/134	Miscigenada	0,9999 ± 0,0009
Rio Branco	31/31	Miscigenada	1,0000 ± 0,0082
Santarém	72/70	Miscigenada	0,9988 ± 0,0027
São Gabriel da Cachoeira	61/57	Nativa	0,9973 ± 0,0038

A população da Ilha de Marajó apresentou uma diminuição muito discreta no valor da diversidade haplotípica quando a análise foi realizada com um grupo menor de marcadores, indicando que ambos os conjuntos de marcadores poderiam ser usados para discriminar indivíduos dessa população. Essa diminuição pode ser explicada pela menor capacidade de discriminação com esse número de marcadores, o que pode ser observada pelo aumento no número de haplótipos compartilhados; diminuindo o número de haplótipos únicos com o *Yfiler Plus*, que foi 78, para 75, na mesma população.

Uma maior diversidade haplotípica com o uso dos 27 loci do kit *Yfiler Plus* em relação aos 17 marcadores do kit *Yfiler* já havia sido relatada para algumas populações analisadas por Jannuzzi e colaboradores (2020).

3.3.3.2 Distância Genética (F_{ST})

As distâncias genéticas F_{ST} e as probabilidades de não-diferenciação entre pares de populações foram calculadas entre as populações da Tabela 19 e entre as populações da Região Norte do Brasil, da Tabela 20, utilizando o *software* Arlequin versão 3.5. Os valores obtidos estão apresentados nas Tabelas 21 e 22, respectivamente. Os valores obtidos para as distâncias genéticas podem ser visualizados em duas dimensões através de gráficos MDS, nas Figuras 13 (populações brasileiras e referências africana, europeia e nativa americana) e 14 (populações da Região Norte do Brasil).

As distâncias genéticas baseadas nos valores de F_{ST} indicaram uma homogeneidade entre as populações brasileiras analisadas (12 amostras com dados previamente publicados e a amostra da Ilha de Marajó), exceto pela população de indivíduos pomeranianos do Espírito Santo, que só não apresentaram diferenças significativas com a população do Rio Grande do Sul. Os pomeranianos ou pomeranos descendem da Pomerânia, região situada ao Norte da Polônia e da Alemanha. São conhecidos por preservarem sua cultura original, tradições e a língua. Muitos descendentes emigraram para o Brasil durante o século XIX, principalmente para Santa Catarina e também para o Rio Grande do Sul. Atualmente existe uma grande comunidade de descendentes de pomeranianos na região montanhosa do Estado do Espírito Santo, onde sua língua é ensinada nas escolas da rede pública, havendo forte preservação cultural. Dados originados com os marcadores Y-STR do *kit Yfiler Plus* foram publicados por Stange e colaboradores (2019), demonstrando através dos valores de F_{ST} obtidos que os pomeranianos estão mais próximos geneticamente dos europeus do que dos africanos ou das populações brasileiras miscigenadas, incluindo a própria população do Estado do Espírito Santo. Isso corrobora o resultado encontrado neste trabalho, embora haja distância genética significativa entre os pomeranianos e a população europeia considerada, que é de outra região da Europa (Madri, Espanha).

Na representação bidimensional (Figura 13) é possível visualizar o agrupamento das populações brasileiras quando analisadas com as populações de referência, e perceber que estão mais próximas da europeia, o que está de acordo com a história da colonização do território brasileiro por europeus vindos principalmente da Península Ibérica.

A homogeneidade observada entre as populações brasileiras, exceto pelos Pomeranianos do Espírito Santo, se evidencia por valores de F_{ST} consideravelmente baixos contrastando com valores de probabilidades de não-diferenciação bastante elevados, acima do nível de significância (Tabela 21). Embora haja altas distâncias genéticas entre as populações brasileiras e as referências nativa Americana e africana analisadas, observa-se que as maiores distâncias se encontram entre o Rio Grande do Sul e a população nativa (18,44 %), e entre o Ceará e a africana (19,81 %).

Além dos Pomeranianos do Espírito Santo, a população analisada da Ilha de Marajó apresentou distâncias significativas apenas com a população nativa americana (13,73 %) e a africana (14,82 %). Já com a população europeia o valor de F_{ST} foi de apenas 1,3 %, mas com uma probabilidade de não diferenciação alta. Esses resultados são consistentes com a alta proporção de linhagens masculinas européias encontradas na Ilha de Marajó.

Na comparação realizada entre as populações da região Norte observa-se, pelos índices calculados (Tabela 22), que a amostra de São Gabriel é significativamente diferente de todas as outras, o que pode ser visualizado pelo gráfico bidimensional da Figura 14. Isso pode ser explicado pelas características dessa população, que é nativa e vive isolada geográfica e culturalmente, na região amazônica. Os altos valores de F_{ST} observados nessa comparação com São Gabriel indicam uma maior prevalência das linhagens masculinas europeias em todas as outras populações. Observa-se também distâncias significativas, porém menores, entre Belém e Marajó (1,3 %) e entre Boa Vista e Marajó (1,7 %). Essa menor distância com Belém, em relação a Boa Vista, capital do Estado de Roraima, pode ser explicada pela proximidade geográfica entre a capital Belém e a Ilha de Marajó, ambas pertencentes ao Estado do Pará.

Os dados publicados por Resque e colaboradores (2016) sobre a distribuição de haplogrupos em todas as regiões brasileiras, contrastados com os dados dos SNPs neste trabalho, podem explicar a diferença observada entre Marajó e Belém, uma vez que a proporção de haplogrupos nativos (12 %) e africanos no Marajó (17 %) é maior que a observada por Resque e colaboradores na Região Norte, representada em seu trabalho majoritariamente por indivíduos de Belém (6,6 % de africanos e 8,1 % de nativos).

Este estudo possibilitou a geração de dados de haplótipos do cromossomo Y para os marcadores Y-STR presentes no *kit Yfiler Plus*, contribuindo para a construção do banco de dados da população brasileira. Além disso, é o primeiro estudo genético realizado com marcadores Y-STR na população miscigenada da Ilha de Marajó, e os resultados mostraram que este conjunto de marcadores tem alta diversidade de haplótipos nessa população, tornando-o útil para investigações forenses por fornecer alta discriminação intrapopulacional.

Tabela 21 - Matriz das distâncias genéticas baseadas no F_{ST} (abaixo da diagonal) entre todos os pares de populações comparadas com a população da Ilha de Marajó e os correspondentes valores de probabilidade de não diferenciação p (acima da diagonal) – dados *Yfiler Plus*

	Marajó	Peru NAM	Ceará	ES	ES-POM	Goiás	Kenya AFR	Espanha EUR	Maranhão	Mato Grosso	PE	Piauí	RJ	RS	RN	SP
Marajó	*	<5,00E-6	0,37521	0,07019	<5,00E-6	0,12444	<5,00E-6	0,00089	0,12464	0,02624	0,25631	0,11593	0,1284	0,59014	0,01673	0,06801
Peru-NAM	0,13727	*	<5,00E-6	<5,00E-6	<5,00E-6	<5,00E-6	<5,00E-6	<5,00E-6	<5,00E-6	<5,00E-6	<5,00E-6	<5,00E-6	<5,00E-6	<5,00E-6	<5,00E-6	<5,00E-6
Ceará	0,00065	0,17901	*	0,38273	0,0001	0,39669	<5,00E-6	0,06574	0,19394	0,64885	0,21463	0,23612	0,53272	0,94892	0,46758	0,54945
ES	0,00248	0,14208	0,00036	*	<5,00E-6	0,2676	<5,00E-6	<5,00E-6	0,40808	0,3365	0,12177	0,2579	0,49559	0,66033	0,17909	0,16771
ES-POM	0,02299	0,16229	0,03057	0,02251	*	<5,00E-6	<5,00E-6	<5,00E-6	<5,00E-6	<5,00E-6	<5,00E-6	0,0003	<5,00E-6	0,0393	<5,00E-6	0,0001
Goiás	0,00211	0,13737	0,00027	0,00037	0,02415	*	<5,00E-6	<5,00E-6	0,30918	0,68053	0,41867	0,44471	0,33145	0,69647	0,08356	0,26819
Kenya -AFR	0,1482	0,22553	0,19808	0,15025	0,15281	0,14842	*	<5,00E-6	<5,00E-6	<5,00E-6	<5,00E-6	<5,00E-6	<5,00E-6	<5,00E-6	<5,00E-6	<5,00E-6
Espanha- EUR	0,01299	0,17758	0,00699	0,01327	0,03925	0,01278	0,214	*	0,0001	0,00842	<5,00E-6	0,01059	0,0005	0,08682	0,0003	0,00168
Maranhão	0,0019	0,13808	0,00267	0,00002	0,02559	0,0003	0,14592	0,0126	*	0,11623	0,33106	0,30126	0,76161	0,70646	0,07861	0,19949
Mato Grosso	0,00587	0,15124	-0,00222	0,00033	0,02606	-0,00103	0,17448	0,00854	0,00195	*	0,07128	0,42154	0,46985	0,67835	0,2274	0,33888
PE	0,00097	0,13319	0,0027	0,0013	0,02209	0	0,12292	0,02195	0,00028	0,00348	*	0,30096	0,20018	0,83259	0,01426	0,03633
Piauí	0,0049	0,16639	0,00343	0,00143	0,02845	-0,00017	0,15428	0,01356	0,00111	0,00008	0,0012	*	0,40352	0,80982	0,14969	0,2373
RJ	0,00185	0,14294	-0,00094	-0,00015	0,02068	0,00026	0,148	0,01036	-0,00066	-0,00023	0,00093	0,00009	*	0,76903	0,09712	0,58093
RS	-0,00295	0,1844	-0,01277	-0,00356	0,01782	-0,00432	0,16263	0,01055	-0,00422	-0,00434	-0,00677	-0,00867	-0,00516	*	0,72458	0,7422
RN	0,00537	0,15058	-0,00036	0,00069	0,02017	0,00166	0,16892	0,01098	0,00155	0,00102	0,0046	0,00341	0,0014	-0,00469	*	0,25522
SP	0,00309	0,14532	-0,00111	0,00083	0,02234	0,00055	0,16614	0,00947	0,00081	0,0004	0,00342	0,00195	-0,00042	-0,00486	0,00061	*

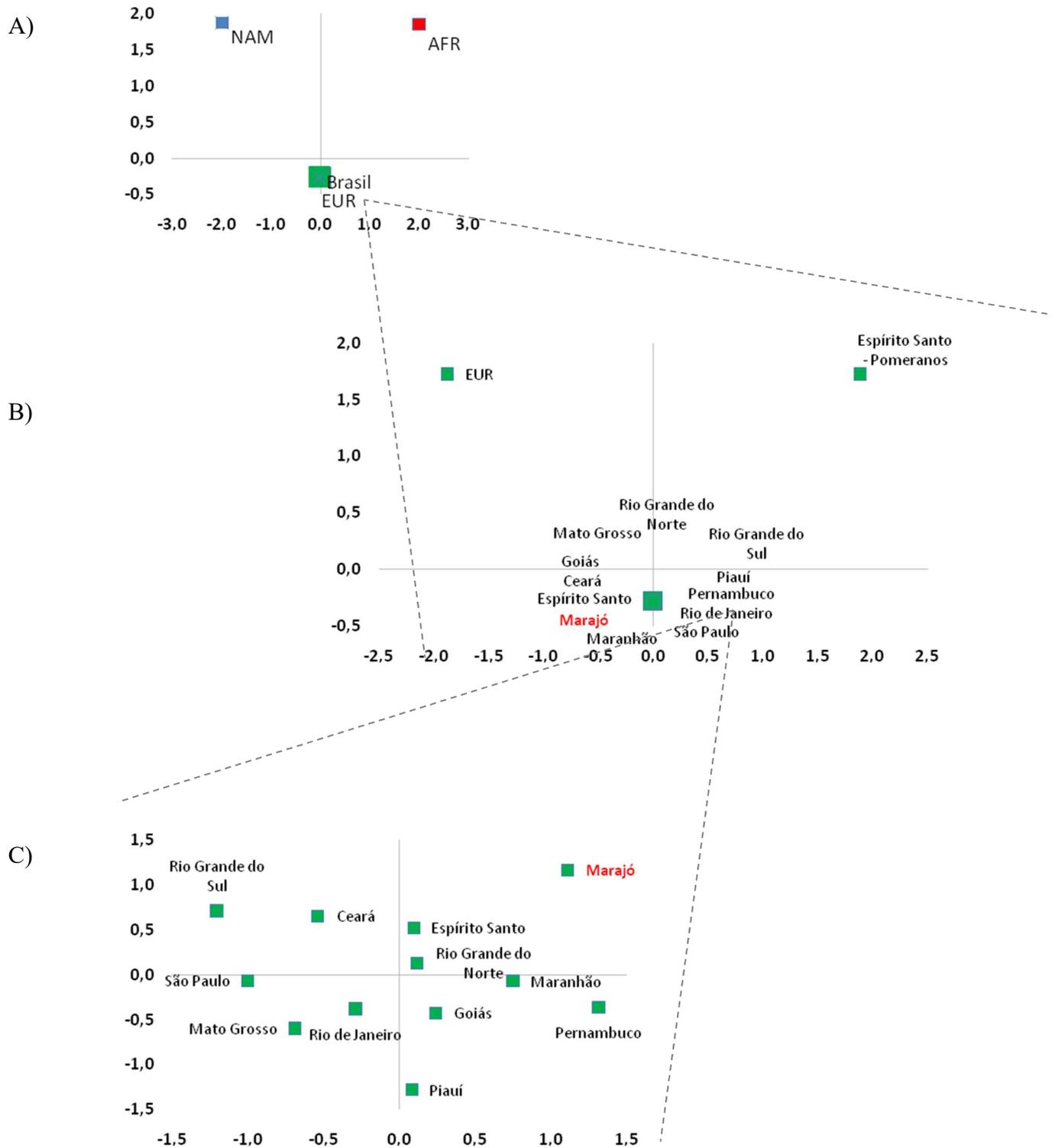
Legenda: Valores de F_{ST} computados com 10100 permutações – diagonal de baixo; Valores de p – diagonal de cima. Valores de significância de $p \leq 0,000417$ após correção de Bonferroni grifados em vermelho, assim como os valores de F_{ST} significativos entre Marajó e as outras populações. Desvio-padrão $\leq 0,0057$. NAM - nativa americana, AFR - africana, EUR - Europeia, ES - Espírito Santo, ES-POM - Espírito Santo - Pomeranos, PE - Pernambuco, RJ - Rio de Janeiro, RS - Rio Grande do Sul, RN - Rio Grande do Norte, SP - São Paulo.

Tabela 22 - Matriz das distâncias genéticas baseadas no F_{ST} (abaixo da diagonal) entre todos os pares de populações da Região Norte do Brasil comparadas com a população da Ilha de Marajó e os correspondentes valores de probabilidade de não diferenciação p (acima da diagonal) - conjunto de 18 marcadores Y-STR

	Marajó	Belém	Boa Vista	Macapá	Manaus	Palmas	Porto Velho	Rio Branco	Santarém	São Gabriel
Marajó	*	0,00010	0,00109	0,03515	0,02554	0,43798	0,01525	0,37234	0,04841	< 5,00E-04
Belém	0,01323	*	0,00030	0,00713	< 5,00E-04	0,03534	< 5,00E-04	0,02297	0,00010	< 5,00E-04
Boa Vista	0,01707	0,01579	*	0,04316	0,00089	0,00842	< 5,00E-04	0,04861	0,00010	< 5,00E-04
Macapá	0,00843	0,00905	0,00836	*	0,08366	0,09217	0,01832	0,27413	0,05683	< 5,00E-04
Manaus	0,00824	0,01854	0,01950	0,00582	*	0,08435	0,00614	0,48451	0,65429	< 5,00E-04
Palmas	-0,00013	0,01075	0,02110	0,00982	0,00920	*	0,77814	0,33393	0,46619	< 5,00E-04
Porto Velho	0,00739	0,01778	0,02304	0,00902	0,01032	-0,00435	*	0,47203	0,14711	< 5,00E-04
Rio Branco	0,00073	0,01177	0,01172	0,00285	-0,00083	0,00199	-0,00066	*	0,49955	< 5,00E-04
Santarém	0,00667	0,01550	0,02728	0,00750	-0,00194	-0,00075	0,00288	-0,00119	*	< 5,00E-04
São Gabriel	0,06769	0,07638	0,09320	0,08923	0,06609	0,07625	0,08164	0,07480	0,05926	*

Legenda: Valores de F_{ST} computados com 10100 permutações – diagonal de baixo; Valores de p – diagonal de cima. Valores de significância de $p \leq 0,001111$ após correção de Bonferroni grifados em vermelho, assim como como os valores de F_{ST} significativos entre Marajó e as outras populações. Desvio-padrão $\leq 0,0052$

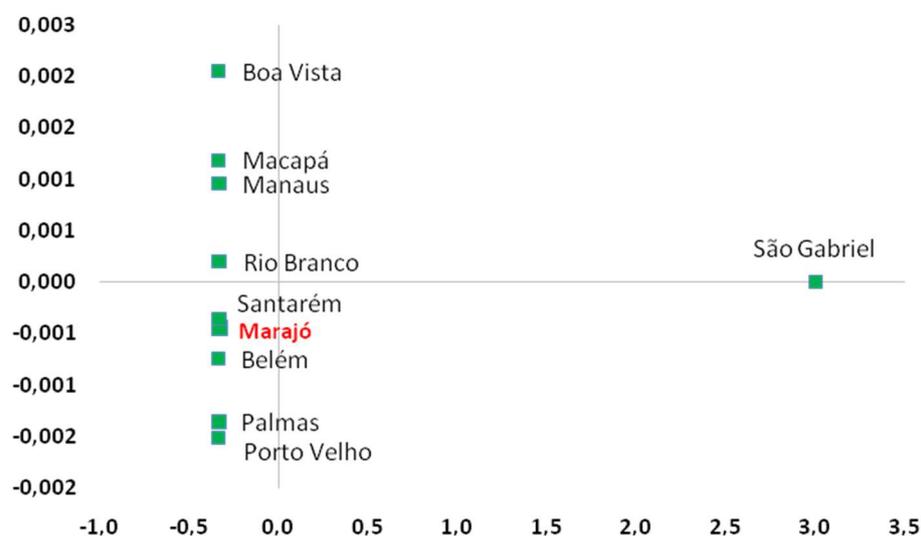
Figura 13 - Gráfico MDS das distâncias genéticas F_{ST} entre as populações brasileiras e populações de referência africana, europeia e nativa americana - haplótipos *Yfiler Plus*



Legenda: Representação gráfica em duas dimensões, das distâncias genéticas entre as populações. A) Populações brasileiras e as populações de referência, Stress 0,0000294; B) Populações brasileiras e a população Europeia, Stress 0,0000039; C) Populações brasileiras, exceto Espírito Santo-Pomeranianos, Stress 0,1697549. AFR -africana, EUR - europeia, NAM - nativa americana.

Fonte: A autora.

Figura 14 - Gráfico MDS das distâncias genéticas F_{ST} entre as populações da região Norte do Brasil - conjunto de 18 marcadores Y-STR



Legenda: Representação gráfica em duas dimensões, das distâncias genéticas entre as populações. Stress 0,0002273.

Fonte: A autora.

3.4 Análise de marcadores do mtDNA

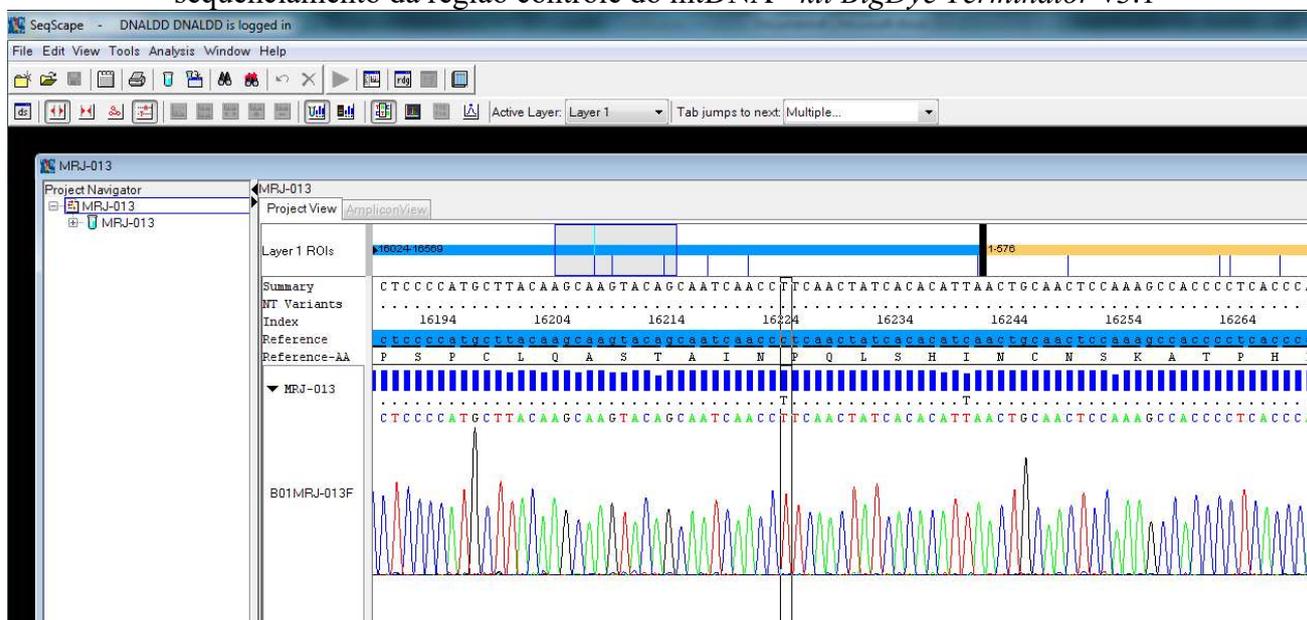
3.4.1 Haplótipos da Região Controle do mtDNA

A Região Controle do mtDNA foi sequenciada para 101 amostras, segundo a metodologia descrita no item 2.5.1 de Materiais e Métodos, e a lista completa dos haplótipos obtidos, assim como o haplogrupo correspondente e sua origem continental está no Apêndice B. A Figura 15 apresenta um trecho de um eletroferograma, representação gráfica do resultado do sequenciamento de uma das amostras analisadas e a comparação de sua sequência com a *rCRS*, no *software SeqScape v2.7*. O conjunto de polimorfismos observados nas sequências obtidas do mtDNA em relação à *rCRS* determinou os haplótipos listados.

Foram encontrados 76 haplótipos distintos nos 101 indivíduos analisados, considerando-se inteiramente a Região Controle. Desses, 62 são haplótipos únicos, com frequência de 0,009901; 7 haplótipos são compartilhados por dois indivíduos (MRJ-003 e MRJ-166; MRJ-050 e MRJ-188; MRJ-060 e MRJ-185; MRJ-128 e MRJ-305; MRJ-134 e

MRJ-142; MRJ-147 e MRJ-148; MRJ-236 e MRJ-237), com frequência de 0,019802; 5 haplótipos são compartilhados por três indivíduos (MRJ-012, MRJ-088 e MRJ-226; MRJ-029, MRJ-049 e MRJ-078-2; MRJ-072, MRJ-120 e MRJ-126; MRJ-085, MRJ-172 e MRJ-194; MRJ-304, MRJ-306 e MRJ-307), com 0,029703 de frequência; 1 haplótipo com frequência de 0,039604, compartilhado por quatro indivíduos (MRJ-056, MRJ-151, MRJ-164, MRJ-176) e 1 compartilhado por 6 indivíduos (MRJ-195, MRJ-197, MRJ-198, MRJ-202, MRJ-204 e MRJ-207), de frequência 0,0594059. Este número de haplótipos difere do número informado na Tabela 24 para a população do Marajó, que após a exclusão das regiões homopoliméricas e diméricas para fins de cálculos de diversidade, passou a ter 73 haplótipos.

Figura 15 - Eletroferograma do perfil genético de uma amostra obtido através do sequenciamento da região controle do mtDNA - *kit BigDye Terminator v3.1*



Legenda: Análise da sequência da região controle do mtDNA de uma amostra deste trabalho no *software SeqScape v2.7*, por comparação com a rCRS. Os picos coloridos referem-se a cada uma das bases (A- verde, C- azul, G -preto e T - vermelho) e na posição destacada observa-se uma transição de C para T em relação à sequência de referência.

Fonte: *Software SeqScape v2.7*

Nesse grupo de haplótipos foi encontrada heteroplasmia de posição em 3 das 101 amostras, o que representa um percentual de 2,97 %, um valor levemente inferior aos percentuais relatados em amostras do Espírito Santo e do Rio de Janeiro também analisadas anteriormente para a Região Controle (DOS REIS *et al.*, 2019; SIMÃO *et al.*, 2018). As heteroplasmias encontradas foram: mistura de C e T, representada pela letra Y, nas amostras MRJ-035 (posição 64Y) e MRJ-144 (posição 199Y) e mistura de A e G, representada por R, na amostra MRJ-107 (posição 234R). Em relação às posições onde ocorreram, apenas a heteroplasmia da posição 234 não está listada como uma posição *hotspot* (SOARES *et al.*,

2009), e apenas a mutação 64Y não se encontra nos segmentos HV-I e HV-II do genoma mitocondrial, onde ocorrem a grande maioria das mutações. Foram observadas mutações nessas três posições em mais de uma ocorrência nas 319 mutações observadas por Irwin e colaboradores (2009) e foi verificada que essas três heteroplasmias estão registradas na base de dados EMPOP (pesquisa em dezembro de 2021).

Uma alta frequência de heteroplasmia de comprimento foi encontrada nessas amostras, associadas a inserções de citosinas (sequências poli - C) em HV-I e HV-II. Adicionalmente, heteroplasmias de comprimento foram observadas pela inserção de repetições do dinucleotídeo AC em HV-III, entre as posições 513 e 525, em algumas amostras e a inserção de citosinas na posição 573 ocorreu em 1 amostra.

3.4.2 Haplogrupos de mtDNA na população da Ilha de Marajó

As linhagens de mtDNA das 101 amostras genotipadas para a região controle foram classificadas em haplogrupos com auxílio da base de dados EMPOP database v4/R13 (PARSON; DÜR, 2007) e confirmadas manualmente na *PhyloTree - mtDNA tree Build 17*. A Tabela 23 mostra o resultado obtido para essas amostras, nas quais foram observados 46 diferentes ramos, sendo 19 de origem continental nativa americana (haplogrupos A, B, C e D), 25 de origem africana (haplogrupos L) e 2 de origem euroasiática (haplogrupos K1b1a e M). Um total de quatro amostras tiveram uma classificação indefinida, podendo pertencer ao haplogrupo euroasiático M ou ao haplogrupo nativo D. Excluindo-se essas amostras do total, foi encontrada uma alta proporção de de linhagens nativa americanas (62,9 %), seguidas de africanas (36,1 %) e de uma minoria de linhagens euroasiáticas (aproximadamente 1%).

Comparando as proporções das ancestralidades continentais encontradas nessa amostra do Marajó com outras populações brasileiras miscigenadas já analisadas para a Região Controle (Tabela 11), observa-se que a única população a apresentar uma maioria de haplogrupos nativos é a do Paraná, como mostra a Figura 16. Nessa população, 49,2 % dos haplogrupos encontrados são nativos, seguidos de 38,5 % europeus e apenas 12,3 % africanos. As populações do Distrito Federal, assim como as do Rio de Janeiro e de São Paulo, apresentam majoritariamente linhagens africanas. Porém, nessas populações do Sudeste, essas vêm seguidas das europeias, enquanto no Centro-Oeste as nativas ocupam o

segundo lugar. Já as populações do Espírito Santo (considerando conjuntamente as macrorregiões da Tabela 11) e de Santa Catarina estão caracterizadas por maioria das linhagens de origem europeia, seguidas de africanas no Espírito Santo, e de nativas em Santa Catarina. Essa heterogeneidade nas proporções das ancestralidades das linhagens maternas no Brasil já foi relatada anteriormente, e resulta do alto nível de miscigenação, relacionado às particularidades dos processos de colonização através de cada região geopolítica do país, que por ter

uma grande extensão geográfica, apresenta aspectos demográficos e antropológicos diversos ao longo de seu território.

O trabalho publicado por Alves-Silva e colaboradores (2000) ao analisarem a região HV-I do mtDNA de indivíduos de quatro regiões do Brasil, mostrou um predomínio de linhagens européias de 39 %, seguidas de 33 % nativas e 28 % africanas. Mas ao considerar cada região isoladamente, foi encontrado um percentual de 54 % de linhagens nativas na Região Norte, predomínio de linhagens africanas no Nordeste, com 44 %, e de linhagens europeias no Sul, chegando a 66 %. Já no Sudeste as proporções dos três grupos foram muito próximas, com 34 % africanas, 33 % nativas e 31 % europeias.

Ancestralidade nativa predominante para o mtDNA já foi relatada para a Região Nordeste, em análise das Regiões HV-I, HV-II e HV-III (43,5 %), sobretudo nos Estados do Piauí, Ceará e Rio Grande do Norte (SCHAAN *et al.*, 2017). Já os Estados de Alagoas e de Pernambuco têm um predomínio do componente africano (ALVES-SILVA, 2000; BARBOSA *et al.*, 2008).

A amostra do Marajó analisada neste estudo foi comparada, em relação à distribuição de haplogrupos, com indivíduos da Região Norte do Brasil também analisados para a Região Controle (NOGUEIRA *et al.*, 2017). Pode-se perceber que, nessa região, com exceção do Estado do Tocantins, que apresenta o componente matrilinear africano um pouco mais alto, todos os outros Estados representados mostram predomínio das linhagens nativas, chegando a 70 % para Roraima (Figura 17).

O resultado obtido para a ancestralidade matrilinear da população da Ilha de Marajó mostra uma coexistência de linhagens com diferentes origens filogeográficas e um predomínio das linhagens nativas, o que está de acordo com estudos realizados para a Região Norte como um todo (ALVES-SILVA *et al.*, 2000; NOGUEIRA *et al.*, 2017).

Dentre os haplogrupos nativo americanos presentes na amostra do Marajó, o de maior frequência é o haplogrupo A (20/61 - 32,8 %), seguido dos haplogrupos B (15/61 - 24,6 %) e C (16/61 - 26,2 %) e em menor proporção o haplogrupo D (10/61 - 16,4 %). Essa distribuição

difere da encontrada para as linhagens da Região Norte analisadas por Nogueira e colaboradores (2007) e por Alves-Silva e colaboradores (2000), em que o haplogrupo C é predominante (30 % e 38 %, respectivamente). Comparando com a população do Paraná, que também apresentou maior componente nativo, nessa população os haplogrupos A, B e C têm proporções bem próximas (13,9 - 15,6 %), enquanto o haplogrupo D também tem menor proporção. Isso também ocorre na população do Rio Grande do Norte. Já No Piauí o haplogrupo nativo predominante é o A, enquanto no Ceará é o haplogrupo C (POLETTTO *et al.*, 2018; SCHAAN *et al.*, 2017).

Tabela 23 - Frequências dos haplogrupos do mtDNA na população da Ilha de Marajó

Haplogrupo	Origem	F.A.	F.R.
K1b1a	Europeia	1	0,00990
L0 (3) - L0a1b (1), L0a1`4 (1), L0d1c(1)	Africana	3	0,02970
L1(4) - L1b1a10(2), L1b1a+189(1), L1c2b1b(4), L1c5(2)	Africana	9	0,08911
L2 (8) - L2a1(1), L2a1+16189+ (16192) (2), L2a1+143+16189+ (16192) (1), L2a1c3a1(1), L2b2(1), L2c1(1), L2c1a(1), L2d1a(1)	Africana	9	0,08911
L3 (10) - L3b(3), L3b2(1), L3d(1), L3e1(2), L3e1d(1), L3e2b(2), L3e2b1(1), L3e3(1), L3e4a(1),	Africana	14	0,13861
A (6) - A2(1), A2+64(12), A2ac(1), A2aI(3), A2an (2), A2h(1)	Nativa Americana	20	0,19802
B(3) - B2(2), B2o(5), B4b(8)	Nativa Americana	15	0,14851
C(5) - C1(6), C1b(6), C1b2(2), C1d(1), C1d+194(1)	Nativa Americana	16	0,15842
D(5) - D1(3), D1a2(1), D1f(2), D1f1(2), D4(2)	Nativa Americana	10	0,09901
M/D1	Eurásia / nativa Americana	4	0,03960

Legenda: Haplogrupos observados na amostra populacional da Ilha de Marajó com suas respectivas frequências absolutas (F.A.) e relativas (F.R.) e as regiões em que são frequentemente encontrados.

Historicamente estima-se que no período de colonização da América do Sul, aproximadamente 2,4 milhões de pessoas pertencentes principalmente aos grupos lingüísticos Tupi e Gê viviam no território que hoje é o Brasil, e que os Tupis se expandiram pela costa brasileira. Atualmente o número de indivíduos nativo americanos está significamente reduzido, devido aos efeitos do proceso de colonização ocorrido por séculos, resultando em forte efeito de deriva gênica e consequente perda da diversidade genética.

Os haplótipos africanos encontrados na Ilha de Marajó estão distribuídos entre os haplogrupos L0, L1, L2 e L3 (Tabela 23). Uma alta diversidade pode ser observada nos ramos L0, L2 e L3, com apenas 2 haplótipos compartilhados por 2 indivíduos (L2a1+16189+(16192)). Já em L1 ocorreram 2 haplótipos compartilhados por 2 indivíduos e 1 haplótipo compartilhado por 4 indivíduos. O mais frequente subhaplogrupo africano encontrado foi o L3e. Esse subhaplogrupo tem altas frequências também no Centro-Oeste africano, e registros históricos relatam que dessa região se originou a maior parte dos escravos vindos da África subsaariana para o Brasil (BANDELT *et al.*, 2001).

Apenas única linhagem de origem europeia foi encontrada, pertencente ao haplogrupo K. Uma busca na base de dados EMPOP (pesquisa em dezembro de 2021) não encontrou haplótipos idênticos, mas identificou haplótipos muito próximos, pertencentes a esse haplogrupo tipicamente europeu, já relatado em outras populações brasileiras (PALENCIA *et al.*, 2010; POLETTO *et al.*, 2018). Desde a chegada dos primeiros europeus o Brasil tem recebido várias ondas migratórias, principalmente da Península Ibérica, o que contribui para um maior componente ancestral europeu, sobretudo nas linhagens masculinas mas também nas de origem matrilinear. O fato da amostragem do presente estudo ter sido feito com pessoas de origem na Ilha de Marajó há três gerações pode ter minimizado a influência de migrações mais recentes para esse território.

3.4.3 Análise estatística – Região Controle do mtDNA

3.4.3.1 Diversidade Haplotípica

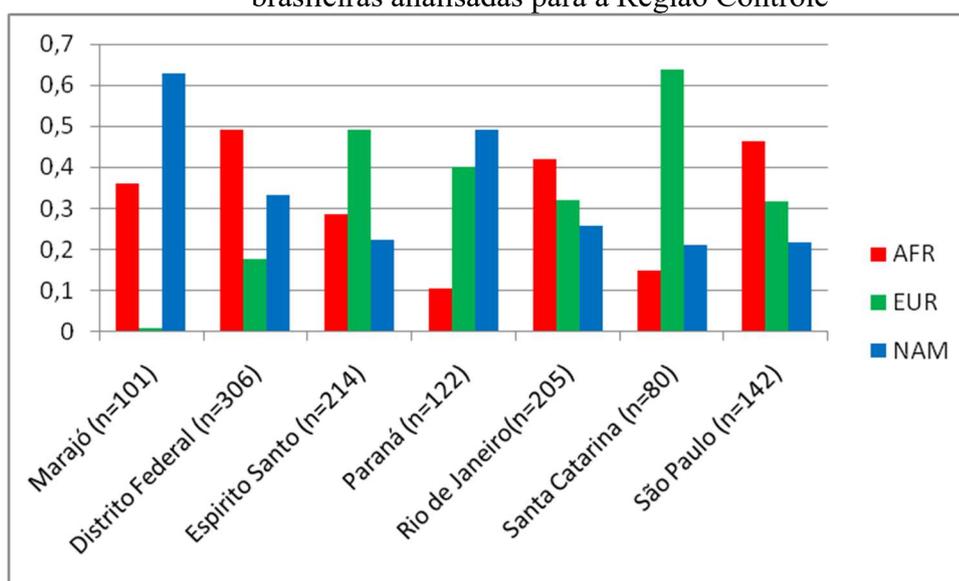
Uma diversidade haplotípica de $0,9897 \pm 0,0036$ foi observada para a Região Controle do mtDNA. Esse valor é inferior aos valores de diversidade observados para todas as populações brasileiras comparadas neste trabalho, superando apenas as populações da Argentina e da Colômbia, entre as populações da América do Sul consideradas (Tabela 24). Em relação às populações brasileiras usadas para comparação com a Ilha de Marajó, são consideradas grandes centros urbanos onde os eventos migratórios são constantes, propiciando intenso fluxo gênico, o que gera uma maior diversidade. Já a Ilha de Marajó, embora não se encontre isolada, devido ao transporte marítimo regular entre a Ilha e a capital

do Estado do Pará, Belém; não é considerada um grande centro urbano. Tais características podem explicar a menor diversidade encontrada. Não existem dados publicados, na literatura, de diversidade haplotípica para a Região Controle do mtDNA em populações da Região Norte do Brasil. O valor encontrado na análise de 48 amostras da Região Norte para HV-I apenas foi de 0,9840 (ALVES-SILVA, 2000); portanto, um valor inferior ao encontrado neste trabalho.

3.4.3.2 Distância Genética (F_{ST})

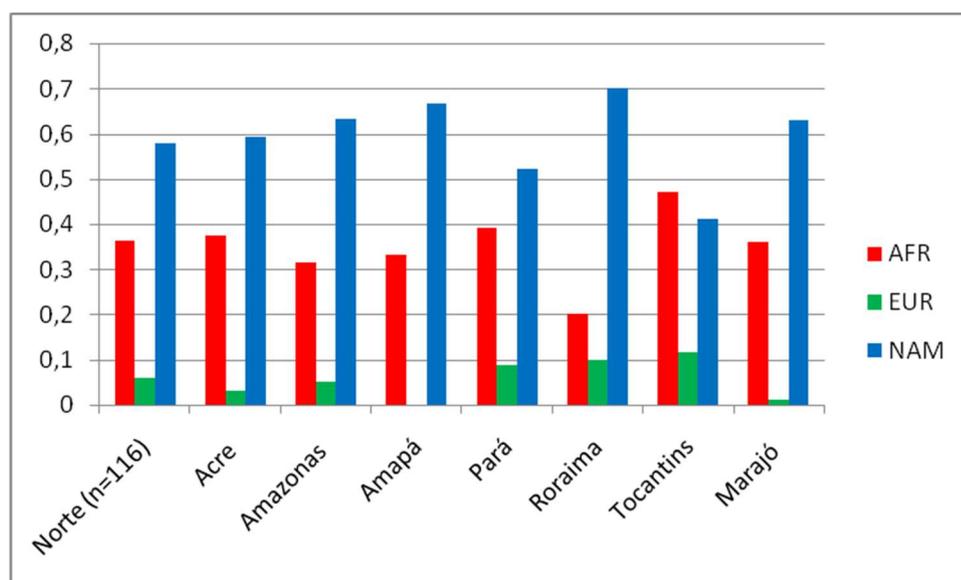
Na comparação entre a população de Marajó e as populações miscigenadas do Brasil e da América do Sul representadas na Tabela 11, quanto às distâncias genéticas, foram obtidos valores de F_{ST} e de probabilidades de não-diferenciação para os pares de populações, utilizando o *software* Arlequin versão 3.5. (Tabela 25). Os valores obtidos para as distâncias genéticas podem ser visualizados em duas dimensões através de gráficos MDS, na Figura 18. Valores de F_{ST} significativos foram observados entre a população da Ilha de Marajó e a maioria das populações analisadas, excetuando as populações da Argentina - Norte e Sul, Espírito Santo - Centro e Norte e Caracas.

Figura 16 - Representação gráfica das proporções de ancestralidades continentais dos haplogrupos de mtDNA de populações brasileiras analisadas para a Região Controle



Legenda: AFR – africana; EUR – européia; NAM – nativa americana
Fonte: A autora.

Figura 17 - Representação gráfica das proporções de ancestralidades continentais dos haplogrupos de mtDNA de populações da Região Norte analisadas para a Região Controle



Legenda: AFR – africana; EUR – européia; NAM – nativa americana
 Fonte: A autora.

Tabela 24 - Valores de Diversidade Haplótipica para populações do Brasil e da América do Sul, com base nos dados obtidos pelo sequenciamento da Região Controle completa

População	Nº de indivíduos/ Nº de haplótipos	Etnia	Diversidade Haplótipica
Marajó	101/73	Miscigenada	0,9897
Argentina, Norte	98/84	Miscigenada	0,9060
Argentina, Sul	47/45	Miscigenada	0,8780
Argentina, Centro	193/176	Miscigenada	0,9370
Colômbia, Medellín	94/48	Miscigenada	0,9407
Distrito Federal	306/264	Miscigenada	0,9988
Equador	156/128	Miscigenada	0,9958
Espírito Santo, Centro	54/53	Miscigenada	0,9993
Espírito Santo,	81/79	Miscigenada	0,9994
Espírito Santo, Norte	17/17	Miscigenada	1,0000
Espírito Santo, Sul	62/59	Miscigenada	0,9994
Paraguai	537/266	Miscigenada	0,9912
Paraná	122/108	Miscigenada	0,9976
Peru, Lima	83/82	Miscigenada	0,9997
Rio de Janeiro	205/184	Miscigenada	0,9994
Santa Catarina	80/64	Miscigenada	0,9930
São Paulo	142/129	Miscigenada	0,9983
Venezuela, Caracas	101/83	Miscigenada	0,9939

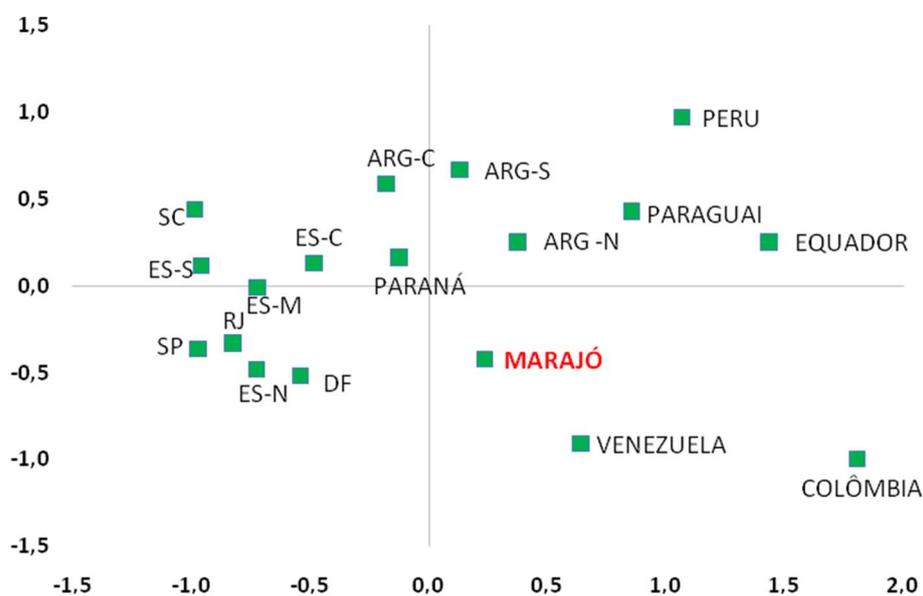
As maiores distâncias observadas em relação à população do Marajó foram com a população Colombiana (0,05899), Santa Catarina (0,04862), Peru (0,04377), Espírito Santo, Sul (0,0364), Equador (0,03545) e Argentina, Centro (0,03434).

Em relação à população de Santa Catarina, a mais distante entre todas as populações brasileiras, o fato de mais de 60 % de suas linhagens serem de origem europeia (Figura 16) podem explicar esse valor, uma vez que no Marajó apenas uma linhagem europeia foi encontrada (Tabela 23). As amostras representativas da região Sudeste ficaram relativamente agrupadas na representação bidimensional (Figura 18) e mostram valores intermediários de distância genética com o Marajó. As menores distâncias genéticas encontradas com o Marajó foram no Distrito Federal (0,01547) e no Paraná (0,01997). A população do Paraná, assim como a do Marajó, tem maior ancestralidade matrilinear nativa, e no Distrito Federal, apesar das linhagens africanas serem maioria, as nativas superam consideravelmente as europeias (Figura 16).

Dentre as populações Sul-Americanas comparadas, a população do Paraguai mostrou a menor distância genética significativa com o Marajó, o que está de acordo com dados publicados por Simão e colaboradores (2021), ao compararem haplótipos de mtDNA de origem nativa Americana entre diversos países da América do Sul. Análises de distâncias genéticas mostraram que as linhagens encontradas no Paraguai eram mais próximas das populações brasileiras do que dos outros países, sugerindo um *background* genético comum, possivelmente devido a migrações de povos nativos, ocorridas em período anterior à chegada dos europeus à América do Sul.

Este estudo possibilitou a geração de dados de haplótipos do mtDNA para a Região Controle completa, contribuindo para a construção do banco de dados da população brasileira. Além disso, é o primeiro estudo genético realizado com marcadores de mtDNA na população miscigenada da Ilha de Marajó. A alta frequência de linhagens matrilineas nativa Americanas reforça as evidências históricas do acasalamento direcional envolvendo homens europeus e mulheres indígenas no Brasil; uma estratégia usada como forma de ocupação (POLETTO *et al.*, 2018).

Figura 18 - Gráfico MDS das distâncias genéticas F_{ST} entre as populações brasileiras, incluindo o Marajó, e populações da América do Sul - RC - mtDNA



Legenda: Representação gráfica em duas dimensões, das distâncias genéticas entre as populações. Stress 0,0796317. ARG-Argentina; C-Centro; S-Sul; N-Norte; M-Metropolitano; ES - Espírito Santo; SC - Santa Catarina; RJ - Rio de Janeiro; SP - São Paulo; DF- Distrito Federal.

Fonte: A autora.

Tabela 25 - Matriz das distâncias genéticas baseadas no F_{ST} (abaixo da diagonal) entre todos os pares de populações comparadas com a população da Ilha de Marajó e os correspondentes valores de probabilidade de não diferenciação p (acima da diagonal) - Região Controle do mtDNA

	Argentina, Norte	Argentina, Centro	Argentina, Sul	Santa Catarina	ES, Sul	ES, Metro	ES, Centro	ES, Norte	Paraná	Rio de Janeiro	Distrito Federal	São Paulo	Colômbia	Peru	Venezuela	Paraguai	Equador	Marajó
Arg, N	*	0,01693	0,03168	<5,00E-6	<5,00E-6	<5,00E-6	0,00129	0,01188	0,03950	<5,00E-6	<5,00E-6	<5,00E-6	<5,00E-6	<5,00E-6	0,00010	0,03346	<5,00E-6	0,00079
Arg, C	0,00797	*	0,04118	0,00040	0,00040	0,00020	0,03584	0,00564	0,05940	<5,00E-6	<5,00E-6	<5,00E-6	<5,00E-6	<5,00E-6	<5,00E-6	<5,00E-6	<5,00E-6	<5,00E-6
Arg, S	0,01383	0,00906	*	<5,00E-6	<5,00E-6	<5,00E-6	0,00446	0,00604	0,00970	<5,00E-6	<5,00E-6	<5,00E-6	<5,00E-6	0,00010	0,00059	0,00535	0,00020	0,00287
SC	0,03948	0,01632	0,0426	*	0,15602	0,22325	0,13078	0,03208	0,00604	0,00010	<5,00E-6	0,00020	<5,00E-6	<5,00E-6	<5,00E-6	<5,00E-6	<5,00E-6	<5,00E-6
ES, Sul	0,03951	0,01979	0,03878	0,00369	*	0,73082	0,59341	0,28086	0,00644	0,04208	0,00218	0,06089	<5,00E-6	<5,00E-6	<5,00E-6	<5,00E-6	<5,00E-6	<5,00E-6
ES, Metro	0,03049	0,01527	0,03399	0,00222	-0,00252	*	0,41719	0,48866	0,01465	0,17513	0,01604	0,13949	<5,00E-6	<5,00E-6	<5,00E-6	<5,00E-6	<5,00E-6	<5,00E-6
ES, Centro	0,0254	0,00813	0,02136	0,00471	-0,00171	0,00025	*	0,27859	0,27057	0,01752	0,00446	0,01980	<5,00E-6	<5,00E-6	<5,00E-6	<5,00E-6	<5,00E-6	0,00050
ES, Norte	0,04362	0,03925	0,04291	0,02237	0,00454	-0,0014	0,00484	*	0,07494	0,81002	0,98069	0,94416	0,00188	0,00020	0,03980	0,00238	0,00099	0,18602
Paraná	0,00716	0,00394	0,01654	0,01074	0,01243	0,00876	0,00171	0,01699	*	<5,00E-6	<5,00E-6	<5,00E-6	<5,00E-6	<5,00E-6	<5,00E-6	0,00010	<5,00E-6	0,00030
RJ	0,04416	0,03372	0,0436	0,01493	0,00625	0,00222	0,00975	-0,00852	0,01775	*	0,01653	0,48827	<5,00E-6	<5,00E-6	<5,00E-6	<5,00E-6	<5,00E-6	<5,00E-6
DF	0,03784	0,03467	0,03782	0,02499	0,01317	0,00718	0,01296	-0,01493	0,01995	0,00361	*	0,18959	<5,00E-6	<5,00E-6	<5,00E-6	<5,00E-6	<5,00E-6	0,00020
SP	0,04833	0,03686	0,04537	0,01609	0,00559	0,00298	0,00957	-0,01285	0,01928	-0,00029	0,00127	*	<5,00E-6	<5,00E-6	<5,00E-6	<5,00E-6	<5,00E-6	<5,00E-6
Colômbia	0,06567	0,07983	0,08827	0,11632	0,11034	0,096	0,07974	0,10408	0,06892	0,0926	0,08214	0,1068	*	<5,00E-6	0,00079	<5,00E-6	0,00020	<5,00E-6
Peru	0,03783	0,04007	0,04492	0,06487	0,07356	0,05965	0,05326	0,08939	0,02934	0,06644	0,06464	0,0735	0,06564	*	<5,00E-6	0,00010	0,01465	<5,00E-6
Venezuela	0,02871	0,05081	0,03781	0,07977	0,06527	0,05374	0,04234	0,03118	0,04094	0,05333	0,03837	0,05885	0,0337	0,07018	*	<5,00E-6	<5,00E-6	0,00832
Paraguai	0,00628	0,02474	0,02099	0,06133	0,06294	0,05139	0,04521	0,06616	0,02048	0,06242	0,05534	0,06705	0,06068	0,0212	0,03846	*	0,00010	<5,00E-6
Equador	0,02967	0,04695	0,04553	0,08135	0,08483	0,07178	0,05913	0,08595	0,03504	0,07733	0,07079	0,08515	0,03573	0,01479	0,045	0,01618	*	<5,00E-6
Marajó	0,01839	0,03434	0,02311	0,04862	0,0364	0,02588	0,02616	0,00941	0,01997	0,02395	0,01547	0,02664	0,05899	0,04377	0,01416	0,02194	0,03545	*

Legenda: Valores de F_{ST} computados com 10100 permutações – diagonal de baixo; Valores de p – diagonal de cima. Valores de significância de $p \leq 0,000327$ após correção de Bonferroni grafados em vermelho, assim como os valores de F_{ST} significativos entre Marajó e as outras populações. Desvio-padrão $\leq 0,51$. Arg, - Argentina, N -

Norte; C - Centro; S - Sul; SC - Santa Catarina; ES, Sul - Espírito Santo, Sul; ES, Metro - Espírito Santo, Metropolitana; ES, Centro - Espírito Santo, Centro; ES, Norte - Espírito Santo, Norte; RJ - Rio de Janeiro; DF - Distrito Federal; SP - São Paulo.

3.5 Análise dos marcadores autossômicos (AIMs)

3.5.1 Genótipos - multiplex de 46 AIM-Indels e inferência de ancestralidade

Os perfis genéticos dos 160 indivíduos não aparentados, que possuíam os 4 avós nativos da Ilha de Marajó, com o multiplex composto de 46 marcadores AIM-Indels descrito por Pereira e colaboradores (2012a) estão descritos no Apêndice C. A Figura 19 apresenta um eletroferograma representativo da tipagem de uma amostra utilizada como controle positivo com esse multiplex.

Esses genótipos foram usados para inferir a ancestralidade global desses indivíduos através do *software* STRUCTURE v2.3.4 (PRITCHARD; STEPHENS; DONNELLY, 2000), considerando, portanto, a herança genética biparental de cada indivíduo. As proporções de ancestralidade considerando as origens continentais africana, europeia e nativa americana também se encontram listadas no Apêndice C.

Para esse grupo populacional como um todo, foi encontrada uma média de contribuições ancestrais de 40,3 % do componente nativo, 35,4 % do componente europeu e, por último, 24,3 % do componente africano (Figura 20). É possível observar uma alta variabilidade intrapopulacional na distribuição das ancestralidades, sugerindo um alto grau de miscigenação (Figura 20). Considerando-se, por exemplo, um valor mínimo de 70% para cada uma das ancestralidades individualmente, são encontrados 9 indivíduos com maior componente europeu, 6 majoritariamente nativos e apenas 1 com o componente africano mais prevalente.

A distribuição das proporções encontradas neste trabalho difere das distribuições encontradas para outras populações da Região Norte, nomeadamente Belém e Manaus, em trabalhos anteriores que utilizaram o mesmo conjunto de marcadores (PEREIRA *et al.*, 2012a; SALOUM DE NEVES MANTA *et al.*, 2013). Para essas populações a contribuição do componente europeu, apesar de ser inferior ao relatado para populações miscigenadas das outras regiões brasileiras, foi predominante (45,9 % - Manaus; 53,7 % - Belém), seguida da contribuição nativa (35,8 % - Manaus; 29,5 % - Belém) e por último, da africana (16,3 % - Manaus; 16,8 % - Belém). A proporção da ancestralidade africana na Ilha de Marajó se mostrou superior ao relatado para as populações de Manaus e Belém para esses marcadores autossômicos, podendo indicar que esse território pode ter sido uma espécie de refúgio para os africanos escravizados.

O estudo conduzido por Saloum de Neves Manta e colaboradores (2013) analisou populações urbanas das cinco regiões geopolíticas brasileiras e encontrou em todas elas uma ancestralidade europeia prevalente, com valores superiores a 74 %, com um gradiente crescente da Região Norte à Região Sul do País. Em segundo lugar, a ancestralidade nativa aparece apenas na Região Norte, chegando a representar o dobro da africana. Já nas outras regiões a ancestralidade africana ocupou a segunda posição, sendo mais prevalente no Nordeste, Rio de Janeiro e Minas Gerais. Nesse estudo, apenas na população nativa Terena, do Mato Grosso do Sul, e na população de Santa Isabel, que é uma população rural isolada na Amazônia, foi encontrada alta ancestralidade nativa America, acima de 75 %. Essas populações, por seu maior isolamento cultural ou geográfico, estão menos expostas ao influxo gênico trazido por indivíduos de outras etnias, como europeus e africanos.

Apesar da proximidade geográfica entre a Ilha de Marajó e a capital Belém, ambas no Estado do Pará, pode-se sugerir, a partir desses resultados, um menor influxo da ancestralidade europeia na população analisada neste trabalho para os AIM-Indels. Possivelmente, essa observação pôde ser melhor percebida devido ao critério de seleção utilizado na amostragem, em que se buscou há três gerações a ancestralidade da Ilha do Marajó. Embora atualmente haja uma certa mobilidade através do sistema de transporte marítimo regular, tal mobilidade há algumas gerações era menor devido às precárias e poucas embarcações. Uma outra análise feita com um total de 233 indivíduos, cujos avós e pais nem sempre eram nascidos na Ilha de Marajó, corrobora essa observação. Nesse grupo foi encontrada um valor de ancestralidade europeia de 39,3 %, ligeiramente maior que a nativa, de 38,1 %. Isso pode ser explicado por que além dos 160 indivíduos com 100 % dos avós nascidos no Marajó, tal grupo continha 27 indivíduos com 3 avós nascidos ali (75 %), 15 com 2 avós (50 %), 6 com apenas 1 dos avós (25 %) e 25 indivíduos com nenhum dos avós nascidos na Ilha (0 %). Sendo assim, tais indivíduos contribuíram nesse grupo, com uma maior ancestralidade europeia do que aqueles cujos avós eram todos nativos do Marajó. No conjunto de origens desses avós não nascidos no Marajó existem outros municípios do Estado do Pará majoritariamente, mas também outros Estados brasileiros, sobretudo do Nordeste e uma pequena minoria europeia.

3.5.2 Comparação das proporções de ancestralidade continental - marcadores autossômicos e os marcadores de linhagem

As proporções de contribuições de ancestralidade para o mtDNA, o cromossomo Y e os AIM-Indels observadas estão representadas graficamente na Figura 21, assim como o valor da média de entre as contribuições das linhagens maternas e paternas da Ilha de Marajó.

A contribuição genética das linhagens matrilineas do Marajó é predominantemente de ancestralidade nativa Americana, revelando a grande influência de mulheres indígenas na formação dessa população. Essa observação corrobora os dados históricos de povoamento desse território por povos de fala Tupi (PACHECO, 2010). As linhagens africanas também têm contribuição relevante (36,1 %), que se deve ao tráfico de escravos trazidos para o Brasil, nos séculos iniciais do processo de colonização. Praticamente não se observa linhagens europeias matrilineas no Marajó.

Em contraste estão as linhagens paternas, para as quais os Y-SNPs mostraram uma predominância de haplogrupos europeus, como ocorre de forma ampla no Brasil (RESQUE *et al.*, 2016). No Marajó as linhagens paternas de origem africana superam as linhagens nativas. Dados históricos evidenciam a substituição dos indígenas brasileiros pelos escravos africanos como mão de obra e o grande declínio populacional ocorrido entre os nativos, devido principalmente a conflitos com colonos e doenças (MORIM, 2014).

Analisando conjuntamente os resultados obtidos dessas linhagens uniparentais é possível observar uma assimetria nos acasalamentos, em termos de ancestralidade, envolvendo homens europeus e mulheres indígenas e também africanas. Tal assimetria é típica de populações miscigenadas da América do Sul, havendo variações ao longo desse subcontinente. Em relação às linhagens nativas, Paraguai, Chile e algumas regiões da Colômbia estão entre as que mais demonstram essa assimetria, enquanto o Brasil se diferencia pela maior contribuição maternal africana (SIMÃO *et al.*, 2021).

A população da Ilha de Marajó também foi analisada com 46 marcadores Indels autossômicos informativos de ancestralidade africana, europeia, asiática e nativa americana. Porém, nessa análise não foram incluídas populações de referência de origem asiática, mas apenas as referências parentais principais da população brasileira. Para populações com as mesmas origens continentais que as populações formadoras da população brasileira, ou seja, africana (Angola), europeia (Portugal) e nativa americana (tribos amazônicas), esse conjunto

de marcadores já havia demonstrado anteriormente alto componente africano (0,970), europeu (0,966) e nativo americano (0,945), respectivamente.

As proporções de ancestralidade obtidas, quando comparados com os valores médios das ancestralidades maternas e paternas, não se mostraram muito diferentes (Figura 21), havendo um pequeno decréscimo na ancestralidade nativa americana e pequeno acréscimo na africana e na europeia.

Valores maiores de ancestralidade europeia em relação ao valor médio das ancestralidades obtidas com os marcadores de linhagem podem ser explicados por eventos múltiplos de fluxo gênico devido a eventos migratórios mais recentes, que diluem o componente ancestral autossômico nativo, porém sem alterar de forma considerável as linhagens maternas uniparentais (SIMÃO *et al.*, 2021; XAVIER *et al.*, 2015). Nessa população da Ilha de Marajó, onde se buscou informações que remetem a três gerações passadas pode-se dizer que essas alterações nos autossomos é pouco evidente, embora populações brasileiras atuais e outras populações Sul Americanas tendam a mostrar um padrão característico que remete à fase inicial da sua colonização: uma ancestralidade nativa inicial e após algumas gerações, um aumento das linhagens matrilineas africanas acompanhado da ancestralidade europeia predominante nas linhagens paternas (ALVES-SILVA *et al.*, 2000).

Para uma interpretação confiável dos resultados genéticos de identificação forense e investigações de parentesco, é importante construir bancos de dados representativos de populações de referência. Neste estudo foram gerados dados haplotípicos para o STRs com o *Kit* comercial *Yfiler Plus*, amplamente utilizado na rotina de laboratórios forenses, e SNPs do Cromossomo Y, assim como da Região Controle completa do mtDNA, para a população urbana da Ilha de Marajó. Sendo assim, tais análises contribuem para o banco de dados brasileiro desses marcadores.

Este trabalho também gerou a primeira contribuição para a base de dados YHRD com a população miscigenada da Ilha de Marajó.

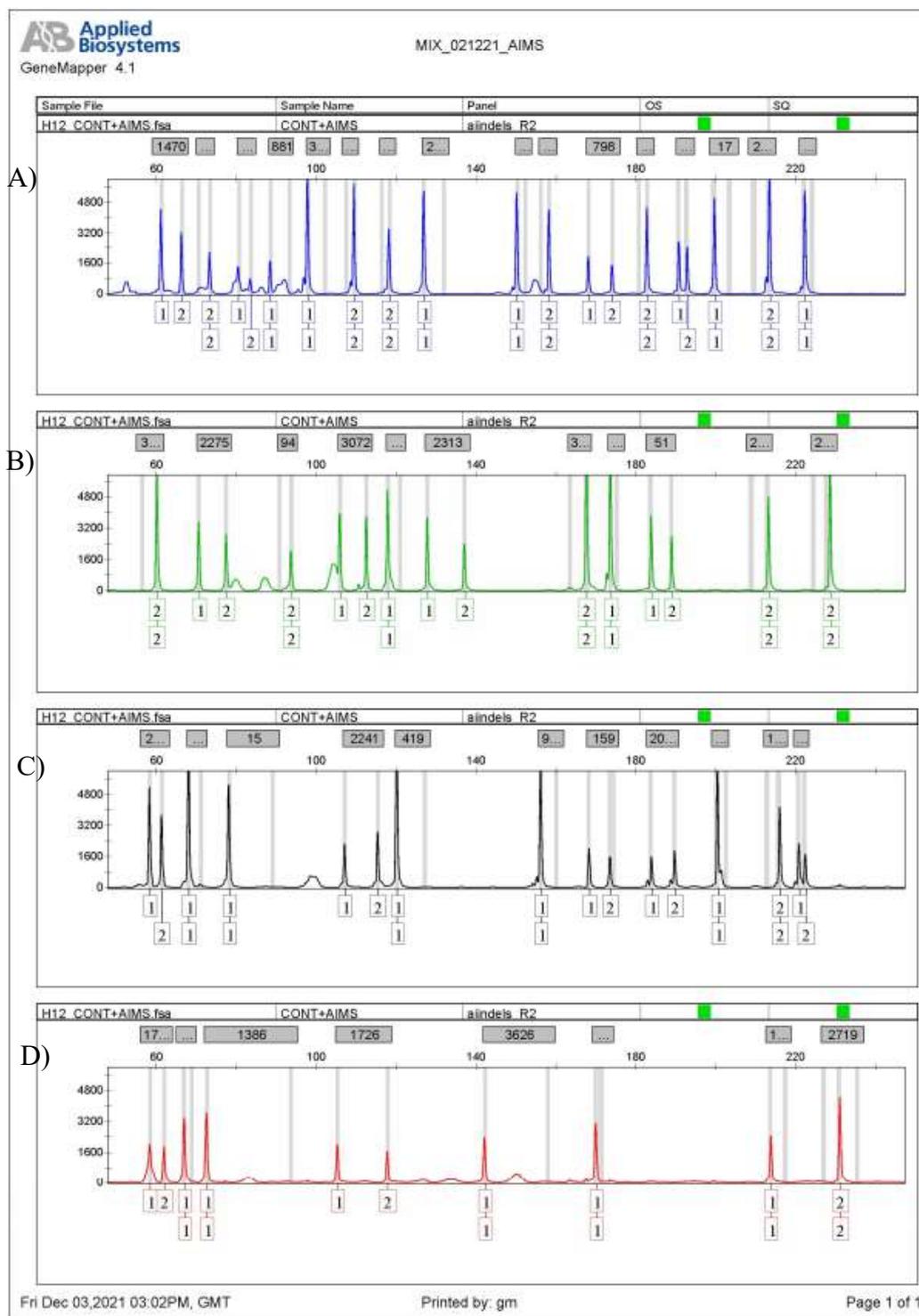
Diversos estudos com marcadores genéticos têm sido realizados, a fim de melhor caracterizar a diversidade étnica do Brasil, elucidando variações ao longo de seu território, e em comunidades que tenham especificidades demográficas e culturais. Para isso, diferentes tipos de marcadores vêm sendo usados, como os marcadores uniparentais, que trazem informações transmitidas ao longo das gerações por via paterna ou materna.

Vários tipos de marcadores autossômicos também têm sido usados na população brasileira, principalmente os STRs, SNPs e os Indels. Em geral, os achados com esses

marcadores encontraram uma influência nativa Americana maior na Região Norte e africana no Nordeste brasileiro. Já a contribuição europeia predomina em todo o Brasil, sobretudo no Sul; embora haja alguma variabilidade entre os resultados encontrados, geralmente devidos a estratégias de amostragens, número e tipo de marcador usado (CALLEGARI-JACQUES *et al.*, 2003; GODINHO *et al.*, 2008; LINS *et al.*, 2010; PENA *et al.*, 2011; SALOUM DE NEVES MANTA *et al.*, 2013).

Os resultados encontrados para o presente estudo com a população da Ilha de Marajó, situada à Região Norte, também mostraram uma ancestralidade nativa Americana maior para essa população do que para outras populações urbanas do Brasil, com os marcadores autossômicos utilizados, corroborando dados de estudos anteriores.

Figura 19 - Eletroferograma do perfil genético de uma amostra controle obtido através da tipagem com o Multiplex de 46 AIM-Indels

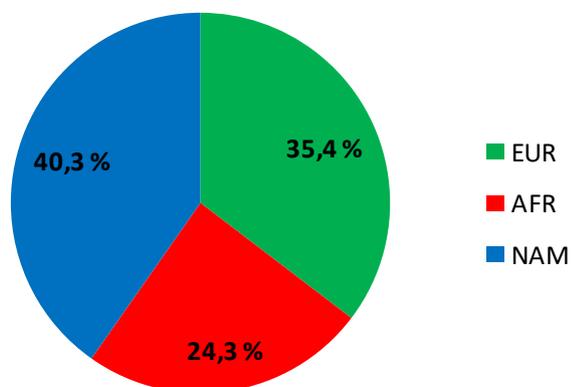


Legenda: Eletroferograma de genotipagem do Multiplex AIM-Indels, e seus 46 *loci*. Picos nomeados pelos algarismos, que identificam o alelo, representam a deleção (alelo 1) ou a inserção (alelo 2). (A) *primers* marcados com 6-FAM; (B) *primers* marcados com VIC; (C) *primers* marcados com NED; (D) *primers* marcados com TAZ. Eixo X - tamanho dos fragmentos em pb; Eixo Y - a intensidade de fluorescência (RFUs).

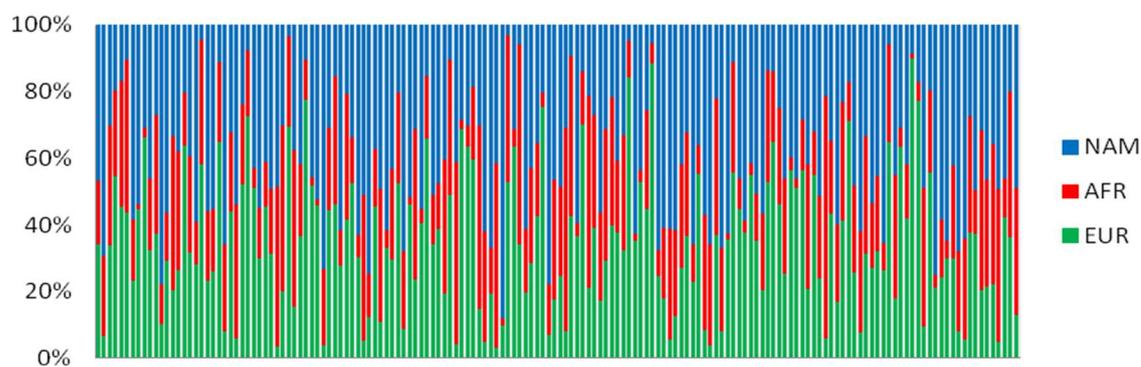
Fonte: Adaptado do *software* GeneMapper v. 4.1

Figura 20 - Representação gráfica das proporções de ancestralidade nativa americana (NAM), europeia (EUR) e africana (AFR) nos 160 indivíduos da população (A) e a variabilidade entre os indivíduos genotipados (B)

A)

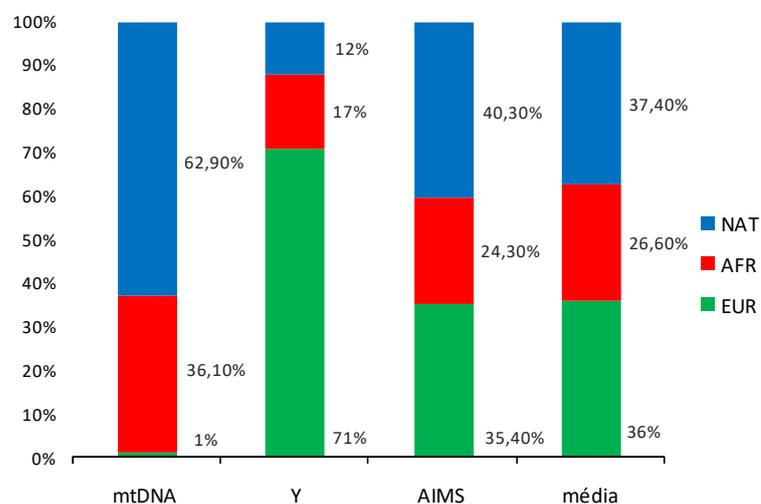


B)



Fonte: A autora, 2022.

Figura 21 - Representação gráfica das proporções de ancestralidade nativa americana (NAM), europeia (EUR) e africana (AFR) na população do Marajó com os marcadores do Cromossomo Y, do mtDNA, dos AIM-Indels, e do valor da média entre Cromossomo Y e mtDNA



Fonte: A autora, 2022.

CONCLUSÕES

Os resultados obtidos durante o desenvolvimento desta tese permitiram alcançar as seguintes conclusões:

- a) a população da Ilha de Marajó possui uma elevada diversidade intra-populacional para o conjunto de marcadores *Yfiler Plus* (0,9959), embora menor que o valor observado para outras populações urbanas brasileiras. Os dados haplotípicos gerados neste estudo, após a genotipagem de 97 indivíduos não aparentados contribuíram para a base de dados de perfis do cromossomo Y do Brasil;
- b) A população da Ilha de Marajó não apresentou distâncias genéticas significativas com as populações miscigenadas brasileiras e com a população europeia para os marcadores do *Yfiler Plus*, estando mais próxima geneticamente desta do que da população nativa americana e da africana;
- c) Os haplogrupos do Cromossomo Y encontrados na população da Ilha de Marajó, definidos por 47 Y-SNPs, foram majoritariamente europeus (71 %), seguidos dos africanos (17 %) e dos nativos americanos (12 %). Isto corrobora dados históricos acerca da colonização do território brasileiro por imigrantes europeus, e a vinda de grande número de indivíduos do continente africano, trazidos como escravos;
- d) A população da Ilha de Marajó possui uma diversidade intra-populacional para a Região Controle do mtDNA (0,9897), valor inferior ao encontrado para outras populações brasileiras comparadas. Os dados dos haplótipos gerados neste estudo contribuirão para a base de dados do Brasil;
- e) A população da Ilha de Marajó apresentou distâncias genéticas significativas com as populações miscigenadas brasileiras do Sudeste e do Sul do Brasil, e com populações da América do Sul, como Colômbia, Peru, Paraguai e Equador, para a Região Controle do mtDNA;
- f) Os haplogrupos majoritários do mtDNA encontrados na população da Ilha de Marajó foram os nativos americanos (62,9 %), seguidos dos africanos (36,1 %) e dos europeus (1 %). Isto corrobora dados históricos acerca dos povos indígenas que habitavam esse território antes da colonização europeia e da vinda de mulheres trazidas como escravas do continente africano;

- g) A população da Ilha de Marajó apresenta, com os 46 marcadores AIM-Indels, uma distribuição de ancestralidade de 40,3 % nativa americana, seguida de 35,4 % europeia e 24,3 % africana. Essa distribuição difere da encontrada para outras populações miscigenadas da Região Norte, nas quais o componente europeu é majoritário. Em outras populações brasileiras analisadas anteriormente, verifica-se um alto predomínio da ancestralidade europeia, com um gradiente crescente de Norte a Sul do país. Os dados genotípicos gerados neste trabalho contribuirão para a base de dados para esses marcadores.

REFERÊNCIAS

- AGUADÉ, M., N.; MIYASHITA, N.; LANGLEY, C. H. Restriction map variation at the zeste-tko region in natural populations of *Drosophila melanogaster*. ***Molecular Biology and Evolution*** . V. 6 (2): p. 123–130, 1989.
- ALLISON, A. C. Polymorphism and natural selection in human populations. ***Cold Spring Harbor Symposia on Quantitative Biology***. 29: 137–149, (1964).
- ALVES-SILVA, J. *et al.* The Ancestry of Brazilian mtDNA Lineages. ***American Journal of Human Genetics***, v. 67, n. 2, p. 444–461, 2000.
- AMIGO, J.; SALAS, A.; PHILLIPS, C. ENGINES: exploring single nucleotide variation in entire human genomes, ***BMC Bioinformatics***. v. 12, p.105, 2011.
- AMORIM, A.; BUDOWLE, B. **Handbook of Forensic Genetics: Biodiversity and Heredity in Civil and Criminal Investigation**. [s.l: s.n.].
- AMORIM, A.; FERNANDES, T., TAVEIRA, N. Mitochondrial DNA in human identification: a review. ***Peer J.***, v. 7, p. e7314, 2019.
- AMORIM, A.; PEREIRA, L. Pros and cons in the use of SNPs in forensic kinship investigation: A comparative analysis with STRs. ***Forensic Science International***, v.150(1):17–21, 2005.
- ANDERSON, S. *et al.* Sequence and organization of the human mitochondrial genome. ***Nature***, 290(5806), p. 457–465, 1981.
- ANDREWS, R. M. *et al.* Reanalysis and revision of the cambridge reference sequence for human mitochondrial DNA. ***Nature Genetics***, v.23, p. 147, 1999.
- ARAGÃO, G. **Characterization of male lineages in the Asháninka from Peru**. Tese de Mestrado – Universidade do Porto, Portugal, 2018.
- ATHEY, T. W. Haplogroup Prediction from Y-STR Values Using a Bayesian-Allele-Frequency Approach. ***Journal of Genetic Genealogy***, v. 2, n. 2, p. 34–39, 2006.
- AUTON, A. *et al.* A global reference for human genetic variation. ***Nature***, 526, 68–74, 2015.
- BAETA, M. *et al.* Mitochondrial diversity in Amerindian Kichwa and Mestizo populations from Ecuador. ***International Journal of Legal Medicine***, v. 126, p. 299-302, 2012.
- BALLANTYNE, K. N. *et al.* Mutability of Y-chromosomal microsatellites: Rates, characteristics, molecular bases, and forensic implications. ***American Journal of Human Genetics***, v. 87, n. 3, p. 341–353, 2010.
- BALLANTYNE, K. N. *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.

BAMSHAD, M. *et al.* Deconstructing the relationship between genetics and race. **Nature Reviews Genetics**, v. 5, p. 598-609, 2004.

BANDELT, H. J. *et al.* Phylogeography of the human mitochondrial haplogroup L3e: a snapshot of African prehistory and Atlantic slave trade. **Annals of Human Genetics**, 65:549–563, 2001

BARBOSA, A. B. G. *et al.* Mitochondrial DNA control region polymorphism in the population of Alagoas State, North-Eastern Brazil. **Journal of Forensic Sciences**, v. 53, n. 1, p. 142–146, 2008.

BATISTA, K. N.; NOGUEIRA, A. F. S. **Povos indígenas do Marajó: os Anajás**. ANAIS do III Colóquio de Letras da FALE/CUMB, Universidade Federal do Pará, Breves, 2016.

BAYE, T.; WILKE, R. Mapping genes that predict treatment outcome in admixed populations. **Pharmacogenomics Journal**, v. 10, 465–477, 2010.

BEGUN, D. J.; AQUADRO, C. F. Levels of naturally occurring DNA polymorphism correlate with recombination rate in *Drosophila melanogaster*. **Nature**, 356: 519–520, 1992.

BEGUN, D. J. *et al.* Population genomics: whole-genome analysis of polymorphism and divergence in *Drosophila simulans*. **PLoS Biology**, 5(11): e310, 2007.

BENDALL, K. E. *et al.* Heteroplasmic point mutations in the human mtDNA control region. **American Journal of Human Genetics**, 59(6):1276–1287, 1996.

BENSON, D. A. *et al.* GenBank. **Nucleic Acids Research**, v. 43: D30-5, 2015.

BERGER, B. *et al.* Reprint of: High resolution mapping of y haplogroup G in Tyrol (Austria). **Forensic Science International: Genetics**, v. 7, n. 6, p. 624–631, 2013.

BHANGALE, T. R. *et al.* Comprehensive identification and characterization of diallelic insertion–deletion polymorphisms in 330 human candidate genes. **Human Molecular Genetics**, v.14, p. 59–69, 2005.

BHANGALE, T. R.; STEPHENS, M.; NICKERSON, D.A. Automating resequencing-based detection of insertion–deletion polymorphisms. **Nature Genetics**, 38, 1457–1462, 2006.

BLACK, W. C. *et al.* Population genomics: genome-wide sampling of insect populations. **Annual Review of Entomology**, 46: 441–469, 2001.

BOBBILO, M. C. *et al.* Amerindian mitochondrial DNA haplogroups predominate in the population of Argentina: Towards a first nationwide forensic mitochondrial DNA sequence database. **International Journal of Legal Medicine**, v.124, p. 263–268, 2010.

BONATTO, S. L.; SALZANO, F. M. A single and early migration for the peopling of

the Americas supported by mitochondrial DNA sequence data. **Proceedings of the National Academy of Sciences USA**, 94: 1866–1871, 1997.

BRANDINI, S. *et al.* The Paleo-Indian entry into South America according to mitogenomes. **Molecular Biology and Evolution**, v. 35, p. 299–311, 2018.

BRINKMANN, B. *et al.* Mutation rate in human microsatellites: Influence of the structure and length of the tandem repeat. **American Journal of Human Genetics**, 62(6):1408–1415, 1998.

BRION, M. *et al.* Hierarchical analysis of 30 Y-chromosome SNPs in European populations. **International Journal of Legal Medicine**, v. 119, n. 1, p. 10–15, 2005.

BUDOWLE, B.; DAAL, A. Forensically relevant SNP classes. **Biotechniques**, v. 44(5), p. 603–608, 2008.

BUTLER, J. M. **Forensic DNA Typing - Biology, Technology and Genetics of STR Markers**, 2nd Edition, Elsevier, 2005.

CALLEGARI-JACQUES, S. M. *et al.* Historical genetics: spatiotemporal analysis of the formation of the Brazilian population. **American Journal of Human Biology**, v.15: 824–834 2003.

CARVALHO, M. *et al.* Analysis of paternal lineages in Brazilian and african populations. **Genetics and Molecular Biology**, v. 33, n. 3, p. 422–427, 2010.

CASILLAS, S.; BARBADILLA, A. Molecular Population Genetics. **Genetics**, v. 205, 1003–1035, 2017.

CASTRO DE GUERRA, D. *et al.* Sequence variation of mitochondrial DNA control region in North Central Venezuela. **Forensic Science International: Genetics**, v. 6, e131–3, 2012.

CAVALLI-SFORZA, L. L. The DNA revolution in population genetics. **Trends in Genetics**, 14(2):60–65, 1998.

CHARLESWORTH, B. Molecular population genomics: a short history. **Genetics Research**, v. 92: 397–411, 2010.

CHARLESWORTH, B.; CHARLESWORTH, D. Population genetics from 1966 to 2016. **Heredity**, v. 118, 2–9, 2017.

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 of the United States of America**, v. 106, n. 48, p. 20174–20179, 2009.

COBLE, M. D. *et al.* Mystery Solved: The Identification of the Two Missing Romanov Children Using DNA Analysis. **PLoS ONE**, 4(3): e4838, 2009.

COLLINS, F.S. *et al.* Construction of a general human chromosome jumping library, with application to cystic fibrosis. **Science**, 235: 1046–1049, 1987.

COLONNA, V. *et al.* A world in a grain of sand: human history from genetic data. **Genome Biology**, 12, 234, 2011.

CRUCIANI, F. *et al.* Strong intra- and inter-continental differentiation revealed by Y chromosome SNPs M269, U106 and U152. **Forensic Science International: Genetics**, v. 5, n. 3, p. 49-52, 2011.

CRUCIANI, F. *et al.* Tracing past human male movements in northern/eastern Africa and western Eurasia: New clues from Y-chromosomal haplogroups E-M78 and J-M12. **Molecular Biology and Evolution**, v. 24, n. 6, p. 1300–1311, 2007.

DE AZEVEDO, D. A. *et al.* Analysis of Y chromosome SNPs in Alagoas, Northeastern Brazil. **Forensic Science International: Genetics Supplement Series**, v. 2, n. 1, p. 421–422, 2009.

DE CAMPOS, F.; DOLHNIKOFF, M. **Atlas: História do Brasil**. 3. ed. [s.l: s.n.].

DE KNIJFF, P. Messages through Bottlenecks: On the Combined Use of Slow and Fast Evolving Polymorphic Markers on the Human Y Chromosome. **American Journal of Human Genetics**, v. 67:1055–1061, 2000.

DOBZHANSKY, T. A review of some fundamental concepts and problems of population genetics. **Cold Spring Harbor Symposia on Quantitative Biology**, v. 20: 1–15, 1955.

DOS REIS, R. S. *et al.* A view of the maternal inheritance of Espírito Santo populations: The contrast between the admixed and Pomeranian descent groups. **Forensic Science International: Genetics**, v. 40, p. 175–181, 2019.

EDSON, S. M. *et al.* Naming the dead - confronting the realities of rapid identification of degraded skeletal remains, **Forensic Science Review**, v. 16, p. 63–90, 2004.

ELLEGREN, H. Microsatellites: simple sequences with complex evolution. **Nature Reviews**, v. 5, n. June, p. 435–445, 2004.

EXCOFFIER, L. ; LISCHER, H. E. L. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. **Molecular Ecology Resources**, 10: 564-567, 2010.

FALUSH, D.; STEPHENS, M.; PRITCHARD, J.K. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. **Genetics**, 164, 1567–1587, 2003.

FAUSTO, B. **História do Brasil**. Edusp, 4 Ed., 639 p., 1996.

FAWCETT, J. A., *et al.* Population genomics of the fission yeast *Schizosaccharomyces pombe*. **PLoS One**, v. 9(8):e104241, 2014.

FIGUEIREDO, R. F. *et al.* Male-specific contributions to the Brazilian population of Espírito Santo. **International Journal of Legal Medicine**, v. 130, n. 3, p. 679-81, 2016.

FONDEVILA, M. *et al.* Revision of the SNPforID 34-plex forensic ancestry test: assay enhancements, standard reference sample genotypes and extended population studies. **Forensic Science International: Genetics**, v. 7, p. 63–74, 2013.

FORSTER, P. Ice Ages and the mitochondrial DNA chronology of human dispersals: a review. **Philosophical transactions of the Royal Society of London. Series B, Biological sciences**, v. 359, n. January, p. 255–264, 2004.

FORSTER, P. *et al.* Origin and evolution of native American mtDNA variation: a reappraisal. **American Journal of Human Genetics**, v. 59, p. 935–945, 1996.

FREGEL, R.; DELGADO, S. HaploSearch: a tool for haplotype-sequence two-way transformation. **Mitochondrion**, v.11, p. 366–367, 2011.

FREIRE, J.R.B. **Da Língua Geral ao Português: para uma história dos usos sociais das línguas na Amazônia**. Tese de Doutorado– Instituto de letras, Universidade do Estado do Rio de Janeiro, p. 239, 2003.

FREITAS, J.M. *et al.* Mitochondrial DNA control region haplotypes and haplogroup diversity in a sample from Brasília, Federal District, Brazil. **Forensic Science International: Genetics**, v. 40, p. 228-230, 2019.

FRIDMAN, C. *et al.* Haplotype diversity in mitochondrial DNA hypervariable region in a population of southeastern Brazil. **International Journal of Legal Medicine**, v. 128, n. 4, p. 589–593, 2014.

FRIEDLAENDER, J. S. *et al.* The genetic structure of Pacific Islanders. **PLoS Genetics**, v.4, e19, 2008.

FRUDAKIS, T. *et al.* A classifier for the SNP-based inference of ancestry. **Journal of Forensic Sciences**, v.48, p. 771–782, 2003.

GABRIEL, M. N. *et al.* Improved MtDNA Sequence Analysis of Forensic Remains Using a “Mini-Primer Set” Amplification Strategy. **Journal of Forensic Sciences**, v. 46, n. 2, p. 247–253, 2001.

GELEHRTER, T. D.; COLLINS, F. S. **Principles of medical genetics**. Williams & Wilkins, Baltimore, 1990.

GENTULA, M. C., NEVSKI, A. **NEVGEN Y-DNA Haplogroup Predictor**. Disponível em: <<http://www.nevgen.org/>>. Acesso em: out. 2019.

GEPPERT, M. *et al.* Identification of new SNPs in native South American populations by resequencing the y chromosome. **Forensic Science International: Genetics**, v. 15, p. 111–114, 2015.

GETTINGS, K. B. *et al.* A 50-SNP assay for biogeographic ancestry and phenotype prediction in the U.S. population. **Forensic Science International: Genetics**, v. 8, p. 101–108, 2014.

- GILL, P. *et al.* DNA commission of the International Society of Forensic Genetics: Recommendations on the evaluation of STR typing results that may include drop-out and/or drop-in using probabilistic methods. **Forensic Science International: Genetics**, 6(6):679–688, 2012.
- GILL, P. *et al.* Identification of the remains of the Romanov family by DNA analysis. **Nature Genetics**, 6(2):130–135, 1994.
- GJERTSON, D. W. *et al.* ISFG: Recommendations on biostatistics in paternity testing. **Forensic Science International: Genetics**, 1(3–4):223–231, 2007.
- GODINHO, N. M. O. *et al.* Regional patterns of genetic admixture in South America. **Forensic Science International: Genetics Supplement Series**, v.1: 329-330, 2008.
- GOLDSTEIN, D. B. *et al.* An evaluation of genetic distances for use with microsatellite loci. **Genetics**, v. 92, p. 6723–6727, 1995.
- GOMES, C. *et al.* Comparative evaluation of alternative batteries of genetic markers to complement autosomal STRs in kinship investigations: Autosomal indels vs. X-chromosome STRs. **International Journal of Legal Medicine**. 126(6):917–921, 2012.
- GOMES, V. *et al.* Digging deeper into East african human y chromosome lineages. **Human Genetics**, v. 127, n. 5, p. 603–613, 2010.
- GOODWIN, S.; MCPHERSON, J. D.; MCCOMBIE, W. R. Coming of age: ten years of next-generation sequencing technologies. **Nature Review Genetics**, v. 17: 333–351, 2016.
- GREENBERG, J. H; TURNER II, C. G.; ZEGURA, S. L. The settlement of the Americas: a comparison of linguistic, dental, and genetic evidence. **Current Anthropology**, v. 4, p. 477–497, 1986.
- GUSMÃO, L.; CARRACEDO, A. Y Chromosome-Specific STRs. **Profiles in DNA**, p. 3– 6, 2003.
- HARPENDING, H.; ROGERS, A. R. Genetic perspectives on human origins and differentiation. **Annual Review of Genomics and Human Genetics**, v.1, p. 361–385, 2000.
- HARRIS, H. Enzyme polymorphisms in man. **Proceedings of the Royal Society B: Biological Sciences**, v. 164: 298–310. 1966
- HARTL, D.L. **Princípios de genética de população**. 3. ed. Ribeirão Preto: FUNPEC, 2008.
- HOLLAND, M. M.; MCQUILLAN, M. R.; O’HANLON, K. A. Second generation sequencing allows for mtDNA mixture deconvolution and high resolution detection of heteroplasmy. **Croatian Medical Journal**, v. 52, p. 299–313, 2011.
- HOLSINGER, K. E.; WEIR, B. S. Genetics in geographically structured populations: defining, estimating and interpreting F_{ST} . **Nature Reviews Genetics**, v. 10, n. 9, p. 639-50, 2009.

HUERTA-SÁNCHEZ, E. *et al.* Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. **Nature**, 512, 194–197, 2014.

IACOVACCI, G. *et al.* Forensic data and microvariant sequence characterization of 27 Y-STR loci analyzed in four Eastern african countries. **Forensic Science International: Genetics**, v. 27, p. 123–131, 2017.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Brasil: 500 anos de povoamento**. Rio de Janeiro, 2000 (<https://brasil500anos.ibge.gov.br/territorio-brasileiro-e-povoamento/historia-indigena/os-numeros-da-populacao-indigena.html>) acesso em 30/10/2021.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Censo demográfico 2010**. Rio de Janeiro, p.1- 245, 2012.

IRWIN, J. A. *et al.* Investigation of heteroplasmy in the human mitochondrial DNA control region: A synthesis of observations from more than 5000 global population samples. **Journal of Molecular Evolution**, v. 68, n. 5, p. 516–527, 2009.

INTERNATIONAL SOCIETY OF GENETIC GENEALOGY. Disponível em: <<http://isogg.org/>>. Acesso em: 22 jan. 2017.

JAKOBSSON, M.; ROSENBERG, N. A. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. **Bioinformatics**, 23, 1801–1806, 2007.

JANNUZZI, J. *et al.* Male lineages in Brazilian populations and performance of haplogroup prediction tools. **Forensic Science International: Genetics**, Volume 44, 102163, 2020.

JOBLING, M. A. Y-chromosomal SNP haplotype diversity in forensic analysis. **Forensic Science International: Genetics**, v. 118, n. 2-3, p. 158-162, 2001.

JOBLING, M. A.; HURLES M.; TYLER-SMITH, C. **Human Evolutionary Genetics: Origins, Peoples & Disease**. Garland Science, New York, 2003

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

KALINOWSKI, S.T. The computer program STRUCTURE does not reliably identify the main genetic clusters within species: simulations and implications for human population structure. **Heredity**, v. 106, p. 625–632, 2011.

KARAFET, T. M. *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.

KAYSER, M. Uni-parental markers in human identity testing including forensic DNA analysis. **Biotechniques**, v. 43(6):S16-S21, 2007.

KAYSER, M. Forensic use of Y-chromosome DNA: a general overview. **Human Genetics**, v. 136, p 621–635, 2017.

KAYSER, M.; DE KNIJFF, P. Improving human forensics through advances in genetics, genomics and molecular biology. **Nature Reviews Genetics**, v. 12(3), p.179–192, 2011.

KERSBERGEN, P. *et al.* Developing a set of ancestry-sensitive DNA markers reflecting continental origins of humans. **BMC Genetics**, v. 10, p. 69, 2009.

KIDD, J. R. *et al.* Analyses of a set of 128 ancestry informative single-nucleotide polymorphisms in a global set of 119 population samples. **Investigative Genetics**, v. 5;2(1):1, 2011.

KIDD, K. K. *et al.* Current sequencing technology makes microhaplotypes a powerful new type of genetic marker for forensics, **Forensic Science International: Genetics**, v. 12, p. 215–224, 2014.

KIMURA, M.; CROW, J. F. The number of alleles that can be maintained in a finite population. **Genetics**, 49: 725–738, 1964.

KING, J. L. *et al.* High-quality and high-throughput massively parallel sequencing of the human mitochondrial genome using the Illumina MiSeq. **Forensic Science International: Genetics**, 12:128–135, 2014.

KINGMAN, J. F. C. On the genealogy of large populations. **Journal of Applied Probability**, v.19: 27–43, 1982a.

KINGMAN, J. F. C. The coalescent. **Stochastic Processes and their Applications**, v. 13: 235–248, 1982b.

KINGMAN, J. F. C. Origins of the coalescent. 1974–1982. **Genetics**, v. 156: 1461–1463, 2000.

KREITMAN, M. Nucleotide polymorphism at the alcohol dehydrogenase locus of *Drosophila melanogaster*. **Nature**, 304: 412–417, 1983.

LAO, O. *et al.* Proportioning whole-genome single-nucleotide-polymorphism diversity for the identification of geographic population structure and genetic ancestry. **American Journal of Human Genetics**, v.78, 680–690, 2006.

LARMUSEAU, M. H. D. *et al.* Towards a consensus Y-chromosomal phylogeny and Y-SNP set in forensics in the next-generation sequencing era. **Forensic Science International: Genetics**, v. 15, p. 39–42, 2015.

LEFFLER, E. M. *et al.* Revisiting an old riddle: what determines genetic diversity levels within species? **PLoS Biology**, 10: e1001388, 2012.

LEVY, S. *et al.* The diploid genome sequence of an individual human. **PLoS Biology**, v. 5, e254, 2007.

LEWONTIN, R. C. **The Apportionment of Human Diversity**. In: Dobzhansky T., Hecht M.K., Steere W.C. (eds) *Evolutionary Biology*. Springer, New York, NY, 1972.

- LEWONTIN, R. C.; HUBBY, J. L. A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. **Genetics**, v. 54: 595–609, 1966.
- LI, J. Z. *et al.* Worldwide human relationships inferred from genome-wide patterns of variation. **Science**, v. 319, p. 1100–1104, 2008.
- LINCH, C. A.; WHITING, D. A.; HOLLAND, M. M. Human hair histogenesis for the mitochondrial DNA forensic scientist. **Journal of Forensic Sciences**, v. 46(4):844–853, 2001.
- LINS, T. C. *et al.* Genetic composition of Brazilian population samples based on a set of twenty-eight ancestry informative SNPs. **American Journal of Human Biology**, v. 22: 187–192, 2010.
- LONDIN, E. R. *et al.* CoAIMs: a cost-effective panel of ancestry informative markers for determining continental origins. **PLoS One**, v.5, e13443, 2010.
- LOTT, M. T. *et al.* mtDNA variation and analysis using MITOMAP and MITOMASTER. **Current Protocols in Bioinformatics**, v.1(123): 1-26, 2013.
- LUIKART, G. *et al.* The power and promise of population genomics: from genotyping to genome typing. **Nature Review Genetics**, v. 4: 981–994, 2003.
- LYNCH, M. **The Origins of Genome Architecture**. Sinauer Associates, Inc., Sunderland, MA, 2007.
- MACAULAY, V. *et al.*, 2005 - Single, Rapid Coastal Settlement of Asia Revealed by Analysis of Complete Mitochondrial Genomes. **Science**, v. 308, p. 1034, 2005.
- MAPA DO TERRITÓRIO DO MARAJÓ. Disponível em: <<http://sit.mda.gov.br/download.php?ac=obterDadosBas&m=1501808>>. Acesso em 20 de novembro de 2021.
- MARTÍN, P. *et al.* A Spanish population study of 17 Y-chromosome STR loci. **Forensic Science International**, v.139, p. 231-5, 2004.
- MERRIWETHER, D. A.; ROTHHAMMER, F.; FERREL, R. E. Genetic variation in the New World: Ancient teeth, bone, and tissue as sources of DNA. **Experientia**, 50, 592–601, 1994.
- METZKER, M. L. Sequencing technologies - the next generation. **Nature Review Genetics**. v. 11: 31- 46, 2010.
- MIKKELSEN, M. *et al.* Massively parallel pyrosequencing of the mitochondrial genome with the 454 methodology in forensic genetics. **Forensic Science International: Genetics**, v. 12:30–37, 2014.
- MILLS, R. E. *et al.* An initial map of insertion and deletion (INDEL) variation in the human genome. **Genome Research**, 16(9):1182–1190, 2006.

MIRANDA-NETO, M. J. **Marajó: Desafio da Amazônia, Aspectos da Reação a Modelos Exógenos de Desenvolvimento**. Record, Rio de Janeiro, 180 pp, 1976.

MISHMAR, D. *et al.* Natural selection shaped regional mtDNA variation in humans. **Proceedings of the National Academy of Sciences USA**, v. 100:171–76, 2003.

MIZUNO, N. *et al.* A forensic method for the simultaneous analysis of biallelic markers identifying Y chromosome haplogroups inferred as having originated in Asia and the Japanese archipelago. **Forensic Science International: Genetics**, v. 4, n. 2, p. 73–79, 2010.

MORIM, J. **Ilha de Marajó**. Pesquisa Escolar Online, Fundação Joaquim Nabuco, Recife. 2014. Disponível em: < [http://basilio.fundaj.gov.br/pesquisaescolar/Ilha de Marajó](http://basilio.fundaj.gov.br/pesquisaescolar/Ilha%20de%20Maraj%C3%B3)>. Acesso em: 04/07/2016.

MORIOT, A. *et al.* Inferring biogeographic ancestry with compound markers of slow and fast evolving polymorphism. **European Journal of Human Genetics**, v. 26:1697–1707, 2018.

MORLING, N. *et al.* Paternity Testing Commission of the International Society of Forensic Genetics: recommendations on genetic investigations in paternity cases. **Forensic Science International**, v. 129(3):148–157, 2002.

MORTON, N. E. Parameters of the human genome. **Proceedings of the National Academy of Sciences USA**, v. 88, n. 17, p. 7474-7476, 1991.

MULLANEY, J. M. *et al.* Small insertions and deletions (INDELs) in human genomes. **Human Molecular Genetics**, v.19 (2), 131-136, 2010.

MULLER, H. J. Our Load of Mutations. **American Journal of Human Genetics**, v. 2(2), p. 111-176, 1950.

MUZZIO, M. *et al.* Software for Y-haplogroup predictions: A word of caution. **International Journal of Legal Medicine**, v. 125, n. 1, p. 143–147, 2011.

MYRES, N. M. *et al.* Y-chromosome short tandem repeat intermediate variant alleles DYS392.2, DYS449.2, and DYS385.2 delineate new phylogenetic substructure in human Y-chromosome haplogroup tree. **Croatian medical journal**, v. 50(3), p. 239–249, 2009.

NACHMAN, M. W.; CROWELL, S. L. Estimate of the mutation rate per nucleotide in humans. **Genetics**, 156(1):297–304, 2000.

NEI, M.; LI, W. H. Mathematical model for studying genetic variation in terms of restriction endonucleases. **Proceedings of the National Academy of Sciences USA**, v. 76: 5269–5273, 1979.

NGUIDI, M. P.R. **Herança genética materna dos principais grupos étnicos da Nigéria**. 2021. 108 f. Dissertação (mestrado em Biociências) - Instituto de Biologia Roberto Alcântara Gomes, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2021.

NOGUEIRA, T. L. S. *et al.* MTDNA control region analysis of North Brazilian population. **Forensic Science International: Genetics Supplement Series**, v. 6, n. September, p. e169–e170, 2017.

NOGUEIRO, I. *et al.* Phylogeographic Analysis of Paternal Lineages in NE Portuguese Jewish Communities. **American Journal of Physical Anthropology**, v. 141, p. 373–381, 2010.

NOGUERA, M. C. *et al.* Colombia's racial crucible: Y chromosome evidence from six admixed communities in the Department of Bolivar. **Annals of Human Biology**, v. 41, n. 5, p. 453–459, 2014.

OHTA, T. Slightly Slightly deleterious mutant substitutions in evolution. **Nature**, 246: 96–98, 1973.

OHTA, T.; GILLESPIE, J. H. Development of neutral and nearly neutral theories. **Theoretical Population Biology**, v.49: 128–142, 1996.

OLIVEIRA, A. M. *et al.* Male lineage strata of Brazilian population disclosed by the simultaneous analysis of STRs and SNPs. **Forensic Science International: Genetics**, v. 13, p. 264-268, 2014.

PAABO, S. Ancient DNA: Extraction, characterization, molecular cloning, and enzymatic amplification. **Proceedings of the National Academy of Sciences U S A**, v. 86(6):1939–1943, 1989.

PACHECO, A. S. **A Conquista do ocidente marajoara: índios, portugueses e religiosos em reinvenções históricas.** In; SCHAAN, D; MARTINS, C. (Orgs). Muito Além dos campos: arqueologia e história na Amazônica marajoara. Belém: GKONORA, 2010.

PAKENDORF, B; STONEKING, M. Mitochondrial DNA and human evolution. **Annual Review of Genomics and Human Genetics**, v. 6:1, 165-183, 2005.

PAKSTIS, A. J. *et al.* Mini-haplotypes as lineage informative SNPs and ancestry inference SNPs. **European Journal of Human Genetics**, v. 20, p. 1148–1154, 2012.

PALENCIA, L. *et al.* Mitochondrial DNA diversity in a population from Santa Catarina (Brazil): Predominance of the European input. **International Journal of Legal Medicine**, v. 124, p. 331-336, 2010.

PALHA, T. J. B. F. *et al.* Male Ancestry Structure and Interethnic Admixture in african-Descent Communities From the Amazon as Revealed by Y-Chromosome Strs. **American Journal of Physical Anthropology**, v.144, p. 471-478, 2011.

PALHA, T. *et al.* Disclosing the genetic structure of Brazil through analysis of male lineages with highly discriminating haplotypes. **PLoS ONE**, v. 7, n. 7, p. 1–8, 2012.

PARSON, W. *et al.* DNA Commission of the International Society for Forensic Genetics: Revised and extended guidelines for mitochondrial DNA typing. **Forensic Science International: Genetics**, v. 13, p. 134–142, 2014.

- PARSON, W. *et al.* Massively parallel sequencing of complete mitochondrial genomes from hair shaft samples. **Forensic Science International: Genetics**, v. 15, p. 8-15, 2015.
- PARSON, W.; BANDELT, H. Extended guidelines for mtDNA typing of population data in forensic science. **Forensic Science International: Genetics**, v.1, p. 13–19, 2007.
- PARSON, W.; DÜR, A. EMPOP-A forensic mtDNA database. **Forensic Science International: Genetics**, v. 1(2) p. 88-92, 2007.
- PENA, S. D. J. *et al.* DNA tests probe the genomic ancestry of Brazilians. **Brazilian Journal of Medical and Biological Research**, v. 42, n. 10, p. 870–876, 2009.
- PENA, S. D. J. *et al.* The genomic ancestry of individuals from different geographical regions of Brazil is more uniform than expected. **PLoS ONE**, v. 6, n. 2, p. e17063, 2011.
- PENA, S. D. J.; SANTOS, F. R.; TARAZONA-SANTOS, E. Genetic admixture in Brazil. **American Journal of Medical Genetics**, 184C:928–938, 2020.
- PEREIRA, L. *et al.* PopAffiliator: online calculator for individual affiliation to a major population group based on 17 autosomal short tandem repeat genotype profile. **International Journal of Legal Medicine**, v. 125, p. 629–636, 2011.
- PEREIRA, N. **A Ilha de Marajó – Estudo Econômico-Social**. Série Estudos Brasileiros nº 8. Ministério da Agricultura. Divisão de Caça e Pesca do D.N.P.A. Rio de Janeiro, 1956.
- PEREIRA, R. *et al.* A new multiplex for human identification using insertion/deletion polymorphisms. **Electrophoresis**, v.30, p. 3682–3690, 2009
- PEREIRA, R. *et al.* Straightforward Inference of Ancestry and Admixture Proportions through Ancestry-Informative Insertion Deletion Multiplexing. **PLoSOne**, v. 2 (1), p e29684, 2012a.
- PEREIRA, R. *et al.* A method for the analysis of 32 X chromosome insertion deletion polymorphisms in a single PCR. **International Journal of Legal Medicine**, 126(1):97–105, 2012b.
- PFAFFELHUBER, P. *et al.* How to choose sets of ancestry informative markers: A supervised feature selection approach. **Forensic Science International: Genetics**, 46, 102259, 2020.
- PHILLIPS, C. Forensic genetic analysis of bio-geographical ancestry. **Forensic Science International: Genetics**, v. 18, p. 49–65, 2015.
- PHILLIPS, C. *et al.* The SNPforID Consortium, inferring ancestral origin using a single multiplex assay of ancestry-informative marker SNPs. **Forensic Science International: Genetics**, v. 1, p. 273–280, 2007.
- PHILLIPS, C. *et al.* D9S1120, a simple STR with a common Native American-specific allele: forensic optimization locus characterization and allele frequency studies. **Forensic Science**

International: Genetics, v. 3, p. 7–13, 2008.

PHILLIPS, C. *et al.* Analysis of global variability in 15 established and 5 new European Standard Set (ESS) STRs using the CEPH human genome diversity panel. **Forensic Science International: Genetics**, v. 5, p. 155–169, 2011.

PHILLIPS, C. *et al.* Development of a novel forensic STR multiplex for ancestry analysis and extended identity testing. **Electrophoresis**, v. 34, p. 1151–1162, 2013.

PHILLIPS, C. *et al.* Building a forensic ancestry panel from the ground up: the EUROFORGEN Global AIM-SNP set. **Forensic Science International: Genetics**, v.11, p. 13–25, 2014a.

PHILLIPS, C. *et al.* New turns from old STaRs: enhancing the capabilities of forensic short tandem repeat analysis. **Electrophoresis**, v.35, 3173–3187, 2014b.

PHILLIPS, K.; MCCALLUM, N.; WELCH, L. A comparison of methods for forensic DNA extraction: Chelex-100® and the QIAGEN DNA Investigator Kit (manual and automated). **Forensic Science International: Genetics**, v. 6, p. 282–285, 2012.

PICKRELL, J. K.; REICH, D. Toward a new history and geography of human genes informed by ancient DNA. **Trends in Genetics**, v. 30, 377–389, 2014.

PINTO, N.; SILVA, P. V.; AMORIM, A. General derivation of the sets of pedigrees with the same kinship coefficients. **Human Heredity**, v. 70(3):194–204, 2010.

POLETTI, M. M. *et al.* Mitochondrial DNA control region diversity in a population from Parana state – increasing the Brazilian forensic database. **International Journal of Legal Medicine**, v. 133:347–351, 2018.

PRIETO, L. *et al.* The GHEP-EMPOP collaboration on mtDNA population data - A new resource for forensic casework. **Forensic Science International: Genetics**, v. 5, 146–151, 2011.

PRITCHARD, J. K., STEPHENS M.; DONNELLY, P. Inference of Population Structure Using Multilocus Genotype Data. **Genetics**, v. 155(2), p. 945–59, 2000.

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

QUINTANA-MURCI, L.; KRAUSZ, C.; MCELREAVEY, K. The human Y chromosome: function, evolution and disease. **Forensic science international**, v. 118, p. 169–181, 2001.

RAJEEVAN, H. *et al.* Introducing the forensic research/reference on genetics knowledge base, FROG-kb. **Investigative Genetics**. v.3:18, 2012a.

RAJEEVAN, H. *et al.* ALFRED: an allele frequency resource for research and teaching. **Nucleic Acids Research**, v. 40, D1010–1015, 2012b.

RAY, N. *et al.* A Statistical Evaluation of Models for the Initial Settlement of the American Continent Emphasizes the Importance of Gene Flow with Asia. **Molecular Biology and Evolution**, v. 27(2):337–345, 2010.

RÊGO, J. J. D. do. **Marcadores do cromossomo Y no estudo de linhagens paternas da população brasileira**. 2019. 140f. Tese de Doutorado – Programa de Pós-Graduação em Biociências, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2019.

REICH, D. *et al.* Denisova admixture and the first modern human dispersals into Southeast Asia and Oceania. **American Journal of Human Genetics**, v. 89, p. 516–528, 2011.

REICH, D. *et al.* Reconstructing Native American Population History. **Nature**, August 16; 488(7411): 370–374, 2012.

Relatório Analítico do Território do Marajó. Universidade Federal do Pará. Instituto de Ciências Sociais Aplicadas Belém, 2012

RESQUE, R. *et al.* Male Lineages in Brazil: Intercontinental Admixture and Stratification of the European Background. **Plos One**, v. 11, n. 4, p. e0152573, 2016.

RIBEIRO, D. O povo brasileiro. **São Paulo: Companhia das Letras**, p. 467, 1995.

RIBEIRO-DOS-SANTOS, A. K. *et al.* Dissimilarities in the process of formation of Curiaú, a semi-isolated Afro-Brazilian population of the Amazon region. **American Journal of Human Biology**, v. 14: 440-447, 2002.

ROOTSI, S. *et al.* Phylogeography of Y-chromosome haplogroup I reveals Rootsi 2004. **American Journal of Human Genetics**, v. 75, p. 128–137, 2004.

ROMIGUIER, J. *et al.* Comparative population genomics in animals uncovers the determinants of genetic diversity. **Nature**, 515: 261–263, 2014.

ROSENBERG, N. A. *et al.* Genetic structure of human populations. **Science**, 298: 2381–385, 2002.

ROSENBERG, N. A. *et al.* Informativeness of genetic markers for inference of ancestry. **American Journal of Human Genetics**, v. 73, p. 1402–1422, 2003.

ROSENBERG, N. A. *et al.* Clines, clusters, and the effect of study design on the inference of human population structure. **PLoS Genetics**, v.1, p. 70, 2005.

ROTAS DO TRÁFICO DE ESCRAVOS AFRICANOS PARA O BRASIL. Disponível em: <<http://www.slavevoyages.org/static/images/assessment/intro-maps/01.jpg>>. Acesso em 21 de outubro de 2021.

RUIZ-LINARES, A. *et al.* Admixture in Latin America: Geographic Structure, Phenotypic Diversity and Self-Perception of Ancestry Based on 7,342 Individuals. **PLoS Gene**, v. 10, n. 9, p. e1004572, 2014.

SALAS, A. *et al.* The making of the African mtDNA landscape. **American Journal of Human Genetics**, v. 71:1082–111, 2002.

SALOUM DE NEVES MANTA, F. *et al.* Revisiting the Genetic Ancestry of Brazilians Using Autosomal AIM-Indels. **PLoS ONE**, v. 8, n. 9, p. 1–11, 2013

SALZANO, F. M.; BORTOLINI, M. C. **The Evolution and Genetics of Latin American Populations**. Cambridge University Press, Cambridge, 512 p., 2002.

SALZANO, F. M.; SANS, M. Interethnic admixture and the evolution of Latin American populations. **Genetics and Molecular Biology**, v. 37, n. 1, p. 51-170, 2014.

SAMBROOK, J.; FRITCSH, E.F.; MANIATIS, T. **Molecular cloning: A laboratory manual**. 2. ed. New York : Cold Spring Harbor Laboratory, 1989. v3, 3104p.

SANGER, F., NICKLEN, S, COULSON, A. R. DNA sequencing with chain-terminating inhibitors. **Proceedings of the National Academy of Sciences U S A**, v. 74(12):5463-5467, 1977.

SANTOS, N. P. C. *et al.* Assessing individual interethnic admixture and population substructure using a 48-insertion-deletion (INDEL) ancestry- informative marker (AIM) panel. **Human Mutation**, v. 31, p. 184–190, 2010.

SCHAAN, A. P. *et al.* MtDNA structure: The women who formed the Brazilian Northeast. **BMC Evolutionary Biology**, v. 17, n. 1, p. 1–12, 2017.

SCHAFFNER, S. F. The X chromosome in population genetics. **Nature Reviews Genetics**, 5(1):43–51, 2004.

SCHLÖTTERER, C. *et al.* Sequencing pools of individuals — mining genome-wide polymorphism data without big funding. **Nature Reviews Genetics**, v.15, p. 749–763, 2014.

SCHNEIDER, H.; SALZANO, F. M. Gm allotypes and racial admixture in two Brazilian populations. **Human Genetics**, 53(1), 101–105,1979.

SCHNEIDER, P. M. Basic issues in forensic DNA typing. **Forensic Science International**, 88(1):17–22, 1997.

SHRIVER, M. D. *et al.* Ethnic-affiliation estimation by use of population-specific DNA. **American Journal of Human Genetics**, v. 60, p. 957–964, 1997.

SIMÃO, F. *et al.* Defining mtDNA origins and population stratification in Rio de Janeiro. **Forensic Science International: Genetics**, v. 34, 97-104, 2018.

SIMÃO, F. *et al.* The Ancestry of Eastern Paraguay: A Typical South American Profile with a Unique Pattern of Admixture. **Genes**, v.12, p. 1788, 2021.

SKALETISKY, H. *et al.* The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. **Nature**, v. 423, n. 6942, p. 825–837, 2003.

SOARES, P. *et al.* Correcting for Purifying Selection: An Improved Human Mitochondrial Molecular Clock. **American Journal of Human Genetics**, v. 84, p. 740–759, 2009.

STANGE, V. S. Stratification among European descent and admixed Brazilian populations of Espírito Santo for 27 Y-STRs. **Forensic Science International: Genetics**, v. 41, p. e20 - e22, 2019

STEIPER, M. E.. **Human Evolutionary Biology**. Cambridge University Press, New York, 2010.

STROBL, C. *et al.* Evaluation of the precision ID whole MtDNA genome panel for forensic analyses. **Forensic Science International: Genetics**, v. 35, 21 – 25, 2018.

SUTOVSKY, P. *et al.* Ubiquitin tag for sperm mitochondria. **Nature**, 402(6760):371–372, 1999.

SZIBOR, R. X-chromosomal markers: past, present and future. **Forensic Science International: Genetics**, 1(2):93–99, 2007.

TABOADA-ECHALAR, P. *et al.* The genetic legacy of the pre-colonial period in contemporary Bolivians. **PLoS One**, v. 8, p. e58980, 2013.

TAMM, E. *et al.* Beringian Standstill and Spread of Native American Founders. **PLoS One**, v. 2(9):e829, 2007.

THE 1000 GENOMES PROJECT CONSORTIUM. A map of human genome variation from population-scale sequencing. **Nature**, 467 (7319): 1061-73, 2010.

THE 1000 GENOMES PROJECT CONSORTIUM. An integrated map of genetic variation from 1,092 human genomes. **Nature**, 491 (7422): 56-65, 2012.

THE 1000 GENOMES PROJECT CONSORTIUM. A global reference for human genetic variation. **Nature**, 526 (7571): 68-74, 2015.

THE 1001 GENOMES CONSORTIUM. 1,135 Genomes Reveal the Global Pattern of Polymorphism in *Arabidopsis thaliana*. **Cell**, 166(2): 481–491, 2016.

TINEO, D.H. *et al.* Genetic characterization of 27 Y-STR loci in the native population of Ashaninka from Peru. **Forensic Science International: Genetics Supplement Series**, v. 5 e220–e222, 2015.

TORRONI, A. *et al.* Harvesting the fruit of the human mtDNA tree. **Trends in Genetics**, v. 22(6):339-345, 2006.

TULLY, G. *et al.* Considerations by the European DNA profiling (EDNAP) group on the working practices, nomenclature and interpretation of mitochondrial DNA profiles. **Forensic Science International**, v. 124, p. 83–91, 2001.

UNDERHILL, P. A., KIVISILD, T. Use of y chromosome and mitochondrial DNA population structure in tracing human migrations. **Annual Review of Genetics**, v. 41:539-564, 2007.

URQUHART, A. *et al.* Variation in short tandem repeat sequences-a survey of twelve microsatellite loci for use as forensic identification markers. **International Journal of Legal Medicine**, v. 107, p.13-20, 1994.

VALVERDE, L. *et al.* New clues to the evolutionary history of the main European paternal lineage M269: dissection of the Y-SNP S116 in Atlantic Europe and Iberia. **European journal of human genetics**, April, p. 1–5, 2015.

VAN OVEN, 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.

VAN OVEN, M.; KAYSER, M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. **Human mutation**, v. 30, n. 2, p. 386–394, 2009.

VERÍSSIMO, J. **Scenas da Vida Amazônica**. Ed. Lemmert, Rio de Janeiro, 1899.

VIGNAL, A. *et al.* A review on SNP and other types of molecular markers and their use in animal genetics. **Genetics Selection Evolution**, v. 34(3):275–305, 2002.

Voyages: The Trans-Atlantic Slave Trade Database, www.slavevoyages.org/assessment/estimates. Acessado em 02 de dezembro de 2021.

WALSH, P. S.; METZGER, D. A.; HIGUCHI, R. Chelex-100 as a Medium for Simple Extraction of DNA for PCR- Based Typing from Forensic Material. **BioTechniques**, v. 10, p. 506–513, 1991.

WANG, S. *et al.* Genetic variation and population structure in Native Americans. **PLoS Genetics**, v. 3, n. 11, p. 2049–2067, 2007.

WEBER, J. L. *et al.* Human diallelic insertion/deletion polymorphisms. **American journal of human genetics**, v. 71, n. 4, p. 854-862, 2002.

WESTEN, A. S. *et al.* Tri-allelic SNP markers enable analysis of mixed and degraded DNA samples. **Forensic Science International: Genetics**, v. 3, p. 233–241, 2009.

WHEELER, D. A. *et al.* The complete genome of an individual by massively parallel DNA sequencing. **Nature**, 452, 872–876, 2008.

WILLUWEIT S.; ROEWER, L. The new Y Chromosome haplotype reference database. **Forensic Science International: Genetics**, v. 15:43–48, 2015.

WILSON, M. R. *et al.* Validation of mitochondrial DNA sequencing for forensic casework analysis. **International Journal of Legal Medicine**, v. 108:68-74, 1995.

WOODING, S. *et al.* Directional migration in the Hindu castes: inferences from mitochondrial, autosomal and Ychromosomal data. **Human Genetics**, 115: 221–29, 2004.

XAVIER, C. *et al.* Admixture and Genetic Diversity Distribution Patterns of

NonRecombining Lineages of Native American Ancestry in Colombian Populations. **PLoS ONE**, v. 10, n. 3, p. e0120155, 2015.

YANG, N. *et al.* Examination of ancestry and ethnic affiliation using highly informative diallelic DNA markers: Application to diverse and admixed populations and implications for clinical epidemiology and forensic medicine. **Human Genetics**, 118(3–4):382–392, 2005.

Y CHROMOSOME CONSORTIUM. A nomenclature system for the tree of human Y chromosomal binary haplogroups. **Genome Research**, v.12:339–348, 2002.

ZAUMSEGEL, D.; ROTHSCHILD, M. A.; SCHNEIDER, P. A 21 marker insertion deletion polymorphism panel to study bio-geographical ancestry. **Forensic Science International: Genetics**, v.7, p. 305–312, 2013.

APÊNDICE A – Haplótipos *Yfiler Plus* de indivíduos da Ilha de Marajó (continua)

Amostra	DYS576	DYS389I	DYS635	DYS389II	DYS627	DYS460	DYS458	DYS19	YGATAH4	DYS448	DYS391	DYS456	DYS390	DYS438	DYS392	DYS518	DYS570	DYS437	DYS385	DYS449	DYS393	DYS439	DYS481	DYF387/S1	DYS533	(FA)	(FR)
MRJ 008	17	13	21	30	20	10	16	15	11	21	11	15	24	9	12	40	18	15	15, 19	31	12	13	22	38, 40	10	1	0,0103
MRJ 012	20	13	22	30	19	11	14	13	11	20	10	15	25	11	14	42	16	14	14, 17	30	13	11	24	34, 38	12	1	0,0103
MRJ 018	17	13	22	27	21	10	16	17	12	21	10	14	23	10	11	37	18	15	12, 12	29	13	11	23	35, 38	12	1	0,0103
MRJ 023	14	13	21	30	20	11	18	16	11	21	10	15	22	10	10	42	15	16	15, 15	31	13	12	21	39, 39	10	1	0,0103
MRJ_025	17	13	22	29	20	10	16	14	11	18	9	16	25	11	14	39	18	14	16, 18	28	13	12	22	35, 38	11	4	0,0412
MRJ_219	17	13	22	29	20	10	16	14	11	18	9	16	25	11	14	39	18	14	16, 18	28	13	12	22	35, 38	11		
MRJ_220	17	13	22	29	20	10	16	14	11	18	9	16	25	11	14	39	18	14	16, 18	28	13	12	22	35, 38	11		
MRJ_221	17	13	22	29	20	10	16	14	11	18	9	16	25	11	14	39	18	14	16, 18	28	13	12	22	35, 38	11		
MRJ 036	19	13	23	29	23	11	17	14	12	19	11	16	22	12	13	40	17	15	11, 14	29	13	13	22	35, 37	12	1	0,0103
MRJ 043	18	12	23	29	22	10	16	14	11	19	11	16	24	12	13	39	17	15	11, 15	30	13	11	27	35, 35	13	1	0,0103
MRJ 044	21	13	23	30	23	10	17	15	14	20	11	15	24	12	13	41	17	15	11, 16	31	13	15	20	35, 36	12	1	0,0103
MRJ 045	18	13	23	28	21	11	16	14	12	19	11	15	24	12	13	40	17	15	11, 14	29	13	12	23	35, 36	12	1	0,0103
MRJ 047	15	14	21	31	20	10	18	15	12	20	10	15	24	10	11	39	18	16	12, 13	30	13	11	26	34, 39	12	1	0,0103
MRJ 050	16	13	23	29	22	11	17	14	11	19	10	15	24	12	13	38	18	14	11, 14	29	13	11	22	35, 36	12	1	0,0103
MRJ 053	19	13	24	30	21	10	18	14	12	19	11	15	24	11	13	38	17	15	11, 15	25	13	13	22	34, 35	12	1	0,0103
MRJ 056	16	12	21	29	19	10	15	14	11	19	10	14	22	10	11	40	20	16	14, 17	28	13	12	26	37, 39	11	1	0,0103
MRJ_059	19	13	25	30	22	11	16	14	12	17	11	16	24	12	13	39	18	15	11, 14	30	13	11	22	35, 36	12	2	0,0206
MRJ_264	19	13	25	30	22	11	16	14	12	17	11	16	24	12	13	39	18	15	11, 14	30	13	11	22	35, 36	12		

APÊNDICE A – Haplótipos *Yfiler Plus* de indivíduos da Ilha de Marajó (continua)

Amostra	DYS576	DYS389I	DYS635	DYS389II	DYS627	DYS460	DYS458	DYS19	YGATAH4	DYS448	DYS391	DYS456	DYS390	DYS438	DYS392	DYS518	DYS570	DYS437	DYS385	DYS449	DYS393	DYS439	DYS481	DYF387S1	DYS533	(FA)	(FR)
MRJ 060	20	13	20	29	20	11	18	14	13	20	10	15	24	10	11	38	19	14	17, 19	35	13	11	21	36, 37	12	1	0,0103
MRJ 068	18	14	22	31	19	11	14	14	12	19	10	15	23	11	14	39	16	14	14, 16	28	14	13	25	36, 39	12	1	0,0103
MRJ 071	14	12	21	28	20	10	15	14	11	20	10	14	23	10	11	38	19	16	13, 15	28	13	11	25	37, 38	11	1	0,0103
MRJ 072	19	13	23	29	23	11	18	15	12	18	11	17	24	11	13	37	18	14	11, 13	29	12	13	22	35, 36	12	1	0,0103
MRJ 076	16	14	23	30	23	10	18	14	13	19	11	16	24	12	14	39	16	15	11, 15	28	13	12	21	35, 35	12	2	0,0206
MRJ_188	16	14	23	30	23	10	18	14	13	19	11	16	24	12	14	39	16	15	11, 15	28	13	12	21	35, 35	12		
MRJ 082	17	11	24	27	24	11	17	15	12	19	10	16	24	12	14	38	18	15	11, 15	31	13	13	23	36, 37	12	1	0,0103
MRJ 083	18	13	23	29	19	10	16	15	12	19	11	16	24	12	13	39	17	15	11, 14	28	13	12	22	35, 36	12	1	0,0103
MRJ 091	15	13	21	31	22	10	16	15	12	21	10	15	21	11	11	40	19	14	16, 17	28	13	11	27	36, 38	11	1	0,0103
MRJ 104	18	12	20	28	22	11	16	16	13	21	10	15	22	11	11	38	17	16	14, 15	27	14	11	24	36, 39	9	1	0,0103
MRJ 105	15	13	20	29	18	10	18	17	13	20	10	15	21	11	11	40	23	14	15, 17	31	14	12	24	36, 36	12	1	0,0103
MRJ 106	17	12	22	28	19	10	16	15	11	20	10	14	22	10	11	37	19	16	13, 14	27	13	14	25	37, 38	11	1	0,0103
MRJ 120	16	12	21	30	18	11	15	15	12	21	10	15	23	10	11	38	19	16	14, 15	28	14	14	23	36, 39	10	1	0,0103
MRJ 122	16	12	22	28	19	11	16	14	11	19	10	14	24	10	13	39	21	15	14, 14	28	13	11	27	37, 39	11	1	0,0103
MRJ 126	16	15	21	32	19	11	16	16	11	21	10	17	21	11	11	42	15	15	17, 17	31	14	12	25	38, 39	11	1	0,0103
MRJ 129	16	12	21	30	18	11	15	15	12	21	10	15	23	10	11	38	19	16	14, 15	28	14	13	23	36, 39	10	1	0,0103
MRJ 132	18	13	23	28	21	11	14	14	13	19	10	17	24	12	14	38	17	14	11, 14	28	12	12	22	36, 36	12	1	0,0103
MRJ 134	18	14	23	30	23	10	18	14	12	19	11	14	22	10	13	39	17	15	10, 15	28	13	12	22	35, 36	12	1	0,0103
MRJ 136	18	14	21	31	21	10	20	16	12	21	10	15	21	11	11	41	20	14	15, 16	29	14	12	25	36, 39	12	1	0,0103
MRJ 137	16	12	21	28	21	10	15	14	12	21	10	15	22	10	11	39	17	16	13, 15	28	13	11	22	37, 38	9	1	0,0103
MRJ 140	17	12	21	29	18	11	16	15	12	21	10	17	22	11	11	38	20	16	15, 15	27	14	13	22	39, 39	9	1	0,0103

APÊNDICE A – Haplótipos *Yfiler Plus* de indivíduos da Ilha de Marajó (continua)

Amostra	DYS576	DYS389I	DYS635	DYS389II	DYS627	DYS460	DYS458	DYS19	YGATAH4	DYS448	DYS391	DYS456	DYS390	DYS438	DYS392	DYS518	DYS570	DYS437	DYS385	DYS449	DYS393	DYS439	DYS481	DYF387S1	DYS533	(FA)	(FR)
MRJ 144	17	13	23	29	24	11	16	14	11	19	11	17	23	12	13	40	17	15	11, 14	29	13	12	21	35, 37	12	1	0,0103
MRJ 147	16	15	21	32	19	11	16	16	11	21	10	17	21	11	11	42	15	15	17, 17	31	14	12	25	38, 38	11	1	0,0103
MRJ 148	13	12	24	28	23	11	19	14	11	19	10	16	25	11	11	40	17	14	19, 20	32	13	11	24	37, 38	12	1	0,0103
MRJ 164	16	12	21	28	20	10	14	14	11	20	10	14	22	10	11	41	19	16	14, 14	29	13	12	26	36, 38	11	1	0,0103
MRJ 169	16	13	21	31	20	10	16	15	12	21	10	15	21	11	11	38	19	14	16, 17	26	13	11	28	37, 39	13	1	0,0103
MRJ 177	17	12	22	29	20	10	19	16	12	21	10	15	21	11	11	38	21	14	16, 17	31	13	13	29	37, 39	12	1	0,0103
MRJ 179	18	13	22	31	19	9	14	14	11	20	10	15	23	10	11	38	20	14	16, 18	31	13	12	23	35, 37	12	1	0,0103
MRJ 180	19	13	23	31	16	11	15	15	13	20	11	16	25	11	11	40	17	14	11, 14	32	13	10	23	37, 37	11	1	0,0103
MRJ 183	18	13	22	29	20	11	15	15	13	19	11	17	24	12	13	36	13	15	11, 13	29	12	12	23	35, 36	12	1	0,0103
MRJ 185	18	13	23	29	22	11	18	15	12	19	10	16	22	12	13	37	17	15	11, 14	30	12	12	21	35, 37	13	1	0,0103
MRJ 191	16	13	23	29	22	11	16	15	12	19	12	16	24	11	13	37	18	14	11, 14	28	13	12	22	36, 37	12	1	0,0103
MRJ 194	16	14	23	30	21	10	17	14	11	18	10	15	24	12	14	37	17	14	11, 14	29	13	12	22	36, 36	12	1	0,0103
MRJ 195	15	14	22	31	21	10	15	13	12	19	10	16	23	10	14	37	19	14	14, 17	29	13	12	25	34, 38	12	1	0,0103
MRJ 197	18	14	23	30	22	11	16	14	12	19	11	15	24	12	13	39	17	15	11, 14	30	13	12	22	35, 38	12	1	0,0103
MRJ 198	18	14	23	30	22	11	16	14	12	19	11	15	24	12	13	39	17	15	11, 14	30	13	13	22	35, 38	12	2	0,0206
MRJ 199	18	14	23	30	22	11	16	14	12	19	11	15	24	12	13	39	17	15	11, 14	30	13	13	22	35, 38	12		
MRJ 200	16	14	21	31	20	11	16	17	11	21	11	16	21	11	11	39	18	14	16, 19	33	16	12	24	38, 39	11	1	0,0103
MRJ 201	14	13	20	32	20	10	17	16	12	21	10	15	20	11	12	39	19	14	16, 17	27	13	12	26	37, 40	11	1	0,0103
MRJ 202	17	14	20	32	21	11	20.2	14	11	20	11	15	23	10	11	40	16	14	13, 20	24	12	11	25	0	11	1	0,0103
MRJ 208	17	12	21	28	14.2	10	16	14	11	20	10	14	23	10	11	39	19	16	14, 14	28	13	12	27	37, 38	11	1	0,0103
MRJ 213	18	13	22	31	18	10	17	13	11	19	10	15	24	11	14	36	16	13	15, 16	28	13	12	26	33, 38	11	1	0,0103

APÊNDICE A – Haplótipos *Yfiler Plus* de indivíduos da Ilha de Marajó (continua)

Amostra	DYS576	DYS389I	DYS635	DYS389II	DYS627	DYS460	DYS458	DYS19	YGATAH4	DYS448	DYS391	DYS456	DYS390	DYS438	DYS392	DYS518	DYS570	DYS437	DYS385	DYS449	DYS393	DYS439	DYS481	DYS387S1	DYS533	(FA)	(FR)
MRJ_224	17	13	23	28	22	10	17	14	12	19	10	15	25	12	13	37	17	15	11, 14	30	13	12	22	35, 36	12	2	0,0206
MRJ_238	17	13	23	28	22	10	17	14	12	19	10	15	25	12	13	37	17	15	11, 14	30	13	12	22	35, 36	12		
MRJ_225	15	13	21	30	20	11	16	17	12	19	10	15	23	9	14	37	18	14	15, 18	38	13	11	24	37, 38	12	1	0,0103
MRJ_226	17	13	23	28	21	10	18	14	12	20	10	17	24	12	13	40	17	15	11, 14	30	13	11	23	35, 36	14	1	0,0103
MRJ_236	16	13	21	30	19	11	15	13	11	20	10	17	23	10	11	38	19	14	16, 16	33	13	12	24	35, 37	13	1	0,0103
MRJ_237	17	13	23	28	22	10	17	15	12	19	10	15	25	12	13	37	17	15	11, 14	30	13	12	22	35, 36	12	1	0,0103
MRJ_246	18	13	23	29	19	10	13	14	11	20	10	15	24	11	14	38	17	14	13, 14	26	13	12	24	36, 37	11	1	0,0103
MRJ_247	17	11	24	27	24	11	17	14	12	19	10	16	24	12	14	38	18	15	11, 15	31	13	13	23	36, 37	12	1	0,0103
MRJ_252	17	13	23	29	19	10	13	14	11	20	10	15	24	11	14	38	17	14	13, 14	26	13	12	24	35, 37	11	1	0,0103
MRJ_249	18	12	23	28	22	12	17	14	11	18	10	16	24	12	13	38	18	14	11, 15	30	13	13	22	35, 36	12	3	0,0309
MRJ_250	18	12	23	28	22	12	17	14	11	18	10	16	24	12	13	38	18	14	11, 15	30	13	13	22	35, 36	12		
MRJ_251	18	12	23	28	22	12	17	14	11	18	10	16	24	12	13	38	18	14	11, 15	30	13	13	22	35, 36	12		
MRJ_254	19	14	23	30	19	11	17	14	12	19	11	17	24	13	13	38	15	15	12, 13	29	13	12	22	35, 36	12	1	0,0103
MRJ_255	13	12	24	28	23	11	18	14	11	19	10	16	25	11	11	40	17	14	19, 19	32	13	11	24	37, 38	13	4	0,0412
MRJ_304	13	12	24	28	23	11	18	14	11	19	10	16	25	11	11	40	17	14	19, 19	32	13	11	24	37, 38	13		
MRJ_306	13	12	24	28	23	11	18	14	11	19	10	16	25	11	11	40	17	14	19, 19	32	13	11	24	37, 38	13		
MRJ_307	13	12	24	28	23	11	18	14	11	19	10	16	25	11	11	40	17	14	19, 19	32	13	11	24	37, 38	13		
MRJ_256	17	11	24	27	24	11	17	14	12	20	10	16	24	12	14	38	18	15	11, 15	31	13	13	23	36, 37	12	1	0,0103
MRJ_261	14	13	20	31	20	9	17	15	12	20	10	15	21	12	12	37	20	14	17, 18	28	13	12	26	38, 39	11	1	0,0103
MRJ_262	16	13	23	29	23	11	17	14	12	18	12	16	24	12	13	36	17	15	11, 14	29	13	12	22	35, 35	13	1	0,0103

APÊNDICE A – Haplótipos *Yfiler Plus* de indivíduos da Ilha de Marajó (conclusão)

Amostra	DYS576	DYS389I	DYS635	DYS389II	DYS627	DYS460	DYS458	DYS19	YGATAH4	DYS448	DYS391	DYS456	DYS390	DYS438	DYS392	DYS518	DYS570	DYS437	DYS385	DYS449	DYS393	DYS439	DYS481	DYF387S1	DYS533	(FA)	(FR)
MRJ 270	17	13	23	30	24	11	17	14	12	19	11	15	24	12	13	37	19	14	12, 14	29	13	12	22	36, 37	12	1	0,0103
MRJ 271	18	14	21	30	18	11	17	13	12	20	9	15	24	10	11	45	22	14	13, 14	32	13	10	27	36, 37	11	1	0,0103
MRJ 273	19	13	20	30	21	11	17	15	11	19	10	14	25	9	11	39	20	15	14, 16	30	12	11	23	37, 37	12	1	0,0103
MRJ 275	20	13	23	29	21	11	17	15	13	20	11	15	24	12	13	37	18	15	11, 11	30	13	11	22	36, 36	12	1	0,0103
MRJ 278	18	13	23	29	22	11	16	15	11	20	11	15	24	12	13	38	17	15	12, 15	30	13	14	22	34, 36	12	1	0,0103
MRJ 279	18	13	23	29	20	11	18	13	12	19	11	14	23	12	13	36	17	15	11, 14	30	13	11	23	36, 36	12	1	0,0103
MRJ 282	17	13	25	29	23	11	17	15	12	19	10	15	24	12	13	38	17	14	11, 15	30	13	12	22	35, 37	12	1	0,0103
MRJ 283	16	13	21	29	23	10	14	15	11	20	9	15	23	9	11	39	20	14	13, 16	32	12	11	22	37, 40	12	1	0,0103
MRJ 288	18	13	21	28	24	11	18	14	12	19	11	15	24	12	13	40	21	14	12, 15	28	12	13	22	34, 35	13	1	0,0103
MRJ 290	16	13	19	30	18	11	16	16	11	21	10	16	21	10	11	39	19	14	16, 17	34	15	12	24	38, 39	12	1	0,0103
MRJ 292	18	14	24	30	20	11	17	14	12	19	10	14	24	12	13	38	18	14	11, 15	30	13	12	22	35, 36	12	1	0,0103
MRJ 293	15	14	22	31	21	10	15	13	12	19	10	16	23	10	14	37	18	14	14, 17	29	13	12	25	34, 38	12	1	0,0103
MRJ 295	20	13	25	29	22	11	18	14	11	20	11	13	24	12	13	39	17	15	11, 14	29	13	13	22	36, 36	12	1	0,0103
MRJ 299	18	14	23	30	22	11	16	14	12	19	11	15	24	12	13	39	17	15	11, 14	30	13	13	21	35, 38	12	1	0,0103
MRJ 305	18	13	24	29	21	10	16	14	12	19	11	17	24	12	13	37	17	15	11, 14	28	13	13	22	36, 36	11	1	0,0103
MRJ 308	17	13	23	29	22	11	18	14	12	19	11	16	24	12	13	38	18	15	12, 14	34	13	12	23	35, 36	11	1	0,0103
MRJ 310	19	13	23	29	23	11	16	14	11	18	11	17	24	12	13	38	17	14	11, 15	29	13	12	23	35, 36	12	1	0,0103

Legenda: Nesta tabela, estão listados os códigos das amostras, juntamente com seus respectivos haplótipos, suas frequências absolutas (FA) e suas frequências relativas (FR).
Os dados desta tabela foram gerados através da leitura dos eletroferogramas gerados pela eletroforese capilar no sequenciador automático ABI 3500 após amplificação por PCR com o *kit Yfiler Plus*.

APÊNDICE B – Haplogrupos e Haplótipos da Região Controle do mtDNA de indivíduos da Ilha de Marajó e sua origem continental (continua)

Amostra	HG	Origem	n	Haplótipo
MRJ_00A	L0a1b	AFR	1	16129A 16148T 16168T 16172C 16187T 16188G 16189C 16223T 16230G 16278T 16293G 16311C 16320T 93G 95C 185A 189G 236C 247A 263G 315.1C 523del 524del
MRJ_003	B2o	NAM	2	16092C 16183C 16189C 16217C 16519C 73G 263G 315.1C 499A
MRJ_166				
MRJ_008	L3b2	AFR	1	16124C 16223T 16278T 16362C 16390A 16527T 73G 263G 315.1C 523del 524del
MRJ_010	L2b2	AFR	1	16114A 16129A 16213A 16223T 16390A 73G 146C 150T 152C 182T 183G 195C 198T 204C 263G 309.1C 315.1C
MRJ_012	A2+(64)	NAM	3	16111T 16129A 16183C 16189C 16193.1C 16223T 16290T 16319A 16362C 16456A 16519C 64T 73G 143A 146C 153G 235G 263G 309.1C 315.1C 523del 524del
MRJ_088				
MRJ_226				
MRJ_013	D4	NAM	1	16223T 16239T 16286T 16325C 16362C 73G 210G 263G 315.1C 489C
MRJ_018	M / D1	EURASIA NAM	1	16192T 16223T 16239T 16325C 16362C 73G 210G 228A 263G 315.1C 489C
MRJ_023	C1b2	NAM	1	16223T 16298C 16325C 16327T 16519C 73G 152C 249del 290del 291del 309.1C 315.1C 489C 493G 523del 524del
MRJ_024	L2a1	AFR	1	16189C 16223T 16278T 16294T 16309G 16390A 16519C 73G 146C 152C 195C 263G 315.1C
MRJ_025	A2+(64)	NAM	1	16111T 16129A 16183C 16189C 16193.1C 16223T 16290T 16319A 16362C 16456A 16519C 64T 73G 143A 146C 153G 235G 263G 309.1C 309.2C 315.1C 523del 524del
MRJ_029	M / D1	EURASIA NAM	3	16223T 16239T 16325C 16362C 73G 210G 228A 263G 315.1C 489C
MRJ_049				
MRJ_078_2				
MRJ_032	A2+(64)	NAM	1	16111T 16223T 16290T 16319A 16362C 16519C 64T 73G 146C 153G 235G 263G 309.1C 309.2C 315.1C 523del 524del
MRJ_033	L3h1b2	AFR	1	16129A 16172C 16174T 16192T 16218T 16223T 16256A 16311C 16362C 16519C 73G 151T 152C 189C 195C 263G 294C 315.1C 523del 524del
MRJ_034	A2	NAM	1	16111T 16223T 16290T 16319A 16362C 73G 146C 153G 235G 263G 309.1C 315.1C 523del 524del
MRJ_035	C1b	NAM	1	16223T 16298C 16325C 16327T 64Y 73G 146C 151T 249del 263G 290del 291del 315.1C 489C 493G 523del 524del 573.1C 573.2C 573.3C 573.4C 573.5C
MRJ_036	A2+(64)	NAM	1	16111T 16189C 16223T 16290T 16319A 16362C 64T 73G 146C 153G 235G 263G 309.1C 315.1C 485C 523del 524del

APÊNDICE B – Haplogrupos e Haplótipos da Região Controle do mtDNA de indivíduos da Ilha de Marajó e sua origem continental (continua)

Amostra	HG	Origem	n	Haplótipo
MRJ_044	C1d+194	NAM	1	16051G 16093C 16223T 16298C 16325C 16327T 73G 194T 195C 249del 263G 290del 291del 315.1C 489C 523del 524del
MRJ_045	D1f	NAM	1	16142T 16179T 16219G 16223T 16249C 16325C 16362C 16519C 73G 263G 279C 309.1C 315.1C 489C
MRJ_047	L2a1c3a1	AFR	1	16223T 16278T 16290T 16294T 16309G 16390A 16519C 73G 143A 146C 152C 195C 198T 263G 309.1C 315.1C
MRJ_048	B2	NAM	1	16183C 16189C 16193.1C 16217C 16261T 16284G 16519C 73G 152C 263G 309.1C 315.1C 499A
MRJ_050	L1c5	AFR	2	16114A 16129A 16187T 16189C 16223T 16278T 16311C 16360T 16519C 73G 151T 152C 182T 186A 189C 247A 263G 291T 297G 315.1C
MRJ_188				316A 523del 524del
MRJ_053	C1b2	NAM	1	16223T 16298C 16325C 16327T 16519C 73G 249del 290del 291del 309.1C 315.1C 489C 493G 523del 524del
MRJ_054	L2a1+143 +16189+(16192)	AFR	1	16153A 16189C 16223T 16278T 16293C 16294T 16390A 16519C 73G 143A 146C 152C 195C 263G 264T 315.1C 534T
MRJ_056	L1c2b1b	AFR	4	16129A 16187T 16188T 16189C 16214T 16223T 16265C 16278T 16286A 16291T 16294T 16311C 16360T 16519C 16527T 73G 151T 152C
MRJ_151				182T 186A 189C 195C 198T 247A 263G 297G 309.1C 315.1C 316A 513A
MRJ_164				
MRJ_176				
MRJ_060	L2a1+16189 +(16192)	AFR	2	16189C 16192T 16223T 16278T 16294T 16296T 16309G 16390A 16519C 73G 146C 152C 195C 263G 309.1C 315.1C
MRJ_185				
MRJ_063	B4b	NAM	1	16150T 16183C 16189C 16193.1C 16217C 16519C 73G 199C 263G 309.1C 309.2C 315.1C 499A
MRJ_064	A2+(64)	NAM	1	16075C 16111T 16223T 16290T 16319A 16362C 16519C 64T 73G 146C 153G 235G 252C 309.1C 315.1C 523del 524del
MRJ_068	D1f1	NAM	1	16142T 16179T 16223T 16295T 16325C 16362C 16497G 16519C 73G 263G 309.1C 315.1C 489C 524.1A 524.2C
MRJ_071	B4b	NAM	1	16086C 16183C 16189C 16193.1C 16217C 16519C 73G 263G 315.1C 499A
MRJ_072	B2o	NAM	3	16092C 16183C 16189C 16193.1C 16217C 16519C 73G 263G 315.1C 499A
MRJ_120				
MRJ_126				
MRJ_073	A2h	NAM	1	16111T 16223T 16290T 16319A 16362C 16526A 64T 73G 146C 153G 195C 211G 235G 263G 309.1C 315.1C 523del 524del
MRJ_076	L3e1	AFR	1	16223T 16234T 16327T 73G 150T 189G 200G 263G 309.1C 315.1C
MRJ_082	L2c1a	AFR	1	16223T 16261T 16278T 16318G 16390A 16399G 73G 146C 150T 152C 182T 195C 198T 263G 309.1C 315.1C 325T 523del 524del

APÊNDICE B – Haplogrupos e Haplótipos da Região Controle do mtDNA de indivíduos da Ilha de Marajó e sua origem continental (continua)

Amostra	HG	Origem	n	Haplótipo
MRJ_083	L3b	AFR	1	16124C 16223T 16278T 16362C 16519C 73G 239C 263G 315.1C 523del 524del
MRJ_085	D1	NAM	3	16223T 16266T 16325C 16362C 16380T 73G 263G 315.1C 489C
MRJ_172				
MRJ_194				
MRJ_091	L3b	AFR	1	16124C 16223T 16278T 16311C 16362C 16519C 73G 263G 315.1C 523del 524del
MRJ_104	A2+(64)	NAM	1	16111T 16129A 16183C 16189C 16223T 16290T 16311C 16319A 16362C 16519C 73G 146C 153G 235G 263G 309.1C 309.2C 315.1C 523del 524del
MRJ_105	L2c1	AFR	1	16223T 16278T 16318G 16390A 73G 93G 146C 150T 152C 182T 195C 198T 263G 309.1C 315.1C 325T 523del 524del
MRJ_106	L3b	AFR	1	16093C 16124C 16223T 16278T 16362C 16519C 73G 263G 315.1C 523del 524del
MRJ_107	A2an	NAM	1	16111T 16188T 16223T 16290T 16319A 16352C 16362C 16519C 73G 146C 153G 178G 234R 235G 263G 309.1C 315.1C 472G 523del 524del
MRJ_110	D1f1	NAM	1	16142T 16179T 16223T 16295T 16325C 16362C 16497G 16519C 73G 263G 309.1C 309.2C 315.1C 489C 524.1A 524.2C
MRJ_115	D4	NAM	1	16172C 16223T 16325C 16362C 16519C 73G 195C 263G 309.1C 309.2C 315.1C 489C
MRJ_122	A2+(64)	NAM	1	16111T 16129A 16183C 16189C 16193.1C 16223T 16290T 16319A 16362C 16519C 73G 146C 153G 235G 263G 309.1C 315.1C 523del 524del
MRJ_123	B4b	NAM	1	16183C 16189C 16193.1C 16217C 16519C 73G 146C 199C 263G 309.1C 309.2C 315.1C 499A
MRJ_128	B4b	NAM	2	16093C 16182C 16183C 16189C 16193.1C 16217C 16362C 16519C 73G 263G 309.1C 309.2C 315.1C 499A 524.1A 524.2C 524.3A 524.4C
MRJ_305				
MRJ_129	C1b	NAM	1	16223T 16298C 16325C 16327T 73G 249del 263G 290del 291del 309.1C 315.1C 489C 493G 523del 524 del
MRJ_130	L0a1'4	AFR	1	16129A 16148T 16168T 16172C 16187T 16188G 16189C 16223T 16230G 16311C 16320T 93G 185A 189G 236C 247A 263G 315.1C 523del 524del
MRJ_131	D1a2	NAM	1	16172C 16223T 16258G 16259T 16325C 16362C 16519C 73G 143A 263G 315.1C 489C
MRJ_132	B4b	NAM	1	16051G 16129A 16183C 16189C 16193.1C 16217C 16519C 73G 152C 263G 309.1C 315.1C 499A
MRJ_134	L1b1a10	AFR	2	16126C 16172C 16187T 16189C 16223T 16264T 16270T 16278T 16311C 16519C 73G 151T 152C 182T 185T 195C 247A 263G 315.1C 357G 523del 524del
MRJ_142				
MRJ_136	L3e1	AFR	1	16223T 16327T 73G 150T 189G 200G 263G 309.1C 315.1C
MRJ_137	A2+(64)	NAM	1	16111T 16192T 16223T 16290T 16319A 16362C 16519C 64T 73G 146C 153G 199C 235G 263G 309.1C 315.1C 523del 524del

APÊNDICE B – Haplogrupos e Haplótipos da Região Controle do mtDNA de indivíduos da Ilha de Marajó e sua origem continental (continua)

Amostra	HG	Origem	n	Haplótipo
MRJ_140	B2	NAM	1	16182C 16183C 16189C 16217C 16261T 16284G 16519C 73G 152C 263G 309.1C 315.1C 499A
MRJ_141	C1b	NAM	1	16126C 16223T 16270T 16298C 16325C 16327T 71A 73G 249del 263G 290del 291del 309.1C 315.1C 489C 493G 523del 524del
MRJ_144	L2d1a	AFR	1	16093C 16129A 16183C 16189C 16193.1C 16278T 16300G 16311C 16354T 16390A 16399G 16519C 73G 146C 150T 195C 199Y 263G 315.1C 456T
MRJ_147	C1b	NAM	2	16172C 16223T 16298C 16325C 16327T 42C 73G 139C 207A 249del 263G 290del 291del 309.1C 309.2C 315.1C 489C 493G 523del 524del
MRJ_148				
MRJ_156	L3e1d	AFR	1	16176T 16223T 16327T 73G 150T 152C 189G 200G 263G 315.1C
MRJ_165	L1b1a+189	AFR	1	16126C 16187T 16189C 16223T 16264T 16270T 16278T 16311C 16327T 16519C 73G 152C 182T 185T 189G 195C 247A 263G 315.1C 357G 523del 524del
MRJ_169	B4b	NAM	1	16183C 16189C 16217C 16519C 73G 263G 315.1C 499A 524.1A 524.2C
MRJ_179	L3e3	AFR	1	16093C 16223T 16265T 16519C 73G 150T 195C 263G 315.1C 523del 524del
MRJ_180	K1b1a	EUR	1	16223T 16224C 16311C 16319A 16463G 16519C 73G 152C 263G 309.1C 315.1C
MRJ_183	A2an	NAM	1	16111T 16188T 16223T 16290T 16319A 16352C 16362C 16519C 73G 146C 153G 178G 235G 263G 309.1C 315.1C 472G 523del 524del
MRJ_191	A2+(64)	NAM	1	16223T 16290T 16319A 16362C 16519C 64T 73G 146C 153G 235G 263G 309.1C 309.2C 315.1C 523del 524del
MRJ_193	L0d1c	AFR	1	16187T 16189C 16223T 16230G 16234T 16243C 16311C 16519C 73G 146C 152C 195C 247A 315.1C 456T 498del 523del 524del
MRJ_195	C1	NAM	6	16051G 16189C 16223T 16265C 16294T 16298C 16325C 16327T 73G 249del 263G 290del 291del 315.1C 489C 492G 523del 524del
MRJ_197				
MRJ_198				
MRJ_202				
MRJ_204				
MRJ_207				
MRJ_196	A2+(64)	NAM	1	16111T 16223T 16290T 16319A 16362C 16519C 64T 73G 146C 153G 235G 263G 309.1C 315.1C 523del 524del
MRJ_200	L3d	AFR	1	16124C 16153A 16223T 16291T 73G 152C 263G 309.1C 315.1C 523del 524del
MRJ_205	A2ac	NAM	1	16105C 16111T 16213A 16223T 16290T 16293G 16319A 16362C 64T 73G 146C 153G 235G 263G 315.1C 523del 524del
MRJ_236	L3e2b	AFR	2	16172C 16189C 16223T 16320T 16519C 73G 150T 152C 195C 199C 263G 315.1C
MRJ_237				

APÊNDICE B – Haplogrupos e Haplótipos da Região Controle do mtDNA de indivíduos da Ilha de Marajó e sua origem continental
(conclusão)

Amostra	HG	Origem	n	Haplótipo
MRJ_246	L3e2b1	AFR	1	16164G 16172C 16189C 16223T 16320T 16519C 73G 146C 150T 195C 263G 315.1C 523del 524del
MRJ_261	L3e4a	AFR	1	16051G 16223T 16264T 16519C 73G 150T 263G 309.1C 315.1C 493G 523del 524del
MRJ_288	C1d	NAM	1	16051G 16223T 16294T 16298C 16325C 16327T 73G 249del 263G 290del 291del 315.1C 489C 523del 524del
MRJ_290	D1f	NAM	1	16142T 16179T 16219G 16223T 16249C 16325C 16362C 16519C 73G 263G 279C 309.1C 309.2C 315.1C 489C
MRJ_293	C1b	NAM	1	16223T 16271C 16292T 16298C 16311C 16325C 16327T 73G 249del 263G 290del 291del 309.1C 309.2C 315.1C 489C 493G 498T 523del 524del
MRJ_299	B4b	NAM	1	16150T 16183C 16189C 16193.1C 16217C 16519C 73G 199C 263G 309.1C 315.1C 499A
MRJ_304	A2a1	NAM	3	16111T 16126C 16223T 16259T 16290T 16319A 16362C 16519C 73G 146C 153G 235G 263G 315.1C 523del 524del
MRJ_306				
MRJ_307				

APÊNDICE C - Genótipos dos 46 AIM-Indels de indivíduos da Ilha de Marajó e as proporções de ancestralidade continental (continua)

Amostra	1470	777	196	881	3122	548	659	2011	2929	593	798	1193	1871	17	2538	1644	3854	2275	94	3072	772	2313	397	1636	51	2431	2264	2256	128	15	2241	419	943	159	2005	250	1802	1607	1734	406	1386	1726	3626	360	1603	2719	EUR	AFR	NAM							
MRJ_054	1	1	2	?	1	2	1	1	1	2	2	2	2	1	2	1	2	1	1	1	1	2	2	1	1	2	2	1	2	1	1	1	1	1	1	1	1	1	2	2	1	2	1	2	1	0,062	0,404	0,535								
	2	1	2	?	2	2	2	2	1	2	2	2	2	1	2	1	2	2	2	2	1	1	2	2	1	2	2	2	1	2	2	2	2	2	2	2	1	2	1	1	2	2	2	3	2	1	0,524	0,239	0,237							
MRJ_056	2	1	2	1	1	1	2	2	1	2	1	2	1	1	1	1	1	1	1	2	1	1	1	1	1	1	2	2	2	1	1	1	1	1	1	1	1	2	2	2	2	1	2	2	1	2	2	0,729	0,198	0,073						
	2	2	2	1	1	2	1	1	1	2	2	2	2	1	1	1	2	2	2	2	2	1	2	1	1	1	2	1	1	2	1	1	1	1	2	1	2	1	2	1	1	1	2	1	1	1	1	2	0,513	0,060	0,428					
MRJ_057	2	2	2	1	1	2	1	1	1	2	1	2	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	2	1	2	2	2	2	2	2	2	2	2	2	2	2	1	1	2	1	1	1	1	0,303	0,149	0,548					
	2	2	2	1	1	2	2	1	1	2	2	2	2	1	2	2	2	2	2	2	2	2	2	2	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0,457	0,134	0,408				
MRJ_060	1	1	1	1	1	1	2	1	1	2	2	2	1	1	2	2	2	1	2	1	1	1	1	1	1	1	1	2	1	2	1	1	1	1	1	1	1	2	1	1	1	2	1	1	1	2	1	1	1	0,317	0,192	0,491				
	2	2	1	2	1	2	2	1	1	2	2	2	2	1	2	2	2	2	2	2	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0,038	0,480	0,481			
MRJ_063	1	1	1	1	1	1	1	1	1	2	2	1	2	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	2	1	1	1	1	1	1	1	0,204	0,495	0,301				
	2	2	2	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0,698	0,268	0,033		
MRJ_064	1	1	1	1	1	1	1	1	1	2	2	2	1	1	1	2	2	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0,156	0,471	0,374			
	2	2	2	1	1	2	2	1	1	2	2	2	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0,369	0,214	0,417		
MRJ_068	2	1	1	1	1	2	1	1	1	2	2	2	2	1	2	2	1	1	1	1	1	1	2	1	1	2	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0,369	0,214	0,417		
	2	2	1	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0,369	0,214	0,417
MRJ_072	1	2	1	1	1	1	2	1	1	2	1	1	2	1	2	1	2	2	1	2	1	2	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0,369	0,214	0,417	
	2	2	1	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0,369	0,214
MRJ_073	2	2	1	1	1	1	2	1	1	2	1	2	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0,369	0,214	0,417
	2	2	2	1	1	2	2	2	1	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0,369	0,214
MRJ_075	2	1	1	1	1	1	1	1	1	2	2	2	1	2	2	1	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0,369	0,214	0,417
	2	2	2	1	1	2	2	1	1	2	2	2	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0,369	0,214
MRJ_076	2	1	2	1	1	1	1	2	1	2	2	2	1	1	2	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0,369	0,214	0,417
	2	1	2	1	2	2	2	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0,369	0,214

APÊNDICE C - Genótipos dos 46 AIM-Indels de indivíduos da Ilha de Marajó e as proporções de ancestralidade continental (continua)

Amostra	1470	777	196	881	3122	548	659	2011	2929	593	798	1193	1871	17	2538	1644	3854	2275	94	3072	772	2313	397	1636	51	2431	2264	2256	128	15	2241	419	943	159	2005	250	1802	1607	1734	406	1386	1726	3626	360	1603	2719	EUR	AFR	NAM			
MRJ_078_2	1	1	1	1	1	2	2	1	1	1	2	1	1	1	1	1	2	2	2	1	1	2	1	1	1	2	1	2	2	2	1	1	1	1	1	1	2	2	1	1	1	1	1	1	1	1	1	2	0,778	0,118	0,103	
MRJ_078_2	2	2	2	2	1	2	2	2	1	2	2	2	2	2	2	1	2	2	2	1	1	2	2	1	2	2	2	2	2	2	2	2	2	1	1	2	2	1	2	2	2	1	1	2	2	2	2	2	2	0,520	0,025	0,454
MRJ_079_2	2	1	2	1	1	2	2	1	1	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	1	1	2	2	2	2	2	2	2	0,459	0,019	0,522	
MRJ_083	2	2	2	1	1	2	2	1	2	2	2	2	2	2	2	2	2	2	2	1	2	2	2	1	2	1	2	2	2	2	2	2	2	1	2	2	2	2	2	2	1	2	1	2	2	2	2	0,043	0,227	0,730		
MRJ_084	2	1	1	1	1	1	1	1	1	2	1	2	2	1	2	2	2	2	2	1	1	1	1	1	1	1	2	2	1	2	1	1	1	1	1	1	1	1	2	1	1	1	1	2	1	1	2	0,446	0,249	0,305		
MRJ_085	2	2	2	1	1	2	2	2	1	2	2	2	2	2	2	1	2	2	2	2	2	2	2	1	2	2	2	2	2	2	2	2	1	2	2	1	2	2	1	2	2	1	2	1	2	2	2	0,464	0,385	0,151		
MRJ_086	1	2	1	1	1	1	2	1	1	1	1	1	1	2	2	2	2	1	2	1	1	2	1	2	1	2	1	1	2	1	1	2	1	1	2	1	1	1	1	1	2	1	1	1	1	1	1	0,281	0,105	0,613		
MRJ_088	2	1	2	1	1	2	2	2	2	2	2	2	2	1	2	2	2	2	2	1	1	1	1	1	1	2	2	2	2	2	2	2	1	1	1	1	2	2	2	2	2	1	2	2	2	2	2	2	0,419	0,377	0,204	
MRJ_091	1	2	1	1	1	2	1	1	1	2	2	2	2	2	2	1	1	2	1	1	1	1	1	1	2	1	2	1	2	1	2	1	1	1	1	1	1	1	2	1	1	1	2	1	1	1	1	1	0,528	0,136	0,337	
MRJ_093	2	2	2	1	1	2	2	2	1	2	2	2	2	2	2	2	2	2	2	1	1	2	1	1	2	2	2	2	2	2	2	1	2	1	1	1	2	1	2	1	2	2	2	2	2	2	2	2	2	0,304	0,069	0,627
MRJ_095	2	1	1	1	1	2	1	1	1	2	2	2	1	2	2	1	2	2	1	1	1	2	1	1	1	2	2	1	2	1	2	1	1	1	2	2	2	2	2	2	1	1	1	1	1	1	1	1	0,056	0,436	0,508	
MRJ_096	2	2	2	2	1	2	2	2	1	2	2	2	2	1	2	1	2	2	2	2	2	2	2	1	2	1	2	2	2	2	2	2	2	1	2	2	2	2	2	2	2	1	2	2	2	2	2	2	2			

APÊNDICE C - Genótipos dos 46 AIM-Indels de indivíduos da Ilha de Marajó e as proporções de ancestralidade continental (continua)

Amostra	1470	777	196	881	3122	548	659	2011	2929	593	798	1193	1871	17	2538	1644	3854	2275	94	3072	772	2313	397	1636	51	2431	2264	2256	128	15	2241	419	943	159	2005	250	1802	1607	1734	406	1386	1726	3626	360	1603	2719	EUR	AFR	NAM										
MRJ_097	1	1	1	1	1	2	2	1	1	2	1	2	2	1	1	1	2	2	1	1	1	1	1	1	1	1	2	2	1	2	1	1	1	2	2	1	1	1	1	1	2	2	2	2	1	1	1	0,126	0,130	0,744									
	2	1	2	1	1	2	2	1	1	2	2	2	2	1	2	2	2	2	1	1	1	1	2	2	1	1	2	2	1	2	2	2	1	1	1	2	2	2	2	1	2	1	1	1	2	2	3	2	2	2	0,456	0,176	0,368						
MRJ_099	2	1	2	1	1	2	2	1	1	2	1	2	2	1	1	1	1	2	2	1	1	1	1	1	1	1	1	2	2	1	2	1	1	1	1	1	1	2	2	2	1	1	1	1	1	1	1	1	1	1	0,113	0,397	0,489						
	2	2	2	1	1	2	2	1	1	2	2	2	2	1	2	1	2	2	2	2	1	1	2	2	1	1	2	2	2	2	2	2	2	1	2	2	2	2	1	2	2	2	2	2	2	3	2	2	2	0,334	0,052	0,613							
MRJ_101	2	1	1	1	1	2	1	1	1	1	2	2	1	1	2	1	1	1	1	1	1	1	1	1	1	1	2	2	1	2	1	1	1	1	1	1	1	1	2	1	1	1	1	2	1	1	1	1	1	1	0,300	0,271	0,429						
	2	2	2	1	1	2	1	2	1	2	2	2	2	2	2	2	2	2	2	2	2	1	2	2	2	2	2	2	2	2	2	2	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0,529	0,271	0,200					
MRJ_102	1	1	2	1	1	1	1	1	1	2	1	2	2	1	1	1	2	2	1	1	1	1	1	1	1	1	1	1	1	2	1	2	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	0,092	0,231	0,677					
	2	?	2	1	1	2	1	2	1	2	2	2	2	1	2	2	2	2	2	2	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0,465	0,022	0,513					
MRJ_104	2	1	1	1	1	1	1	1	1	2	1	1	1	2	1	1	2	2	2	2	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0,237	0,451	0,312				
	2	2	2	1	1	2	2	2	2	2	2	2	2	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0,409	0,042	0,548			
MRJ_105	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	1	1	1	1	1	1	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0,663	0,186	0,152			
	2	2	1	1	2	2	2	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0,344	0,148	0,508		
MRJ_106	2	1	1	1	1	2	1	1	2	2	2	2	1	1	1	1	2	1	1	1	1	1	1	1	1	1	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0,237	0,451	0,312		
	2	2	2	1	1	2	2	1	1	2	2	2	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0,409	0,042	0,548	
MRJ_107	1	1	1	1	1	1	2	1	1	2	2	2	2	1	2	1	2	2	2	2	1	1	1	1	1	1	2	1	2	1	1	2	1	1	1	1	1	1	1	1	1	2	1	1	2	1	1	1	1	1	1	1	1	1	1	0,663	0,186	0,152	
	2	2	2	2	1	2	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0,344	0,148	0,508
MRJ_108	2	2	2	2	1	1	2	1	1	2	1	2	1	1	2	1	2	2	2	1	1	1	1	1	1	1	2	3	1	1	1	1	1	1	1	1	1	1	1	2	1	1	2	1	1	2	1	1	1	1	1	1	1	1	1	0,344	0,148	0,508	
	2	2	2	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0,344	0,148	0,508
MRJ_109	2	1	1	1	1	2	1	1	2	2	2	2	2	1	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0,344	0,148	0,508	
	2	2	2	2	1	2	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0,344	0,148
MRJ_110	1	1	1	1	1	1	2	1	1	2	2	2	2	1	2	1	2	2	2	2	1	1	1	1	1	1	2	1	2	1	1	2	1	1	1	1	1	1	1	1	1	2	1	1	2	1	1	2	1	1	1	1	1	1	1	0,663	0,186	0,152	
	2	2	2	2	2	2	2	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0,344	0,148
MRJ_111	2	2	1	1	1	2	2	1	1	1	2	1	2	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0,344	0,148	0,508

APÊNDICE C - Genótipos dos 46 AIM-Indels de indivíduos da Ilha de Marajó e as proporções de ancestralidade continental (continua)

Amostra	1470	777	196	881	3122	548	659	2011	2929	593	798	1193	1871	17	2538	1644	3854	2275	94	3072	772	2313	397	1636	51	2431	2264	2256	128	15	2241	419	943	159	2005	250	1802	1607	1734	406	1386	1726	3626	360	1603	2719	EUR	AFR	NAM			
MRJ_111	2	2	2	1	1	2	2	1	2	2	2	1	2	1	2	2	2	2	1	1	1	1	1	1	2	2	2	2	1	2	2	2	1	2	2	2	2	2	2	1	1	2	2	1	1	2	2					
MRJ_114	1	1	1	1	1	1	2	1	1	2	2	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	2	1	1	2	1	2	1	1	2	2	1	2	2	2	1	1	2	1	1	1	1	1	0,389	0,137	0,474	
	2	2	2	1	1	2	2	2	1	2	2	2	2	1	2	2	2	2	2	2	2	2	2	2	1	1	2	3	1	2	2	2	2	2	2	2	2	1	2	2	2	1	2	2	1	1	1	2				
MRJ_115	2	2	2	1	1	1	1	1	1	1	1	2	1	1	1	1	1	2	1	1	1	1	1	1	1	2	2	1	1	2	1	2	1	1	1	2	1	2	1	2	2	1	2	2	1	1	1	2	0,195	0,403	0,402	
	2	2	2	1	1	1	2	2	1	2	2	2	2	1	2	2	2	2	2	2	2	2	1	2	2	2	2	2	2	1	2	2	2	1	2	1	2	1	2	2	2	2	2	2	1	1	2					
MRJ_116	1	1	1	1	1	2	2	2	1	2	1	1	2	1	2	1	1	1	2	1	1	1	1	1	2	2	2	1	1	1	2	2	1	2	1	2	1	2	1	2	1	1	2	1	1	1	2	1	0,494	0,403	0,103	
MRJ_116	2	2	2	2	2	2	2	2	1	2	2	2	2	1	2	1	2	2	2	2	2	1	1	2	2	2	2	2	2	2	1	2	2	2	2	1	2	1	2	2	1	1	2	1	1	2	1	2	1	0,494	0,403	0,103
MRJ_120	1	1	1	1	1	1	1	2	1	1	2	2	2	1	2	1	2	1	2	2	1	1	1	1	2	1	2	1	1	2	1	1	1	2	1	2	1	1	1	1	1	1	1	1	2	1	1	2	0,045	0,545	0,410	
	2	2	2	1	1	2	2	2	1	2	2	2	2	1	2	1	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	1	2	2	2	2	2	2	2	2	2	2	2			
MRJ_121	1	2	2	1	1	1	1	1	1	2	1	2	1	2	2	1	2	2	1	1	1	1	1	1	1	1	1	2	1	1	1	2	2	1	1	2	1	2	1	1	1	1	1	2	1	2	1	2	1	0,689	0,027	0,283
	2	2	2	2	1	2	2	2	1	2	1	2	1	2	2	1	2	2	2	2	1	1	2	2	1	1	2	2	2	2	2	2	2	2	2	2	2	2	1	2	2	1	1	2	1	2	1	2	1			
MRJ_122	2	2	2	1	1	2	2	1	1	2	1	1	1	2	1	1	2	2	1	1	1	2	1	1	1	1	2	2	2	1	1	2	1	1	1	2	1	2	1	1	1	1	1	1	2	1	1	2	0,636	0,066	0,298	
	2	2	2	1	2	2	2	1	2	2	2	2	2	2	2	2	1	2	2	2	1	1	2	2	1	1	2	2	2	2	2	2	2	2	2	2	2	1	2	2	2	2	1	1	2	2	2	2	2			
MRJ_123	2	2	1	1	1	1	1	1	1	2	2	2	2	1	1	1	1	2	1	1	1	1	2	1	1	1	1	2	2	1	2	2	1	2	1	2	1	2	1	2	2	1	1	1	1	1	1	1	1	0,598	0,220	0,182
	2	2	2	2	1	2	2	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	1	2	1	2	2				
MRJ_126	1	1	2	2	1	2	1	2	1	1	2	2	2	1	2	2	2	2	2	2	1	1	1	1	1	1	2	2	2	2	2	2	2	1	1	1	1	1	1	2	2	1	1	1	1	2	1	2	0,150	0,551	0,300	
	2	2	2	2	2	2	1	2	1	1	2	2	2	1	2	2	2	2	2	2	2	1	2	1	2	2	2	2	2	2	2	1	2	2	1	2	1	2	1	2	2	2	2	2	2	1	2	2	2			
MRJ_129	2	1	2	1	1	2	1	1	1	2	1	1	2	1	2	1	2	1	2	2	1	1	1	1	1	2	2	1	1	2	1	1	1	1	2	1	2	1	2	1	1	1	1	2	2	1	1	1	0,051	0,331	0,618	
	2	1	2	1	1	2	2	2	1	2	2	2	2	1	2	1	2	1	2	2	1	1	2	2	2	2	2	2	1	2	2	1	2	1	2	1	2	1	2	1	2	1	2	2	2	2	2	2	2	2		
MRJ_130	2	1	2	1	1	1	2	1	1	2	2	2	2	1	2	2	1	1	2	1	1	1	1	2	1	2	2	2	2	2	1	1	2	1	1	1	2	1	2	1	1	1	1	1	1	1	1	1	1	0,196	0,138	0,666
	2	1	2	1	1	2	2	1	1	2	2	2	2	1	2	2	2	2	2	2	1	1	2	2	1	2	2	2	2	2	2	1	2	2	1	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2		
MRJ_131	2	1	1	1	1	2	2	1	1	2	2	2	2	1	2	1	1	1	2	1	1	1	1	2	2	2	2	1	1	2	1	2	1	1	1	2	1	2	1	1	1	2	2	3	1	2	0,036	0,552	0,412			
	2	1	2	2	1	2	2	2	1	2	2	2	2	2	2	2	2	1	2	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	2	2	1	1	2	2	2	2	2	2	2	2			

APÊNDICE C - Genótipos dos 46 AIM-Indels de indivíduos da Ilha de Marajó e as proporções de ancestralidade continental (continua)

Amostra	1470	777	196	881	3122	548	659	2011	2929	593	798	1193	1871	17	2538	1644	3854	2275	94	3072	772	2313	397	1636	51	2431	2264	2256	128	15	2241	419	943	159	2005	250	1802	1607	1734	406	1386	1726	3626	360	1603	2719	EUR	AFR	NAM				
MRJ_203	1	1	2	?	1	1	1	1	1	1	2	2	1	1	2	1	2	1	2	1	1	1	1	1	2	2	2	1	1	2	1	1	1	2	1	2	1	2	1	1	1	1	1	2	1	1	1	0,086	0,345	0,568			
	2	2	2	?	2	1	2	2	1	2	2	2	2	1	2	1	2	2	2	2	1	2	2	2	1	2	2	2	2	2	2	2	2	1	2	2	2	2	2	2	2	2	2	2	2	1	2	2	0,040	0,304	0,656		
MRJ_204	2	1	2	1	1	1	1	1	1	1	2	2	1	1	2	1	2	1	1	1	1	1	1	1	1	1	2	2	1	2	1	2	1	2	1	1	1	1	2	2	1	1	1	2	2	1	2	1	0,373	0,409	0,218		
	2	2	2	1	1	1	1	1	1	1	1	2	2	1	1	2	1	1	2	2	1	1	1	1	1	2	1	1	1	2	1	2	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0,083	0,249	0,668
MRJ_205	2	1	1	1	1	1	1	1	1	2	2	2	1	1	2	1	2	2	1	1	1	1	1	1	1	1	2	2	1	2	1	2	1	2	1	1	1	1	2	2	1	1	1	1	2	2	1	2	1	0,357	0,018	0,625	
	2	2	2	1	1	2	2	2	2	2	1	2	2	2	2	2	2	2	2	2	1	2	2	2	1	1	2	3	2	1	2	2	1	2	1	2	2	2	2	2	2	2	2	2	2	2	1	1	2	0,560	0,331	0,108	
MRJ_213	1	1	1	1	1	2	1	1	2	1	2	1	1	1	2	1	2	2	2	2	1	1	1	1	1	2	2	2	2	1	1	2	1	1	1	2	1	1	1	1	2	1	1	1	2	1	1	1	1	0,452	0,090	0,458	
	2	2	1	2	1	2	2	1	1	2	2	2	2	1	1	1	1	2	2	1	1	2	2	1	1	1	2	2	1	2	2	1	2	1	2	1	1	1	1	1	2	2	2	2	2	2	2	1	1	1	0,378	0,035	0,586
MRJ_217	2	1	1	1	1	1	1	1	1	2	2	2	1	1	1	1	2	2	1	1	1	1	1	1	1	1	2	2	1	2	1	2	1	1	1	1	1	1	1	1	2	2	1	1	1	1	1	1	1	0,552	0,035	0,413	
	2	2	2	?	1	2	2	2	1	2	1	2	2	2	2	2	2	2	2	2	1	1	1	2	1	1	2	1	2	2	1	2	1	1	1	1	2	2	2	2	2	2	2	2	2	2	1	2	2	0,355	0,142	0,502	
MRJ_219	2	1	2	1	1	2	2	1	1	2	1	2	2	1	1	1	2	2	1	1	1	1	1	1	1	1	2	2	1	2	1	2	1	1	1	1	1	1	1	2	2	1	1	1	2	2	1	1	1	0,208	0,227	0,565	
	2	2	2	1	1	2	2	1	1	2	2	2	2	1	2	2	2	2	2	2	1	2	2	1	1	1	1	2	2	2	2	2	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0,533	0,332	0,135
MRJ_219	2	1	2	1	1	2	2	1	1	2	1	2	2	1	1	1	2	2	1	1	1	1	1	1	1	1	2	2	1	2	1	2	1	1	1	1	1	1	1	1	2	2	1	1	1	1	1	1	1	0,533	0,332	0,135	
	2	2	2	1	1	2	2	1	1	2	2	2	2	1	1	2	2	2	2	2	1	2	2	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0,533	0,332
MRJ_220	2	1	?	1	1	2	2	1	1	2	1	2	1	1	1	1	2	2	1	1	1	1	1	1	1	1	2	2	1	2	1	1	1	1	1	1	1	1	1	1	2	2	1	1	1	1	1	1	1	0,533	0,332	0,135	
	2	?	2	1	1	2	2	1	1	2	2	2	2	1	2	1	2	2	2	2	1	2	2	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0,533	0,332	0,135
MRJ_222	1	2	2	1	1	1	2	1	1	2	1	1	1	1	2	1	2	2	2	1	1	1	2	1	1	1	2	1	2	2	1	1	1	1	1	1	1	1	2	2	2	1	2	1	1	1	1	1	1	0,533	0,332	0,135	
	2	2	2	?	1	2	2	2	1	2	1	2	2	2	2	2	2	2	2	2	1	1	1	2	1	1	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0,533	0,332	0,135
MRJ_225	2	1	1	1	1	1	2	1	1	1	1	2	1	1	2	2	2	2	2	2	1	1	1	1	1	1	2	1	2	2	1	2	1	1	1	1	1	2	1	2	2	1	1	1	1	1	1	1	1	1	0,533	0,332	0,135
	2	2	2	1	1	2	2	1	1	2	1	2	2	1	2	2	2	2	2	2	1	2	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0,533	0,332
MRJ_226	2	1	1	1	1	1	2	1	1	2	1	1	1	1	2	1	1	2	2	2	1	1	1	1	1	1	1	2	1	2	1	2	1	1	1	1	1	1	1	1	2	2	1	2	1	1	1	1	1	1	0,533	0,332	0,135
	2	1	1	1	1	1	2	1	2	2	2	2	2	1	2	1	2	2	2	2	1	2	2	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0,533	0,332
MRJ_230	2	2	1	?	1	2	2	1	1	1	2	2	2	2	2	1	1	1	1	1	2	1	2	1	1	1	2	1	1	1	1	2	1	1	1	2	1	1	1	1	1	2	2	1	1	1	1	1	1	0,533	0,332	0,135	
	2	2	1	?	1	2	2	1	1	2	2	2	2	2	2	1	1	2	2	2	2	2	2	2	2	2	2	3	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0,533	0,332	0,135

APÊNDICE C - Genótipos dos 46 AIM-Indels de indivíduos da Ilha de Marajó e as proporções de ancestralidade continental (continua)

Amostra	1470	777	196	881	3122	548	659	2011	2929	593	798	1193	1871	17	2538	1644	3854	2275	94	3072	772	2313	397	1636	51	2431	2264	2256	128	15	2241	419	943	159	2005	250	1802	1607	1734	406	1386	1726	3626	360	1603	2719	EUR	AFR	NAM			
MRJ_251	1	1	2	2	2	1	2	1	1	2	2	2	2	1	1	1	2	2	1	1	1	1	1	1	2	2	1	1	2	1	1	1	1	1	1	1	2	2	1	1	1	1	1	2	1	1	0,416	0,355	0,229			
	2	1	2	2	2	2	2	2	1	2	2	2	2	1	1	2	1	2	1	2	1	1	1	1	2	2	2	2	1	2	2	2	2	2	1	1	2	2	2	1	2	2	2	2	1	1	2	1	1	0,416	0,355	0,229
MRJ_252	2	2	2	1	1	1	2	1	1	2	2	2	1	1	1	1	2	1	1	1	1	1	1	1	1	1	2	1	1	2	2	2	2	1	2	1	1	1	2	2	1	1	1	1	1	1	1	1	1	0,714	0,116	0,169
	2	2	2	2	1	1	2	2	2	2	2	2	1	1	1	1	2	2	2	2	2	1	2	1	2	1	2	2	2	2	2	2	2	2	1	2	1	2	1	2	2	1	1	2	2	1	1	2	2	0,714	0,116	0,169
MRJ_253	1	2	2	2	2	1	2	1	1	2	2	2	1	1	2	1	2	2	1	1	1	1	1	1	1	1	2	2	1	2	1	2	1	1	1	2	1	2	1	1	1	1	1	1	1	2	1	1	0,259	0,261	0,480	
	2	2	2	2	2	2	2	2	1	2	2	2	2	1	1	2	2	2	2	1	1	1	1	1	1	2	2	2	2	2	2	2	2	1	2	2	2	1	2	1	1	1	2	2	2	1	2	2	0,259	0,261	0,480	
MRJ_254	2	1	2	1	1	1	1	1	1	1	2	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	2	2	1	2	1	2	1	1	1	2	1	2	2	1	1	1	1	2	2	1	2	0,082	0,300	0,618	
	2	1	2	1	2	2	2	1	1	1	2	2	2	1	1	2	2	2	2	2	2	2	2	1	1	2	2	2	2	2	2	2	2	2	1	1	1	2	2	2	2	1	2	2	2	2	1	2	0,082	0,300	0,618	
MRJ_255	2	1	1	1	1	1	2	1	1	2	1	2	2	1	2	1	1	1	2	1	1	1	1	1	1	1	1	2	2	2	2	2	1	2	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	0,316	0,353	0,331
	2	2	2	2	1	2	2	2	2	2	1	2	2	1	2	2	2	1	2	2	1	2	2	2	2	2	2	2	2	2	2	2	1	2	2	2	1	2	1	2	1	1	2	2	2	2	2	2	2	0,316	0,353	0,331
MRJ_256	1	1	1	1	1	1	1	1	1	2	2	2	2	1	2	1	2	2	1	1	1	1	1	1	1	2	1	1	1	1	2	1	2	1	1	1	1	2	2	2	1	1	2	1	1	1	2	1	1	0,273	0,194	0,533
	2	1	2	2	2	2	2	2	2	2	2	2	2	1	2	2	2	2	2	1	1	1	1	1	1	2	1	2	2	2	2	1	2	2	2	1	2	2	2	2	2	1	2	2	2	2	2	2	2	0,273	0,194	0,533
MRJ_258	1	1	2	1	2	1	2	1	1	2	2	2	2	1	1	1	2	2	1	1	1	1	1	1	1	1	1	1	1	1	2	1	2	1	1	1	2	1	2	1	1	1	1	1	1	1	2	1	1	0,323	0,225	0,451
	2	2	2	2	2	2	2	2	1	2	2	2	2	1	2	1	2	2	1	2	1	1	1	1	1	1	2	2	2	2	2	2	2	2	1	2	2	2	2	2	2	1	1	2	2	2	1	2	1	0,323	0,225	0,451
MRJ_261	2	1	2	1	1	1	2	1	1	2	2	2	2	1	1	2	2	1	1	1	1	1	1	1	1	1	2	2	1	2	1	2	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	0,266	0,082	0,652
MRJ_261	2	2	2	2	1	2	2	2	1	2	2	2	2	2	2	2	2	2	2	1	2	1	2	1	1	2	2	2	2	1	2	2	2	1	1	1	1	2	2	2	1	1	2	1	2	2	1	2	0,266	0,082	0,652	
MRJ_264	2	2	1	1	1	2	2	1	1	1	1	1	2	2	1	1	1	2	1	1	1	2	1	1	1	1	1	2	2	2	1	1	2	1	1	1	2	1	2	1	1	1	2	1	1	1	1	1	1	0,650	0,294	0,057
	2	2	2	1	1	2	2	2	1	2	2	2	2	2	2	2	1	2	2	2	2	1	2	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	2	1	2	2	1	2	0,650	0,294	0,057	
MRJ_266	1	1	1	1	1	1	2	1	1	1	1	1	2	1	2	2	2	1	2	1	1	1	1	1	1	1	2	1	2	1	2	1	2	2	2	2	2	2	2	1	2	2	2	1	1	2	1	1	1	0,181	0,371	0,448
	2	2	2	1	1	2	2	2	1	2	2	2	2	2	2	2	2	2	2	2	2	1	1	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0,181	0,371	0,448
MRJ_268	1	1	2	1	1	2	2	2	1	2	2	2	1	1	2	1	2	1	2	1	1	1	1	2	1	1	2	1	1	1	2	2	1	1	2	2	2	1	2	2	1	2	2	1	1	1	1	1	1	0,637	0,057	0,305
	1	1	2	1	1	2	2	2	1	2	2	2	2	2	2	2	1	2	1	2	1	1	1	2	1	1	2	2	1	1	1	2	2	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0,637	0,057	0,305
MRJ_271	1	1	2	1	1	1	2	1	1	1	1	2	1	1	2	1	2	2	2	1	1	1	1	1	1	1	2	2	1	1	1	1	1	1	1	1	1	2	2	1	1	2	2	1	1	1	1	1	1	0,422	0,164	0,415
	1	1	2	1	2	1	2	1	1	2	1	2	2	2	2	2	2	2	2	2	1	2	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0,422	0,164	0,415

APÊNDICE C - Genótipos dos 46 AIM-Indels de indivíduos da Ilha de Marajó e as proporções de ancestralidade continental (continua)

Amostra	1470	777	196	881	3122	548	659	2011	2929	593	798	1193	1871	17	2538	1644	3854	2275	94	3072	772	2313	397	1636	51	2431	2264	2256	128	15	2241	419	943	159	2005	250	1802	1607	1734	406	1386	1726	3626	360	1603	2719	EUR	AFR	NAM																	
MRJ_275	1	1	1	1	1	1	2	1	1	2	1	2	2	1	1	1	2	2	1	1	1	1	1	1	1	1	1	3	1	1	1	1	1	1	1	2	1	1	1	2	1	2	1	1	1	1	2	0,900	0,015	0,085																
	2	2	2	1	1	2	2	2	1	2	2	2	2	2	1	1	2	2	2	2	1	1	2	2	1	2	2	2	2	2	1	2	2	1	2	1	2	2	2	2	2	1	2	2	1	2	1	2	2	0,900	0,015	0,085														
MRJ_278	1	1	1	1	1	1	2	1	1	2	1	2	2	2	1	1	2	2	1	1	1	1	1	1	1	2	2	1	2	2	2	1	1	1	1	2	1	2	2	2	2	1	1	1	1	1	1	1	1	1	0,775	0,056	0,169													
	2	2	1	1	1	2	2	2	2	2	1	2	2	2	2	2	2	2	2	2	1	1	2	1	1	2	2	2	2	2	2	2	2	2	1	2	1	2	2	2	2	2	1	2	2	2	2	2	2	2	0,775	0,056	0,169													
MRJ_283	1	1	1	1	1	2	1	1	1	2	1	2	2	1	2	2	1	1	1	1	1	2	1	1	1	2	1	2	1	2	1	2	1	1	1	1	1	2	1	1	2	1	1	2	1	1	1	1	1	1	1	0,098	0,416	0,486												
	2	1	1	2	1	2	2	2	1	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	2	1	1	2	1	2	1	1	2	2	2	2	1	2	2	2	2	0,098	0,416	0,486												
MRJ_285	1	1	1	1	1	1	2	1	1	2	1	2	2	2	1	1	2	1	1	1	1	1	1	1	1	1	1	1	2	1	2	1	1	1	2	1	1	1	2	1	1	2	2	2	2	1	1	?	?	?	?	?	?	?	?	0,558	0,247	0,195								
	1	1	2	?	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0,558	0,247	0,195										
MRJ_288	2	1	2	1	1	1	2	1	1	2	1	1	1	1	2	1	2	1	1	1	1	1	1	1	1	1	1	2	1	1	2	1	1	1	1	1	1	1	1	2	2	1	1	1	1	2	1	1	1	1	1	1	1	0,216	0,036	0,748										
	2	2	2	1	1	2	2	2	1	2	2	2	2	2	1	2	1	2	2	2	2	1	1	2	1	1	1	2	2	1	2	1	2	1	2	1	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0,216	0,036	0,748								
MRJ_290	2	1	2	1	1	1	2	1	1	2	2	1	2	1	2	1	2	1	1	1	1	1	1	2	1	2	2	1	1	1	2	1	1	1	1	1	1	1	1	2	2	2	2	1	1	1	1	1	1	1	1	1	1	1	1	0,246	0,173	0,581								
	2	2	2	2	1	2	2	1	2	2	2	2	2	1	2	2	2	1	1	1	1	1	1	2	1	2	2	2	2	1	2	2	2	1	1	1	1	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0,246	0,173	0,581						
MRJ_291	1	1	2	1	1	1	2	1	1	2	2	2	2	1	2	1	2	1	1	1	1	1	1	1	1	1	1	2	2	1	2	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0,302	0,054	0,644								
	2	2	2	2	1	2	2	2	1	2	2	2	2	2	2	2	1	2	2	2	2	1	1	2	1	1	1	1	2	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0,302	0,054	0,644							
MRJ_293	2	1	2	1	1	1	1	2	1	1	2	1	2	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0,302	0,279	0,419							
	2	2	2	1	1	2	2	2	1	2	2	2	2	1	2	2	2	2	2	2	2	1	1	2	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0,302	0,279	0,419					
MRJ_295	1	1	?	1	1	1	1	1	1	2	2	1	2	1	2	2	2	2	1	1	1	1	1	1	1	1	1	1	2	1	2	2	2	1	1	1	2	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0,085	0,238	0,677					
	2	2	2	2	1	1	2	2	1	2	2	2	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0,085	0,238	0,677				
MRJ_299	1	1	2	1	1	2	1	1	1	1	2	2	1	1	2	2	2	1	2	2	1	1	1	1	1	2	1	2	1	2	1	2	1	1	1	1	2	1	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0,060	0,302	0,637					
	2	2	2	2	1	2	2	2	1	2	2	2	1	1	2	2	2	2	2	2	2	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0,060	0,302	0,637				
MRJ_300	1	1	1	1	1	2	1	1	1	1	1	2	2	1	2	1	2	2	1	1	1	1	2	1	1	1	1	2	2	1	1	1	1	1	1	1	2	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0,381	0,347	0,272				
	2	2	2	1	2	2	2	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0,381	0,347	0,272			
MRJ_302	1	2	2	1	1	2	2	1	1	2	1	2	1	1	2	2	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	2	1	2	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0,377	0,131	0,491			
	2	2	2	1	1	2	2	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0,377	0,131

APÊNDICE D - Pôster exibido no Congresso *Haploid Markers Conference* - Polônia, maio/2018



Genetic characterization of Y and mtDNA lineages in the admixed population of Marajó Island, Northern of Brazil.

S. Loiola¹, F. Simão¹, L.M.L. Pinheiro², L. Gusmão¹, E.F. Carvalho¹

¹DNA Diagnostic Laboratory (LDD), State University of Rio de Janeiro (UERJ), Brazil; ²Department of Natural Sciences, Federal University of Para (UFPA), Brazil
slvia.labdna@gmail.com



Introduction



Figure 1: Location of Marajó Island

The Island of Marajó is located in the north of Pará state, Brazil. It is the largest fluvial archipelago in the world, bathed by the Atlantic Ocean and the Amazon and Tocantins rivers (Fig 1). At the time of Portuguese colonization, around 30 different indigenous groups inhabited the Island. However, by the end of 18th century, all indigenous populations virtually disappeared from the region. Currently, the island is inhabited by an admixed population, with Native American, European and African genetic contributions. The lineage markers located in Y chromosome and mitochondrial DNA (mtDNA), with exclusive paternal and maternal inheritances, respectively, are useful to reconstruct populations' history and for ancestry inferences. The aim of this work was to contribute to the knowledge about the current genetic ancestry of Marajó Island population through the study Y chromosome and mtDNA markers. We intend to better understand the maternal and paternal contributions from Native Americans, Europeans and Africans, in the formation of the current admixed population from Marajó.

Results and Discussion

Y-STRs

For the 27 Yfiler Plus markers, a haplotype diversity of 0.9969 ± 0.0032 was observed, with 58 unique haplotypes, 4 shared by two individuals and 2 shared by three individuals. This value is low when compared with the one obtained for the population of Rio de Janeiro (0.9997) [1].

Table 1: Haplotype Diversity for 18 Y-STRs

Population	Diversity
Marajó	0.9969±0.0032
Belém	1.0000±0.0002
Boa Vista	1.0000±0.0026
Macapá	1.0000±0.0027
Manaus	1.0000±0.0022
Palmas	1.0000±0.0086
Porto Velho	0.9999±0.0009
Rio Branco	1.0000±0.0062
Santarém	0.9988±0.0027
São Gabriel	0.9973±0.0038

Since the Y-STR data available for other populations from the North region of Brazil do not include all markers from the Yfiler Plus kit [2,3], a new analysis was performed for a subset of 18 Y-STRs (Table 1). The population of Marajó has the lowest value of haplotype diversity, in comparison to Belém, Boa Vista, Macapá, Manaus, Palmas, Porto Velho, Rio Branco, Santarém, or even with a sample of Native Americans from São Gabriel da Cachoeira (Amazonia).

Table 2: F_{ST} genetic distances among Marajó and other North populations (below diagonal) and the corresponding non-differentiation p -values (above diagonal).

	Marajó	Belém	Boa Vista	Macapá	Manaus	Palmas	Porto Velho	Rio Branco	Santarém	São Gabriel
Marajó	*	0.0004	0.0013	0.0159	0.0182	0.4330	0.0233	0.2832	0.0697	0.0000
Belém	0.0137	*	0.0002	0.0068	0.0000	0.0361	0.0000	0.0254	0.0001	0.0000
Boa Vista	0.0192	0.0157	*	0.0431	0.0007	0.0097	0.0001	0.0556	0.0001	0.0000
Macapá	0.0122	0.0090	0.0083	*	0.0874	0.0914	0.0185	0.2759	0.0580	0.0000
Manaus	0.0102	0.0185	0.0195	0.0058	*	0.0685	0.0079	0.4667	0.6597	0.0000
Palmas	-0.0001	0.0107	0.0211	0.0098	0.0092	*	0.7772	0.3336	0.4708	0.0000
Porto Velho	0.0076	0.0177	0.0230	0.0090	0.0103	-0.0043	*	0.4734	0.1492	0.0000
Rio Branco	0.0025	0.0117	0.0117	0.0028	-0.0008	0.0019	-0.0008	*	0.5001	0.0000
Santarém	0.0062	0.0155	0.0272	0.0075	-0.0019	-0.0007	0.0028	-0.0013	*	0.0000
São Gabriel	0.0653	0.0763	0.0932	0.0892	0.0660	0.0762	0.0816	0.0748	0.0592	*

Note: significant level is 0.0011 (with Bonferroni's correction), 50175 permutations, $\alpha \leq 0.0024$

Genetic distances between Marajó and other admixed populations from the North were low, except for São Gabriel da Cachoeira, indicating a predominating European male lineages contribution (Table 2).

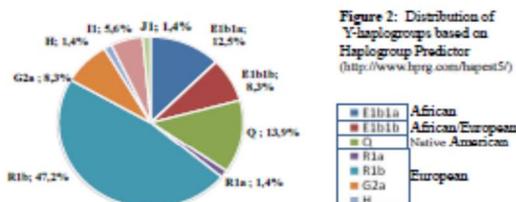


Figure 2: Distribution of Y-haplogroups based on Haplogroup Predictor (<http://www.hpg.com/hapset5/>)

Haplogroup prediction showed a distribution of 13.9% native American haplogroups, 13.8% African and 72.3% European (Fig. 2). This is a preliminary result that needs to be confirmed by the analysis of Y-SNPs.

mtDNA

In the analysis of mtDNA complete control region, a high haplotypic diversity was observed (0.9833 ± 0.0097), although it is lower than that for the metropolitan region of Rio de Janeiro [4]. To increase the number of populations in pairwise comparisons, a new analysis was performed using available data from HVS-I [5], and the F_{ST} genetic distances are represented in Table 3. Significant differences were found between Marajó and samples from the South, Southeast, Northeast and North regions of Brazil.

	Marajó	Southeast	South	Northeast	North
Marajó	*	0.0000	0.0000	0.0007	0.0047
Southeast	0.0400	*	0.0058	0.5717	0.0819
South	0.0898	0.0179	*	0.0127	0.0346
Northeast	0.0371	-0.0017	0.0194	*	0.0904
North	0.0295	0.0080	0.0144	0.0100	*

Table 3: F_{ST} genetic distances between Marajó and other admixed Brazilian populations (below diagonal) and non-differentiation p -values (above diagonal).

Note: significant level is 0.005, 10100 permutations, $\alpha \leq 0.0049$

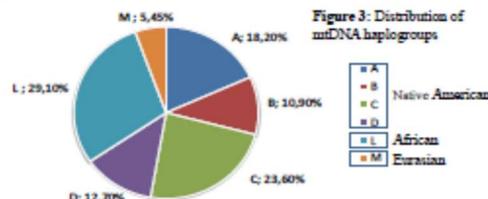


Figure 3: Distribution of mtDNA haplogroups

The mtDNA haplogroup composition revealed a high contribution of Native American maternal ancestry (65%). The African haplogroups represented 29% of the samples and only 5% where from eurasian maternal origin (Fig. 3).

Conclusion

The results of this study are in accordance with historical data, showing that Marajó population results from an admixture processes with a strong maternal ancestry contribution from Native Americans and African slaves, and a paternal ancestry that is predominantly European. A low haplotypic diversity was found for the Y chromosome, but it is relatively high for the mtDNA control region. Significant genetic distances were found between Marajó and all Brazilian populations in terms of maternal inheritance, and for the Y chromosome, significantly high genetic distances were only observed with the native Amazon population of São Gabriel da Cachoeira.

Methodology: Blood samples were collected from 72 unrelated males living in Marajó with local ancestry for at least three generations. The project was approved by the ethical committee of UERJ. DNA was extracted using Chelex resin or by organic extraction. All samples were genotyped for the Yfiler® Plus markers (Thermo Fisher Scientific). In 35 samples, the complete mtDNA control region was amplified and sequenced using the Big Dye Terminator v3.1 cycle sequencing kit (Applied Biosystems). Separation and detection was carried out on an ABI 3500 (Applied Biosystems). Sequences were compared to the revised Cambridge Reference Sequence (rCRS) with the software SeqScape v2.7. Haplogroups were attributed via EMPOP. Statistical parameters were calculated using Arlequin 3.5 software [6].

References: [1] Jannuzzi et al. *Int J Legal Med*. 2018. Epub ahead of print. [2] Palha et al. *PLoSOne*. 2012; 7 (7): e40007; [3] Purps et al. *Forensic Sci Int Genet*. 2014; 12:12-23; [4] Simão et al. *Forensic Sci Int Genet*. 2018; 34: 97-104; [5] Alves-Silva et al. *Am J Hum Genet*. 2000; 67(2):444-61; [6] Excoffier et al. *Evol Bioinform Online*. 2005; 1: 47-50.

Acknowledgements: We thank all sample donors and people who helped in sample collection. This work was supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES and DNA Program - State University and Justice Court of Rio de Janeiro (UERJ/TJRJ) - Brazil.

APÊNDICE E - *Proceedings* do 28º Congresso - Sociedade Internacional de Genética Forense 2019 – Praga, República Tcheca
(<https://doi.org/10.1016/j.fsigss.2019.09.118>)

Forensic Science International: Genetics Supplement Series 7 (2019) 313–314



Contents lists available at ScienceDirect

Forensic Science International: Genetics Supplement Series

journal homepage: www.elsevier.com/locate/fsigss



Genetic characterization and ancestry of the admixed population of Marajó Island, northern of Brazil, with autosomal and lineage markers



S. Martins^{a,*}, F. Simão^a, L.M.L. Pinheiro^b, M. Nguidi^a, L. Deccache^a, L. Gusmão^a, E.F. Carvalho^a

^a DNA Diagnostic Laboratory (LDD), State University of Rio de Janeiro (UERJ), Brazil

^b Department of Natural Sciences, Federal University of Pará (UFPA), Brazil

ARTICLE INFO

Keywords:
Y-STRs
mtDNA
AIMs
Maternal inheritance

ABSTRACT

To investigate the genetic background of the Marajó population, samples from unrelated individuals were selected, with matrilineal or patrilineal local ancestry for at least three generations. A total of 72 male samples were genotyped for 27 Y-STRs, 65 for 46 AIMs and 55 were sequenced for the mtDNA control region. Low haplotype diversities were obtained for Y-STRs and mtDNA, when compared to other Brazilian admixed populations, most probably due to the small size of the population and the geographic isolation of the island. Based on HVI, significant genetic distances were found between Marajó and Brazilian populations from the North, Northeast, South and Southeast regions. For mtDNA, Native American haplogroups were the most prevalent, followed by the African and European haplogroups. For the autosomal markers, a 40.7% European ancestry was found, followed by 33.2% of Native American and 26.1% of African contributions.

1. Introduction

The Island of Marajó is in the north of Pará state, about 90 km from the capital, Belém. It is the largest fluvial-maritime archipelago in the world, bathed by the Atlantic Ocean and the Amazon and Tocantins rivers. Around 30 different indigenous nations inhabited the Island at the time of Portuguese colonization, but by the end of the 18th century these groups had virtually disappeared. Currently, the island is inhabited by an admixed population, with Native American, European and African genetic contributions [1].

The genetic markers are transmitted from both parents after recombination, except for the Y chromosome and mitochondrial DNA (mtDNA), which have exclusively paternal and maternal inheritance, respectively. Although lineage markers are useful in the reconstruction of the evolutionary history and geographical correlations of populations, they do not reveal the overall ancestry of an individual, which can only be accessed with by the study of autosomal Ancestry Informative Markers (AIMs).

The aim of this work was to contribute to the knowledge about the genetic ancestry of the population of Marajó Island, using Y-STRs, mtDNA and AIMs.

2. Material and methods

A total of 111 blood samples were collected in FTA papers, from unrelated individuals from Marajó, under written informed consent from the participants. The project was approved by the Ethical Committee of the State University of Rio de Janeiro (CAAE:0067.0.228.000-09). DNA was extracted using Chelex® resin [2] or a standard phenol-chloroform protocol [3].

A total of 72 males who had paternal grandfather born on the island were selected for Y-STRs genotyping. The 27 Y-STRs from the Yfiler® Plus kit (Thermo Fisher Scientific, Waltham, MA, USA) were genotyped, following manufacturer's recommendations.

The mtDNA control region was amplified in 55 samples from individuals with maternal grandmother born in Marajó. PCR reactions were carried out using the primers L15997-H639 [4]. PCR products were purified with Illustra™ ExoProStar™ (GE Healthcare). The BigDye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems, Foster City, CA) was used for sequencing, with primers L15997, H639, L16555, H016, L16268 and L314 [4]. Sequencing products were cleaned using Sephadex G-50 resin [3]. Sequences were compared to the revised Cambridge Reference Sequence (rCRS) with the software SeqScape v2.7 (Applied Biosystems). Mitochondrial haplotypes were assigned to haplogroups using EMPOP v4/R12 (<http://www.empop.online>) [5] and confirmed in Phylotree (<http://www.phylotree.org>) [6].

* Corresponding author.

E-mail address: silvia.labdna@gmail.com (S. Martins).

<https://doi.org/10.1016/j.fsigss.2019.09.118>

Received 10 September 2019; Accepted 30 September 2019

Available online 01 October 2019

1875-1768/ © 2019 Published by Elsevier B.V.

Table 1
 F_{ST} genetic distances between Marajó and other admixed Brazilian populations (below diagonal) and non-differentiation p-values (above diagonal).

	Marajó	Southeast	South	Northeast	North
Marajó	*	0.00000	0.00000	0.00079	0.00475
Southeast	0.04009	*	0.00584	0.57173	0.08197
South	0.08983	0.01796	*	0.01277	0.03465
Northeast	0.03716	-0.00177	0.01947	*	0.09049
North	0.02959	0.00801	0.01447	0.01001	*

*Significant level is 0.005 (after applying Bonferroni's correction); s.e. \leq 0.0049.

An analysis of 46 AIM-Indels was performed for 65 individuals who had both maternal and paternal grandparents born in Marajó. The full set of AIMs was amplified in a single PCR reaction as described in Pereira et al. [7].

In all cases, PCRs were carried out in a Veriti™ 96-Well Thermal Cycler (Applied Biosystems) and separation and detection were performed on an ABI 3500 Genetic Analyser (Applied Biosystems). Both Y-STRs and AIM-Indels were genotyped using the GeneMapper® Software V. 4.1 (Applied Biosystems).

For Y-STRs and mtDNA, haplotype/haplogroup diversities and genetic distance were calculated using the Arlequin 3.5 software [8]. For the AIM-Indels, estimates of individual and global ancestry were obtained using the STRUCTURE v2.3.3 software [9]. African, European and Amerindian populations from the HGDP-CEPH human diversity panel (<http://www.hagsc.org/hgdp/>) were used as references [7].

3. Results and discussion

For the 27 YfilerPlus markers, a haplotype diversity of 0.9969 ± 0.0032 was observed, with 58 unique haplotypes, 4 shared by two individuals and 2 shared by three individuals. This value is low when compared with the one obtained in Rio de Janeiro (0.99997) [10]. A new analysis was performed for a subset of 18 Y-STRs, in order to include other populations from the North region of Brazil [11,12]. Marajó presented the lowest value of haplotype diversity (0.9961 ± 0.0033), in comparison to Belém, Boa Vista, Macapá, Manaus, Palmas, Porto Velho, Rio Branco and Santarém (HD values \geq 0.9988), or even with the Amazonian Natives from São Gabriel da Cachoeira (0.9973 ± 0.0038). Based on Y-STRs, genetic distances were calculated between Marajó and other 8 admixed populations from Belém, Boa Vista, Macapá, Manaus, Palmas, Porto Velho, Rio Branco and Santarém [10], and one Native American population from São Gabriel da Cachoeira [12], all in the Brazilian North region. Low F_{ST} s were found between Marajó and the other northern populations, except for São Gabriel da Cachoeira ($F_{ST} = 0.06537$; non-differentiation p-value $< 5E-06$), pointing to a predominating European male contribution.

In the analysis of mtDNA control region, a high haplotypic diversity was observed (0.9833 ± 0.0097), although lower than the observed in Rio de Janeiro [13].

Using data available in the literature for HVI [14], genetic distances were calculated between Marajó and 4 regions of Brazil (Table 1). Significant differences were found between Marajó and samples from the South, Southeast, Northeast and North regions. The mtDNA haplogroup composition revealed a high Native American maternal ancestry (65%). The African haplogroups represented 29% of the samples and only 5% where from Eurasian maternal origin.

In the analysis of AIM-Indels, a proportion of 26.1% African, 33.2% Native American and 40.7% European contributions were estimated.

The African ancestry in Marajó is higher than in other populations from the North region, such as Santa Isabel, Manaus and Belém [15]. On the other hand, the European contribution is lower, except when compared to the population of Santa Isabel [15].

4. Conclusions

This study represents the first report on the genetic characterization of the Marajó population. The results are in accordance with historical data, showing that this population results from admixture processes with a strong Native American and African maternal contributions, and a paternal ancestry that is predominantly European.

A low haplotypic diversity was found for the Y chromosome and mtDNA, most probably due to the small size of the population and the geographic isolation of the island.

For the autosomal markers, the African contribution to this population was higher in other populations in the North region, indicating that Marajó Island may have become a refuge for African slaves.

Financial support

This work was supported by FAPERJ (Fundação de Amparo à Pesquisa do Rio de Janeiro), CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) and CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico).

Declaration of Competing Interest

None.

References

- [1] J. Morim, *Ilha de Marajó*, Pesquisa Escolar Online, Fundação Joaquim Nabuco, Recife. < http://basilio.fundaj.gov.br/pesquisaescolar/index.php?option=com_content&view=article&id=10449%3AIlha-de-marajoc&id=44%3AIlha-de-Itemid=1.
- [2] P.S. Walsh, D.A. Metzger, R. Higuchi, Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material, *Biotechniques* 10 (4) (1991) 506–513.
- [3] J. Sambrook, E.F. Fritsch, T. Maniatis, *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harb. Lab. Press, New York, 1989.
- [4] R.S. Dos Reis, F. Simão, V. Dos Santos Strange, et al., A view of the maternal inheritance of Espírito Santo populations: the contrast between the admixed and Pomeranian descent groups, *Forensic Sci. Int. Genet.* 40 (2019) 175–181.
- [5] W. Pámon, A. DÜR, EMPPOP—a forensic mtDNA database, *Forensic Sci. Int. Genet.* 1 (2) (2007) 88–92.
- [6] M. van Oven, M. Kayser, Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation, *Hum. Mutat.* 30 (2009) 386–394.
- [7] R. Pereira, C. Phillips, N. Pinto, et al., Straightforward inference of ancestry and admixture proportions through ancestry-informative insertion deletion multiplexing, *PLoS One* 7 (1) (2012) e29684.
- [8] L. Excoffier, H.E. Lischer, Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows, *Mol. Ecol. Resour.* 10 (2010) 564–567.
- [9] J.K. Pritchard, M. Stephens, P. Donnelly, et al., Inference of population structure using multilocus genotype data, *Genetics* 155 (2) (2000) 945–959.
- [10] J. Jannuzzi, P. Domingues, F. Simão, et al., Genetic characterization of Rio de Janeiro for different Y-STR sets, *Int. J. Legal Med.* 132 (5) (2018) 1313–1315.
- [11] T. Palha, L. Guzmán, E. Rubelso-Rodrigues, et al., Disclosing the genetic structure of Brazil through analysis of male lineages with highly discriminating haplotypes, *PLoS One* 7 (2012) e40007.
- [12] J. Purps, S. Siegen, S. Sabine, et al., A global analysis of Y-chromosomal haplotype diversity for 23 STR loci, *Forensic Sci. Int. Genet.* 12 (2014) 12–23.
- [13] F. Simão, A.P. Ferreira, E.F. de Carvalho, et al., Defining mtDNA origins and population stratification in Rio de Janeiro, *Forensic Sci. Int. Genet.* 34 (2018) 97–104.
- [14] J. Alves-Silva, M. da Silva Santos, P.F. Guimarães, et al., The ancestry of Brazilian mtDNA lineages, *Am. J. Hum. Genet.* 67 (2) (2000) 444–461.
- [15] F. Saloum de Neves Manta, R. Pereira, R. Vianna, et al., Revisiting the genetic ancestry of Brazilians using autosomal AIM-Indels, *PLoS One* 8 (9) (2013) e75145.

APÊNDICE F - Artigo publicado na revista *Forensic Science International: Genetics*
(<https://doi.org/10.1016/j.fsigen.2019.102163>)

Forensic Science International: Genetics 44 (2020) 102163



Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsigen



Male lineages in Brazilian populations and performance of haplogroup prediction tools



Juliana Jannuzzi^{a,*}, Julyana Ribeiro^a, Clarice Alho^b, Grasielly de Oliveira Lázaro e Araújo^c, Regina Cicarelli^d, Heitor Simões Dutra Corrêa^e, Suelen Ferreira^f, Cíntia Fridman^g, Verónica Gomes^{h,i}, Sílvia Loiola^a, Mariana Flavia da Mota^c, Ândrea Ribeiro-dos-Santos^j, Carlos Antonio de Souza^k, Rossana Santiago de Sousa Azuly^l, Elizeu F. Carvalho^a, Leonor Gusmão^a

^a DNA Diagnostic Laboratory (LDD), State University of Rio de Janeiro (UERJ), Rio de Janeiro, Brazil

^b Laboratory of Human and Molecular Genetics, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, Brazil

^c Laboratório de Biologia e DNA Forense, Superintendência de Polícia Técnico-Científica do Estado de Goiás, Goiás, Brazil

^d UNESP-Universidade Estadual Paulista, Faculdade de Ciências Farmacêuticas, Laboratório de Investigação de Parentalidade-NAC, São Paulo, Brazil

^e Coordenadoria de Perícias em Biologia Molecular, POLITEC – Perícia Oficial e Identificação Técnica, Mato Grosso, Brazil

^f Faculdade Pitágoras, São Luís, Brazil

^g Departamento de Medicina Legal, Ética Médica e Medicina Social e do Trabalho da Faculdade de Medicina da USP, Universidade de São Paulo (USP), São Paulo, Brazil

^h IPATIMUP-Institute of Pathology and Molecular Immunology from the University of Porto, Portugal

ⁱ IIS-Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Porto, Portugal

^j Postgraduate Program in Genetics and Molecular Biology, Laboratory of Human and Medical Genetics, Federal University of Pará, Belém, Brazil

^k Secretaria de Defesa Social Pernambuco, Pernambuco, Brazil

^l Serviço de Endocrinologia e Metabolismo do Hospital Universitário da Universidade Federal do Maranhão, Maranhão, Brazil

ARTICLE INFO

Keywords:

Y-STRs
Yfiler Plus kit
Haplogroup prediction
Admixed populations

ABSTRACT

The use of Y-chromosomal genetic markers in forensic investigations demands the establishment of reliable and representative DNA databases of different reference populations. The genetic characterization of the Y chromosome variation in human populations requires the analyses of haplotype frequencies allied to haplogroup determination. The present study aimed to contribute to the Brazilian database by providing 1,382 Yfiler Plus individual profiles, from 11 Brazilian states. The Yfiler Plus markers showed high haplotype diversities in all Brazilian populations (> 0.9970), allowing high intra-population discrimination in forensic investigations. Pairwise genetic distances showed a homogeneity between Brazilian populations ($F_{ST} < 0.0043$; non-differentiation p-values ≥ 0.0212), indicating that admixed populations from Brazil can be represented in a single Yfiler Plus haplotype database, for forensic purposes. The performance of Haplogroup Predictor and NevGen software in haplogroup prediction based on Yfiler Plus and Yfiler haplotypes was evaluated in a subset of 416 Brazilian samples that were also genotyped for 51 Y-SNPs. In 25% of the samples, no classification or errors were found for at least one of the prediction tools or marker sets. NevGen presented lower error rates (5.52% and 8.65% with Yfiler Plus and Yfiler, respectively) than Haplogroup Predictor (16.11% with Yfiler Plus and 13.70% with Yfiler). In conclusion, both haplogroup prediction tools can be useful to direct the SNP typing, but present large error rates to be used in forensic analysis, especially in predicting African haplogroups in admixed South American populations.

1. Introduction

Genetic markers of Y chromosome are informative and useful in characterizing male lineages, since a large part of its extension, named as non-recombining portion of the Y chromosome (NRY), does not

recombine with X chromosome. Thus, unless mutations occur, NRY is inherited and transmitted intact from generation to generation as a haplotype. Genetic polymorphisms of the NRY are widely used in forensic genetics, mainly to complement the investigation in cases in which the analysis of autosomal DNA is not informative, as for example:

* Corresponding author.

E-mail address: jannuzzi.juliana@gmail.com (J. Jannuzzi).

<https://doi.org/10.1016/j.fsigen.2019.102163>

Received 13 June 2019; Received in revised form 25 August 2019; Accepted 19 September 2019

Available online 15 October 2019

1872-4973/ © 2019 Elsevier B.V. All rights reserved.

(a) in sexual assault cases, to identify the male genetic profile in mixtures of male-female biological traces; (b) in familial searching, when there is no connection between the trace and the suspects; and (c) to infer paternal biogeographic origin of unknown trace donors, helpful in guiding the investigation when no suspects are available. These polymorphisms can also be useful in paternal kinship investigations, in the study of human migrations and in genealogical researches [1–4].

Short Tandem Repeat markers of this chromosome (Y-STRs) display high mutation rates, in the order of 10^{-3} , exhibiting high variability and power of discrimination. Therefore, Y-STR defined haplotypes allow for a high intrapopulation discrimination [3,5–7]. As a result, Y-STRs are widely used in forensic and paternity investigations and in short time evolutionary studies or in micro geographic coverage studies [1–3,8,9]. Single Nucleotide Polymorphisms of the Y chromosome (Y-SNPs) exhibit low mutation rates (about 10^{-9}) and characterize unique molecular events in human evolutionary history. Therefore, these markers define haplogroups with high geographical specificity and well-defined origins, being widely used in ancestry studies and to detect substructure in admixed populations [10,11].

The genetic characterization of human populations requires the analyses of haplotype frequencies allied to haplogroup determination. As mutation rates of SNPs and STRs are very different, a haplogroup can comprise many Y-STR haplotypes [12]. Despite this, Y-STR haplotypes can be used to infer haplogroups, to assist SNP selection in haplogroup classification. There are software available to determine the most likely haplogroups based on Y-STR profiles (e.g., Vadim Urasin's YPredictor - <http://predictor.ydna.ru/>; The haplogroup classifier of Arizona University - Schlecht et al., 2008; Jim Cullen's World Haplogroup & Haplogroup-I Subclade Predictor - <http://members.bex.net/jtcullen515/haplotest.htm>).

A free software widely used is the Haplogroup Predictor (<http://www.hprg.com/hapest5/>) that uses Bayesian allele-frequency approaches to estimate the probability of a given haplotype belonging to a certain haplogroup [13,14]. This software considers the observed allele frequencies for each marker in the haplotype, for several haplogroups, which are deposited in public databases [12,13,15]. The adopted Bayesian approach requires the indication of the possible geographic origin of the haplotypes analyzed, since their frequencies may vary according to the region considered. Currently, all versions of the software, except the 27-haplogroup program version, offer three options of probable regions of origin (Northeast European, Eastern European and Mediterranean) plus Equal Priorities, which considers all previous ones in the same proportion and can also be used for the remaining (not-considered) regions, such as those from the American continent [13,14]. The 27-haplogroup program version offers the same four options plus The South Asia.

NevGen Y DNA Predictor [16] is another free software available online and in desktop version, which used as inspiration the Athey's Haplogroup Predictor software. NevGen also uses the Bayesian allele-frequency approach, but, differently from Haplogroup Predictor, it correlates the values of different pairs of STRs in the probability calculations of subclades when needed (<http://www.nevgen.org/AboutNevGen.html>).

The haplotype data used for prediction by NevGen comes from public Family Tree DNA (FTDNA) projects and only deep SNP-tested haplotypes are considered for predictor's statistics. The statistical calculations used by this program do not need to distribute 100% of probability to supported subclades, designating some of them as "unsupported subclades". Another characteristic of NevGen that differs from the Haplogroup Predictor is that in NevGen it is not necessary to indicate the probable origin region of the entered haplotype.

The use of Y-chromosomal genetic markers in forensic investigations demands the establishment of reliable and representative DNA databases of different reference populations. These databases are highly relevant to correctly provide haplotype frequency information for forensic and kinship analysis. Moreover, in order to avoid errors in

probabilistic calculations, it is also important to investigate population substructure, which is widely present in American countries.

The European colonization process in Brazil initiated in 1500, when the Portuguese arrived, and the admixture between European and Native-American began after that [17,18]. During centuries, the arrival of African slaves and immigrants from different places in Europe, contributed to intensify the admixture of Brazilian population, characterized by an important genetic contribution from three main groups: European, African and Native-American populations [11,19,20]. Currently, admixture patterns in Brazilian populations vary throughout the five geopolitical regions of the country, namely South, Southeast, North, Northeast and Central-West, although all regions being characterized by a high European paternal ancestry [21,22].

Studies on Y-chromosomal STR markers in Brazilian populations have been performed during the past years, in order to analyze the genetic diversity of male lineages throughout the territory. These studies comprised different sets of Y-STRs, most of them present in forensic kits such as the Yfiler (AmpFISTR® Yfiler™, Applied Biosystems™) and PPY23 (PowerPlex® Y23, Promega) kits [8,20,23–26]. There is only one study carried out with Yfiler Plus (Yfiler™ Plus kit, Thermo Fisher Scientific), which comprises 26 Y-STRs, focusing in the Rio de Janeiro population [27].

Since the Yfiler Plus kit proved to be highly discriminative and useful in forensic genetics field, and no data is available for most Brazilian populations, the aim of this study was to contribute to the Brazilian database of Yfiler Plus haplotypes from 11 Brazilian states. Moreover, we intended to evaluate the performance of Haplogroup Predictor and NevGen software in haplogroup predictions based on Yfiler Plus and Yfiler haplotypes.

2. Materials and methods

2.1. Samples collection and DNA extraction

A total of 1382 samples from unrelated male individuals were collected from several Brazilian states (see Fig. 1), under written informed consent. These samples were selected from routine paternity cases and from volunteers, from public university and forensic laboratories situated in urban centers. Both the project and the informed consent were approved by the Ethics Committee of the HUPE/UERJ (Hospital Universitário Pedro Ernesto/Universidade do Estado do Rio de Janeiro). Genomic DNA was extracted with standard Chelex and Phenol-Chloroform methodologies.

2.2. Genetic markers and genotyping

Samples were genotyped using the Yfiler™ Plus kit (Thermo Fisher Scientific, Inc., Waltham, MA, USA). The 26 Y-STR markers present in this kit were amplified in a single multiplex PCR followed by capillary electrophoresis, according to the user guide. The Yfiler Plus haplotypes described in this study were submitted to the YHRD database (see accession numbers in Fig. 1).

2.3. Evaluation of haplogroup prediction software

To evaluate the accuracy of haplogroups prediction tools, 416 random samples from Maranhão, São Paulo, Rio de Janeiro and Espírito Santo were selected and genotyped for Y-SNPs. The samples from Rio de Janeiro were a subgroup of those included in a previous study [27]. Concerning the samples from Espírito Santo, data were already available for Y-SNPs [19].

In this study, 51 Y-SNPs were selected, based on the Y chromosome phylogeny [25], to determine haplogroups usually found in South American admixed populations, namely European, sub-Saharan African and Native American haplogroups (Supplementary Fig. S1 shows a phylogenetic tree including the selected markers). The Y-SNPs were

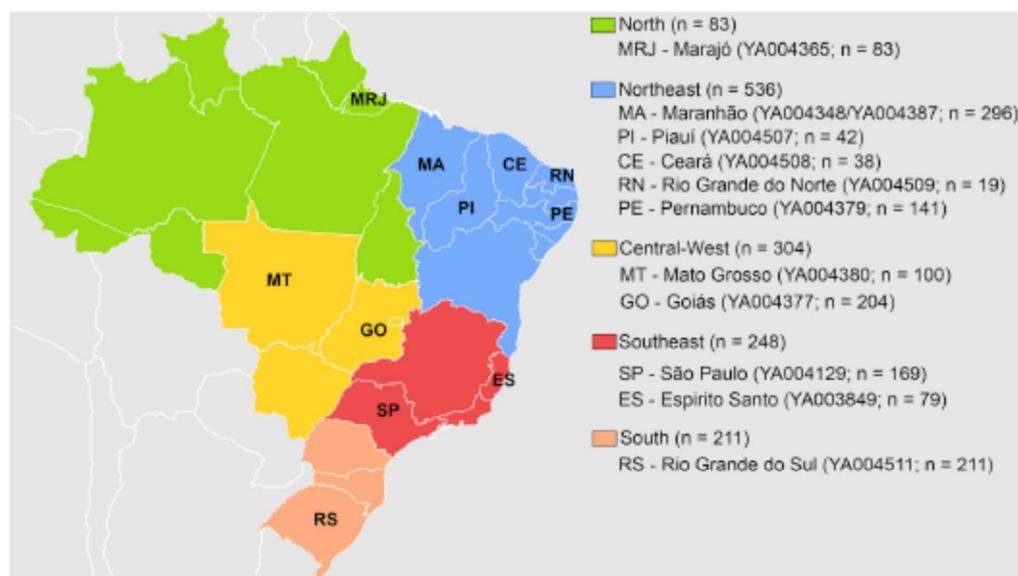


Fig. 1. Brazilian territory and its five geopolitical regions. Samples were organized into groups according to the five geopolitical regions. Brazilian states are designated by two letters, followed by their respective name, YHRD accession number and sample size.

grouped hierarchically in 6 multiplexes and 1 singleplex and typed through PCR and Single Base Extension (SBE) analysis, using the SNaPshot kit (Thermo Fisher Scientific Inc.). Multiplexes were carried out as previously described by Briton et al. [28] (Multiplexes 1 and 2), Gomes et al. [29] (Multiplexes B and E), Resque et al. [11] (Multiplex R) and Miranda-Barros F. M. (personal communication, July 2016) (Multiplex 1 Astan). The M3 (Singleplex Q) was genotyped using the primers described by Noguera et al. [30], and the SBE-primer concentration was 0.4 μ M.

The Yfiler Plus haplotypes from Maranhão, São Paulo, Rio de Janeiro and Espírito Santo were then used in haplogroups prediction using two free online software: Whit Athey's Haplogroup Predictor (<http://www.hprg.com/hapest5/index.html>) and Y-DNA Haplogroup Predictor NevGen (<http://www.nevgen.org>). Haplogroup predictions were also performed for the Yfiler marker set (after the removal of the 10 extra markers from Yfiler Plus), to evaluate the influence of the number of markers in the prediction of haplogroups, using both software cited above.

To obtain probabilities for inferred haplogroups, data from Yfiler Plus and Yfiler haplotypes were organized by FTDNA marker order and submitted to the 27-haplogroup program version of Haplogroup Predictor software, with equal priors. In NevGen software analysis, data was organized in the same way as for Haplogroup Predictor, and parameters were set as "General predictor for 23 or 17 markers" and "FTDNA order". DYS627, DYS518 and DYS387S1 were not included in the analysis, since they are not part of the FTDNA markers. All analyses were made following the models and instructions provided by the authors in the downloadable version of the software. Then, the haplogroups determined by Y-SNP genotyping were compared with those predicted by the software. The error/success rates were estimated by the counting method.

2.4. Intra and interpopulation analysis

Haplotype diversity (HD) and the number of shared haplotypes, pairwise F_{ST} genetic distances and non-differentiation probabilities were calculated with the software Arlequin v3.5 [31]. In population

comparisons, the number of repetitions in DYS389I was subtracted from DYS389II and DYS385 and DYS387S1 markers were excluded from the analysis. The comparisons were made between all Brazilian samples analyzed in this study, previously published data from Rio de Janeiro [27], and 3 samples from Native American, African and European reference populations deposited in the YHRD database: Oxapampa, Peru [Ashaninka] (YA004112 - [32]); Kenya [Bantu Luhya, Other] (YA004206 - [33]); and Madrid, Spain (YA003147). Pairwise genetic distances were visualized in three-dimensional space using multi-dimensional scaling (MDS) analysis.

3. Results and discussion

3.1. Haplotypic overview in the YHRD context

The Yfiler Plus haplotypes of the 1,382 unrelated male individuals from Brazilian populations are shown in Supplementary Table S1. All haplotypes were submitted to the YHRD *Data File Validator* tool and four alleles were not observed in YHRD database: allele 22.3 at DYS390 marker (Supplementary Table S1, sample MT052), allele 6 at DYS533 (2012176), allele 47 at DYS387S1 (IGFEC03) and allele 19 found at DYS460 (IGFEC29). The detection of allele 22.3 was not unexpected since 23.3 and 24.3 are already in YHRD for DYS390. The same applies to the allele 6 at DYS533, since allele 7 is also found in the database. Concerning alleles at DYS387S1 and DYS460 markers, both are far from the upper limit of the allele range. However, this can be explained by the still low representativeness of Yfiler Plus haplotypes in the database, missing alleles of low frequency or only present in populations not studied yet.

Null alleles can usually exist due to the occurrence of deletions or point mutations in the primer binding sites. In this study, null alleles were observed at DYS387S1, DYS533, DYS438, DYS392, YGATAH4, DYS437 and DYS448 loci. All samples were genotyped using another Y-STR kit (Yfiler kit - Applied Biosystems™ - or PPY23 kit - Promega), except for DYS387S1 that is not present in other kits. The same genotyping result was observed for DYS387S1, after a second extraction. For the null allele at DYS438 (sample ID223; Table S1), the genotyping with

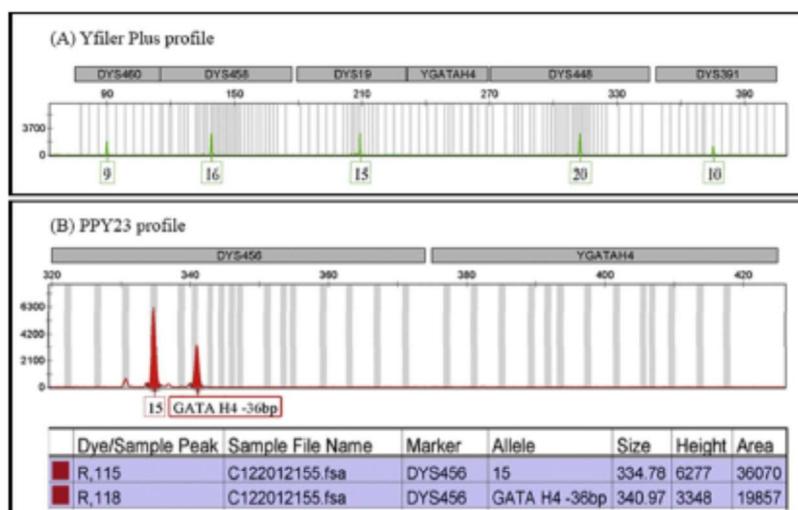


Fig. 2. PPY23 profile of the sample 2012155, showing two alleles at DYS456 and no alleles at YGATAH4 allelic range, explained by a large deletion at YGATAH4 allele.

the Yfiler kit showed an allele 10. For the sample 2012155 (Table S1), the PPY23 profile has two alleles at DYS456 and no amplification at YGATAH4 allelic range, which is compatible with a deletion of at least 36 bp at YGATAH4, involving the binding sequence for one of the Yfiler Plus primers (Fig. 2). For DYS533, DYS392, DYS437 and DYS448, null alleles were also observed for the PPY23 kit.

The occurrence of additional alleles can be attributed to a locus duplication followed by a mutational event that modifies the number of repetitions. Allele duplications with one or two differences in repeat units were observed in 13 samples. In three cases, duplications were found in DYS19, DYS448 or DYS570. The other 10 samples presented duplications in one to three loci located at the AZFa segment of the Y chromosome: DYS389I/II, DYS635, DYS391, DYS437 and DYS439 [34,35]. This explains the simultaneous duplications observed in these loci, since they are placed in a region that is known to be subject to frequent duplications due to non-homologous recombination between flanking sequences.

3.2. Haplotype diversity

High values of HD (0.9970–1.000) were found with the 26 Y-STRs in

all Brazilian samples, due to the low number of shared haplotypes inside populations (see Table 1). When just considering the 17 STRs from the Yfiler, a slight decrease in the HD was observed, except for São Paulo, Rio Grande do Norte and Ceará, indicating that both marker sets could equally discriminate individuals from the selected samples. An increase in the number of shared haplotypes inside populations was also observed for the Yfiler. This improvement in HD inside populations, obtained with the 26 Y-STRs, has been observed previously in other population samples [27,35–37].

3.3. Pairwise genetic distances

A pairwise genetic distance analysis was performed between the 11 Brazilian populations from this study plus a previously studied sample from Rio de Janeiro [27]. The resulting F_{ST} values were extremely low with non-differentiation probabilities above the significance level (Supplementary Table S2), indicating a high homogeneity among the studied populations. When taking into account reference samples from different continents, it is possible to observe that Brazilian populations are closer to the Europeans (Fig. 3), which is in accordance to the colonization history of this territory by the Europeans, mainly from

Table 1

Total number of haplotypes investigated in 12 Brazilian admixed populations, with the respective values of Haplotype Diversity (HD) and shared haplotypes, within each population, considering Yfiler Plus and Yfiler marker sets.

Population	Region	N	Yfiler Plus Shared haplotypes*	HD	Yfiler Shared haplotypes*	HD	Reference
Goias, Brazil	Central-West	204	1/3 and 5/2	0.9996 ± 0.0005	1/3 and 8/2	0.9995 ± 0.0006	This study
Mato Grosso, Brazil	Central-West	100	1/2	0.9998 ± 0.0015	3/2	0.9994 ± 0.0015	This study
Marajó, Brazil	North	83	3/2	0.9991 ± 0.0021	8/2	0.9976 ± 0.0023	This study
Ceará, Brazil	Northeast	38	0	1.0000 ± 0.0060	0	1.0000 ± 0.0060	This study
Maranhão, Brazil	Northeast	296	8/2	0.9998 ± 0.0003	19/2	0.9996 ± 0.0003	This study
Pernambuco, Brazil	Northeast	141	0	1.0000 ± 0.0008	1/2	0.9999 ± 0.0009	This study
Piauí, Brazil	Northeast	42	0	1.0000 ± 0.0052	1/2	0.9988 ± 0.0055	This study
Rio Grande do Norte, Brazil	Northeast	19	0	1.0000 ± 0.0171	0	1.0000 ± 0.0171	This study
Rio Grande do Sul, Brazil	South	211	8/2	0.9996 ± 0.0005	13/2	0.9994 ± 0.0006	This study
Espirito Santo, Brazil	Southeast	79	1/2	0.9997 ± 0.0021	2/2	0.9994 ± 0.0022	This study
Rio de Janeiro, Brazil	Southeast	258	1/2	0.99997	1/3 and 5/2	0.99976	[27]
São Paulo, Brazil	Southeast	169	3/2	0.9998 ± 0.0007	3/2	0.9998 ± 0.0007	This study
Total_Brazil		1,640	2/3 and 44/2	0.99996	1/5, 5/4, 10/3 and 125/2	0.99990	–

* x/z = x haplotypes shared by z individuals.

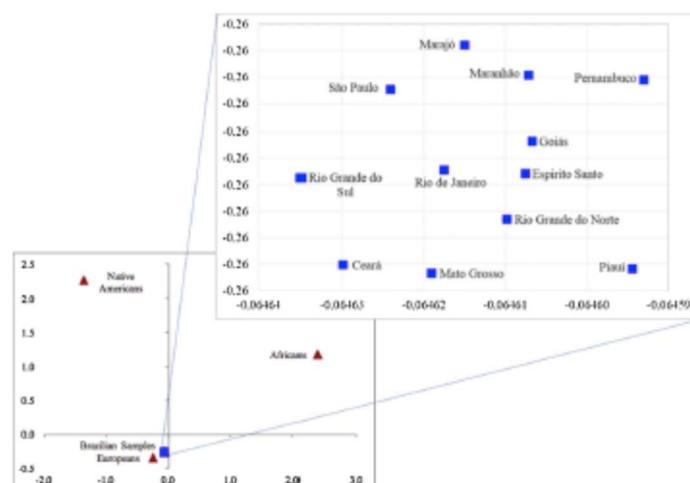


Fig. 3. MDS plot of pairwise F_{ST} genetic distances between Brazilian samples and three reference samples based on the Yfiler Plus haplotypes.

Iberian Peninsula. High genetic distances exist between Brazilian and Native-American and African populations. However, while the distances between Brazilians and natives were similar to those between Europeans and natives, the Brazilians showed lower distances to Africans than the Europeans with Africans.

In summary, concerning Y chromosome inheritance, the results from our study showed that Brazilian populations are genetically closer to the Europeans, followed by the Africans and the Native-Americans, which is in accordance to previous studies [10,19,38]. Furthermore, Admixed Brazilian populations are relatively homogenous, which support the use of a single forensic database for the Yfiler Plus haplotypes.

3.4. Haplogroup prediction tools evaluation

To evaluate the performance of Whit Athey's Haplogroup Predictor (HP) and Y-DNA Haplogroup Predictor NevGen (NG) tools in South American admixed populations, the results obtained with the use of these two software were compared with the haplogroup assignments based on Y-SNP typing. The 416 random samples selected for the haplogroup prediction tools evaluation, as well as their respective haplotypes and Y-SNP based haplogroups are shown in Supplementary Table S3.

Haplogroup predictions with Yfiler Plus (YFPlus) and Yfiler marker sets using both HP and NG tools were in agreement with the known haplogroup for 312 out of the 416 samples tested (75%). For the remaining samples, at least one marker set or software failed to predict the correct haplogroup (see Supplementary Table S4).

Both software were unable to classify 3 samples belonging to the African haplogroups A4(xCDEF), CDEF-M168*(xM213,M96,M174,M130), and one sample from the Asian haplogroup C-M130. Instead, these samples were included inside European haplogroups.

Two samples from haplogroup B2a-M150, were incorrectly assigned by HP to the haplogroup I2a1, while the NG indicated the correct haplogroup. The same happened to D-M174 samples that only were correctly assigned by the NG.

Concerning the 90 samples belonging to lineages inside haplogroup E, the HP failed to assign the correct haplogroup in 15 samples for the Yfiler Plus, and 14 samples with the Yfiler. For the NG, these values decreased to 8 and 12, respectively.

One sample classified as F-M213*(xM201,M170,12f2,M62,M9) by SNP analysis was wrongly assigned by the two software as belonging to E1a or J2 haplogroups.

All sample from G-M201 and I2a-M26 were correctly predicted in all analysis. For the 30 samples from haplogroup I-M170, the HP assigned to the haplogroup G one sample for Yfiler Plus and two for Yfiler; and the NG correctly assigned all samples.

A total of 16 samples, out of the 36 belonging to haplogroups J-12f2*(xM172) and J2-M172, were wrongly assigned with HP based on the Yfiler Plus profiles and 8 samples were wrongly assigned based on the Yfiler. NG was able to predict correctly the haplogroup of 32 and 33 samples, based on their Yfiler Plus and Yfiler profiles, respectively.

The only sample from N-Tat has its haplogroup correctly predicted in all case, except by HP for the Yfiler Plus.

Although HP showed a low error rate when predicting the haplogroup of 8 samples from Q1a2-M3, it is worth noting that this software only allowed classifying these samples as Q. NG was able to correctly include most samples inside Q1a2, although just 4 were classified as Q1a2-M3, three were placed inside other Q1a2 subclades and one was assigned to Q1b.

From the 200 samples from different subclades of haplogroup R, all were assigned to the correct haplogroup by the NG, although one from haplogroup R1b-S116 was ambiguously predict for the Yfiler Plus (R1b/R1a). The HP software showed a higher error rate, with 13 Yfiler Plus and 16 Yfiler profiles being wrongly predicted as belonging to R1a.

Finally, based on Yfiler Plus profiles, the HP did not correctly predict the haplogroup of all seven T-M70 samples, and 6 of them were correctly assigned based on the Yfiler. The NG correctly predicted the haplogroup of all samples with both marker sets.

By evaluating each software with the two markers sets, a high accuracy can be observed for the NG and, in general, the HP performs better with the 17 Yfiler markers (Fig. 4). This can be explained by the fact that the software makes the predictions based on the association of haplotype frequencies inside haplogroups, which are available in the consulted databases. The low representativeness of some haplogroups in the databases prevents its accurate assignment. According to Athey [14], by using the Bayesian approach, the addition of markers (which effectively contribute to the discriminative power) will commonly improve the probability calculation. Usually, about 10–20 markers are enough to give a probability higher than 99% for one of the haplogroups. The fact that the 17 Yfiler markers showed more accuracy can be due to the size and composition of the databases consulted by the software. Since the Yfiler markers are widely used in different laboratories for years, larger databases are available for these markers, contributing for a higher accuracy.

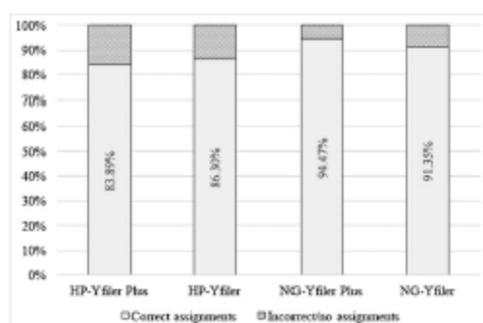


Fig. 4. Comparison of the performance of the Haplogroup Predictor and NevGen Y DNA software, for 416 Brazilian haplotype profiles for the Yfiler and Yfiler Plus markers.

Finally, it is worth noting that for haplogroups E1b1b-M78 and Q1a2-M3, although we have indicated a high error rate for NG, if we consider the same haplogroup classification level of HP (E1b1b and Q), the two software had an equivalent performance.

4. Conclusion

Representative databases from different populations are necessary in order to perform reliable forensic analysis and kinship investigations. Until now, the only Yfiler Plus data available for a Brazilian population is from a Rio de Janeiro sample [27]. In this study, we generated 1382 Yfiler Plus haplotype data from 11 Brazilian states, contributing to the improvement of the Brazilian Yfiler Plus database. Our results showed that this marker set displays high haplotype diversities in all samples from 11 Brazilian states, providing high intra-population discrimination, being suitable for forensic investigations in these populations. Nevertheless, because of the high number of haplotypes expected for the studied set of Y-STRs, other studies are still required to improve the accuracy of match probability estimates in forensic casework.

Consistent with previous studies, the Brazilian populations are homogeneous in terms of their Y-chromosomal background, being genetically closer to the Europeans, followed by the Africans and the Native-Americans. The low stratification of the Y-chromosomal background in Brazilian populations, which has been observed for different Y-STR and Y-SNP marker sets (e.g. [10,11,20,21,38]), supports the representativeness of the database obtained for forensic purposes.

Although haplogroup prediction software are widely used in population genetic researches [12,15,39], it is necessary to evaluate the accuracy of haplogroup predictions based in STR haplotypes. According to Muzzio *et al.* [15], haplogroup prediction software calibrate the frequencies based on profiles deposited in its database, and then, the haplogroups are estimated based on the reference database. Thus, the accuracy of the software prediction can depend on the number of STR profiles associated with determined haplogroup, the STR mutation rates and the deepness of haplogroup ramifications and the occurrence of convergence.

The evaluation of haplogroup prediction tools showed that Y-DNA Haplogroup Predictor NevGen software has a high accuracy than the Haplogroup Predictor, indicating that this software is more suitable in directing the Y-SNP genotyping in admixed populations from Brazil. Even so, these software could not correctly predict some haplogroups present in Brazilian admixed populations, with emphasis to those from African origin. Therefore, haplogroup prediction tools can be very useful to direct the SNP typing and analysis, but still exhibit large error rates in determining the haplogroup frequencies, leading to an underestimation of the African ancestry in American admixed populations.

In summary, results based on prediction tools must be confirmed by

Y-SNP analysis, and are not suitable for reliable estimates of haplogroup composition of the populations or to predict individual ancestry.

Acknowledgments

The authors would like to thank Pablo Martin for kindly providing unpublished Yfiler Plus haplotype data from Spain. JJ was supported by *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001*. JJ, LG (ref. 305330/2016-0) and ARS (ref. 304413/2015-1) were supported by *Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq* and *Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro - FAPERJ (CNE-2018)*. RC was supported by *Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq* (ref. 302198/2016-4) and *Fundação de Amparo à Pesquisa do Estado de São Paulo - FAPESP (2016/10138-8)*. IPATIMUP integrates the I3S research unit, which is partially supported by the Portuguese Foundation for Science and Technology (FCT). VG is supported by FCT, under the program contract provided in Decree-Law no.57/2016 of August 29.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.fsigen.2019.102163>.

References

- M. Kayser, L. Roewer, M. Hedman, L. Henke, J. Henke, S. Brauer, C. Krüger, M. Krawczak, M. Nagy, T. Dobosz, R. Szilhor, P. de Knijff, M. Stoneking, A. Sajantila, Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in father/son pairs, *Am. J. Hum. Genet.* 66 (2000) 1580–1588, <https://doi.org/10.1086/302905>.
- L. Quintana-Murci, C. Krausz, K. McElreavey, The human Y chromosome: function, evolution and disease, *Forensic Sci. Int.* 118 (2001) 169–181.
- K.N. Ballantyne, M. Goedbloed, R. Fang, O. Schaap, O. Lao, A. Wollstein, Y. Choi, K. Van Duijn, M. Vermeulen, S. Brauer, R. Decorte, M. Poetsch, N. Von Wurmb-Schwark, P. De Knijff, D. Labuda, H. Vezina, H. Knoblauch, R. Lessig, L. Roewer, R. Plösch, T. Dobosz, L. Henke, J. Henke, M.R. Furtado, M. Kayser, Mutability of Y-chromosomal microsatellites: rates, characteristics, molecular bases, and forensic implications, *Am. J. Hum. Genet.* 87 (2010) 341–353, <https://doi.org/10.1016/j.ajhg.2010.08.006>.
- M. Kayser, Forensic use of Y-chromosome DNA: a general overview, *Hum. Genet.* 136 (2017) 621–635, <https://doi.org/10.1007/s00439-017-1776-9>.
- K.N. Ballantyne, V. Keerl, A. Wollstein, Y. Choi, S.B. Zuniga, A. Ralf, M. Vermeulen, P. De Knijff, M. Kayser, A new future of forensic Y-chromosome analysis: rapidly mutating Y-STRs for differentiating male relatives and paternal lineages, *Forensic Sci. Int. Genet.* 6 (2012) 208–218, <https://doi.org/10.1016/j.fsigen.2011.04.017>.
- J.M. Butler, *Fundamentals of Forensic DNA Typing*, (2009).
- M. Kayser, R. Kittler, A. Erler, M. Hedman, A.C. Lee, A. Mohyuddin, S.Q. Mehdi, Z. Rosser, M. Stoneking, M.A. Jobling, A. Sajantila, C. Tyler-Smith, A comprehensive survey of human Y-chromosomal microsatellites, *Am. J. Hum. Genet.* 74 (2004) 1183–1197, <https://doi.org/10.1086/421531>.
- A.C. De Souza Góes, E.F. De Carvalho, L. Gomes, D.A. Da Silva, É.H. Fonseca Gil, A.A. Amorim, L. Gusmão, Population and mutation analysis of 17 Y-STR loci from Rio de Janeiro (Brazil), *Int. J. Legal Med.* 119 (2005) 70–76, <https://doi.org/10.1007/s00414-004-0504-x>.
- L. Gusmão, A. Carracedo, Y Chromosome-Specific STRs, Profiles DNA, (2003), pp. 3–6.
- A.M. Oliveira, P.M. Domingues, V. Gomes, A. Amorim, J. Jannuzzi, E.F. De Carvalho, L. Gusmão, Male lineage strata of Brazilian population disclosed by the simultaneous analysis of STRs and SNPs, *Forensic Sci. Int. Genet.* 13 (2014), <https://doi.org/10.1016/j.fsigen.2014.08.017>.
- R. Resque, L. Gusmão, M. Geppert, L. Roewer, T. Falha, L. Alvarez, Á. Ribeiro-dos-Santos, S. Santos, Male lineages in Brazil: Intercontinental admixture and stratification of the European background, *PLoS One* 11 (2016) e0152573, <https://doi.org/10.1371/journal.pone.0152573>.
- C.C. Wang, L.X. Wang, R. Shrestha, S. Wen, M. Zhang, X. Tong, L. Jin, H. Li, Convergence of Y chromosome STR haplotypes from different SNP haplogroups compromises accuracy of haplogroup prediction, *J. Genet. Genomics* 42 (2015) 403–407, <https://doi.org/10.1016/j.jgg.2015.03.008>.
- T.W. Athey, Haplogroup prediction from Y-STR values using a Bayesian-allele-frequency approach, *J. Genet. Genom.* 1 (2005) 1–7.
- T.W. Athey, Haplogroup prediction from Y-STR values using a Bayesian-allele-frequency approach, *J. Genet. Genom.* 2 (2006) 34–39.
- M. Muzzio, V. Ramallo, J.M.B. Motti, M.R. Santos, J.S. Lopez Camelo, G. Bailliet, Software for Y-haplogroup predictions: a word of caution, *Int. J. Legal Med.* 125 (2011) 143–147, <https://doi.org/10.1007/s00414-009-0404-1>.

- [16] A. Cetković, Gentula, M. Nevski, NEVGEN Y-DNA Haplogroup Predictor, (2015) (Accessed November 20, 2018), <http://www.nevgen.org/>.
- [17] D.R. Carvalho-Silva, F.R. Santos, J. Rocha, S.D. Pena, The phylogeography of Brazilian Y-chromosome lineages, *Am. J. Hum. Genet.* 68 (2001) 281–286, <https://doi.org/10.1086/316931>.
- [18] S.D.J. Pena, L. Bastos-Rodrigues, J.R. Pimenta, S.P. Bydlowski, DNA tests probe the genomic ancestry of Brazilians, *Braz. J. Med. Biol. Res.* 42 (2009) 870–876, <https://doi.org/10.1590/S0100-879X2009005000026>.
- [19] R. de, F. Figueiredo, I.B. Ambrósio, D.F. Braganholi, G. Chemale, Ju. Martins, V. Gomes, L. Gusmão, R.M.B. Cicarelli, Male-specific contributions to the Brazilian population of Espírito Santo, *Int. J. Legal Med.* (2015) 6–8, <https://doi.org/10.1007/s00414-015-1214-2>.
- [20] T. Palha, L. Gusmão, E. Ribeiro-Rodrigues, J.F. Guerreiro, A. Ribeiro-dos-Santos, S. Santos, Disclosing the genetic structure of Brazil through analysis of male lineages with highly discriminating haplotypes, *PLoS One* 7 (2012) 1–8, <https://doi.org/10.1371/journal.pone.0040007>.
- [21] S.D.J. Pena, G. di Pietro, M. Fuchshuber-Moraes, J.P. Genro, M.H. Hutz, F. de S.G. Kehdy, F. Kohlrausch, et al., The genomic ancestry of individuals from different geographical regions of Brazil is more uniform than expected, *PLoS One* (2011), <https://doi.org/10.1371/journal.pone.0017063>.
- [22] F. Manta, R. Pereira, R. Vianna, A. Rodolfo Beuttenmüller de Araújo, D. Letícia Góes Gilnã, D. Aparecida da Silva, E. de Vargas Wolgramm, I. da Mota Pontes, J. Ivan Aguiar, M. Ozório Moraes, E. Fagundes de Carvalho, L. Gusmão, Revisiting the genetic ancestry of Brazilians using autosomal AIM-Index, *PLoS One* 8 (2013) 1–11, <https://doi.org/10.1371/journal.pone.0075145>.
- [23] M. Carvalho, P. Brito, V. Lopes, L. Andrade, M.J. Anjos, F.C. Real, L. Gusmão, Analysis of paternal lineages in Brazilian and African populations, *Genet. Mol. Biol.* 33 (2010) 422–427, <https://doi.org/10.1590/S1415-47572010005000067>.
- [24] A.C. Nunes, D.A. Silva, M.A. Teixeira, D.D. Nunes, C.M. Lopes, O.R. Netto, L. Gusmão, E.F. Carvalho, M.M. Moura, Y chromosome comparative analysis of Rondonia with other Brazilian populations, *Leg. Med.* 13 (2011) 161–163, <https://doi.org/10.1016/j.legalmed.2010.12.007>.
- [25] J. Purps, S. Siebert, S. Willuweit, M. Nagy, C. Alves, R. Salazar, S.M.T. Angustiã, L.H. Santos, K. Anslinger, B. Bayer, Q. Ayub, W. Wei, Y. Xue, C. Tyler-Smith, M.B. Bañalluy, B. Martínez-Jarreta, B. Egged, B. Balitzki, S. Tschumi, D. Ballard, D.S. Court, X. Barrantes, G. Bälliger, T. West, B. Berger, H. Niederstätter, W. Parson, C. Davis, B. Budowle, H. Baur, U. Borer, C. Koller, E.F. Carvalho, P.M. Domingues, W.T. Chamoun, M.D. Coble, C.R. Hill, D. Corach, M. Caputo, M.E. D'Amato, S. Davilson, R. Decorte, M.H.D. Larmuseau, C. Ottoloni, O. Rickards, D. Lu, C. Jiang, T. Dobosz, A. Jonkka, W.E. Frank, I. Furac, C. Gehrig, V. Castella, B. Grskovic, C. Haas, J. Wobst, G. Hadzic, K. Drobnic, K. Honda, Y. Hou, D. Zhou, Y. Li, S. Hu, S. Chen, U.-D. Immel, R. Lessig, Z. Jakovski, T. Ilievskia, A.E. Klann, C.C. Garcia, P. de Knijff, T. Kraajenhout, A. Kondili, P. Miniati, M. Vouropoulou, L. Kovacevic, D. Marjanovic, I. Lindner, I. Mansour, M. Al-Azem, A. El Adari, M. Martino, S. Purifuro, L. Locarno, P. Martin, G.M. Luque, A. Alonso, L.S. Miranda, H. Moreira, N. Mizuno, Y. Iwashima, R.S.M. Neto, T.L.S. Nogueira, R. Silva, M. Nastalnczyk-Wulf, J. Edelmann, M. Kohl, S. Nie, X. Wang, B. Cheng, C. Núñez, M.M. de Pancorbo, J.K. Olofsson, N. Morling, V. Onofri, A. Tagliabracci, H. Pamjav, A. Volgyi, G. Barany, R. Pawlowski, A. Maciejewska, S. Pelotti, W. Peptinski, M. Abreu-Glowacka, C. Phillips, J. Cárdenas, D. Rey-Gonzalez, A. Salas, F. Bertéghelli, C. Capelli, U. Toscanini, A. Ficcinini, M. Pignionica, S.L. Baldassarra, R. Floski, M. Konarzewska, E. Jastrzebska, C. Robino, A. Sajanilla, J.U. Palo, E. Guevara, J. Salvador, M.C. De Ungria, J.J.R. Rodriguez, U. Schmidt, N. Schlauderer, P. Saukko, P.M. Schneider, M. Sirker, K.-J. Shin, Y.N. Oh, I. Skitsa, A. Ampati, T.-G. Smith, L.S. de Calvit, V. Stenzl, T. Capai, A. Tillmar, H. Nilsson, S. Turrina, D. De Leo, A. Verzeletti, V. Cortellini, J.H. Wetton, G.M. Gwynne, M.A. Jobling, M.R. Whittle, D.R. Sumita, P. Wolańska-Nowak, R.Y.Y. Yong, M. Krawczak, M. Nothnagel, L. Roewer, A global analysis of Y-chromosomal haplotype diversity for 23 STR loci, *Forensic Sci. Int. Genet.* 12 (2014) 12–23, <https://doi.org/10.1016/j.fsigen.2014.04.008>.
- [26] M.M.S.G. Cardena, A.J. Mansur, A.C. Pereira, C. Fridman, Y haplotype variability in a population of SÃO Paulo state, Brazil, *Forensic Sci. Int. Genet.* 23 (2016) e22–e23, <https://doi.org/10.1016/j.fsigen.2016.05.003>.
- [27] J. Jannuzzi, P. Domingues, F. Simão, A.C. da Silva Chaves, E.F. de Carvalho, L. Gusmão, Genetic characterization of Rio de Janeiro for different Y-STR sets, *Int. J. Legal Med.* (2018), <https://doi.org/10.1007/s00414-018-1776-x>.
- [28] M. Brion, B. Sobrino, A. Blanco-Verea, M.V. Lareu, A. Carracedo, Hierarchical analysis of 30 Y-chromosome SNPs in European populations, *Int. J. Legal Med.* 119 (2005) 10–15, <https://doi.org/10.1007/s00414-004-0439-2>.
- [29] V. Gomes, P. Sánchez-Diz, A. Amorim, A. Carracedo, L. Gusmão, Digging deeper into East African human Y chromosome lineages, *Hum. Genet.* 127 (2010) 603–613, <https://doi.org/10.1007/s00439-010-0808-5>.
- [30] M.C. Nogueira, A. Schwegler, V. Gomes, I. Briceño, L. Álvarez, D. Urcochea, A. Amorim, E. Benavides, C. Silveira, M. Charris, J.E. Bernal, L. Gusmão, Colombia's racial crucible: Y chromosome evidence from six admixed communities in the Department of Bolívar, *Ann. Hum. Biol.* 41 (2014) 453–459, <https://doi.org/10.3109/03014460.2013.852244>.
- [31] L. Excoffier, H. Lischer, Arlequin: an Integrated Software Package for Population Genetics Data Analysis, (2015).
- [32] D.H. Tineo, S. Lotola, F.V. Paredes, L.R. Noll, Y.C. Amaya, F. Simão, E.F. Carvalho, L. Gusmão, Genetic characterization of 27 Y-STR loci in the native population of Ashaninka from Peru, *Forensic Sci. Int. Genet. Suppl. Ser. 5* (2015) e220–e222, <https://doi.org/10.1016/j.fsigs.2015.09.088>.
- [33] G. Iacovacci, E. D'Atanasio, O. Martini, A. Coppa, D. Sellitto, B. Trombetta, A. Berti, F. Cruciani, Forensic data and microvariant sequence characterization of 27 Y-STR loci analyzed in four Eastern African countries, *Forensic Sci. Int. Genet.* 27 (2017) 123–131, <https://doi.org/10.1016/j.fsigen.2016.12.015>.
- [34] E. Bosch, M.A. Jobling, Duplications of the AZPa region of the human Y chromosome are mediated by homologous recombination between HERVs and are compatible with male fertility, *Hum. Mol. Genet.* 12 (2003) 341–347, <https://doi.org/10.1093/hmg/ddg031>.
- [35] O. Garcia, I. Yurrebaso, I.D. Manciador, S. López, S. Alonso, L. Gusmão, Data for 27 Y-chromosome STR loci in the Basque Country autochthonous population, *Forensic Sci. Int. Genet.* (2016), <https://doi.org/10.1016/j.fsigen.2015.09.010>.
- [36] J.K. Olofsson, H.S. Mogensen, A. Buchard, C. Bersting, N. Morling, Forensic and population genetic analyses of Danes, Greenlanders and Somalis typed with the Yfiler® plus PCR amplification kit, *Forensic Sci. Int. Genet.* 16 (2015) 232–236, <https://doi.org/10.1016/j.fsigen.2015.02.006>.
- [37] C. Rapone, E. D'Atanasio, A. Agostino, M. Martano, M.T. Papaluca, F. Cruciani, A. Berti, E. D'Atanasio, A. Agostino, M. Martano, M.T. Papaluca, F. Cruciani, A. Berti, Forensic genetic value of a 27 Y-STR loci multiplex (Yfiler® Plus kit) in an Italian population sample, *Forensic Sci. Int. Genet.* 21 (2015) 1–5, <https://doi.org/10.1016/j.fsigen.2015.11.006>.
- [38] D.R. Carvalho-Silva, E. Tarazona-Santos, J. Rocha, S.D.J. Pena, F.R. Santos, Y chromosome diversity in Brazilians: switching perspectives from slow to fast evolving markers, *Genetica* 126 (2006) 251–260, <https://doi.org/10.1007/s10709-005-1454-z>.
- [39] A. Salas, J.C. Jaime, V. Álvarez-Iglesias, A. Carracedo, Gender bias in the multi-ethnic genetic composition of central Argentina, *J. Hum. Genet.* 53 (2008) 662–674, <https://doi.org/10.1007/s10038-008-0297-8>.

ANEXO A – Termo de consentimento livre e esclarecido

Universidade do Estado do Rio de Janeiro

**Instituto de Biologia Roberto Alcântara Gomes
Departamento de Ecologia
Laboratório de Diagnósticos por DNA**

Consentimento Informado

1- Objetivo do Estudo

Contribuir para a compreensão da variação genética, avaliando-se marcadores genéticos associados à ancestralidade do indivíduo.

2- Procedimentos a Serem Seguidos

Para participar do estudo você deverá assinar o Consentimento Informado.

Precisamos de DNA que será obtido a partir da saliva, ou sangue. A coleta de material na cavidade oral será realizada raspando-se, suavemente o céu da boca, as buchechas e a língua com um bastão com algodão em sua extremidade (suabe estéril). Alternativamente, uma gota de sangue poderá ser coletada do dedo indicador esquerdo, ou direito, usando-se uma lanceta esterilizada.

3- Desconfortos e Riscos

Os procedimentos não causarão nenhum prejuízo para a sua saúde. Também não se corre o risco de adquirir nenhuma doença porque os materiais que serão utilizados estão sem contaminação (estéreis) e/ou são descartáveis.

4- Dados confidenciais

Os pesquisadores envolvidos no estudo irão processar os seus dados pessoais de forma estritamente confidencial. Os resultados do estudo serão incluídos em publicações científicas, as quais você poderá ter acesso. Se assim o desejar, deverá ser indicada essa opção na declaração de consentimento informado. O material biológico será armazenado no Laboratório de Diagnóstico por DNA da UERJ (Universidade do Estado do Rio de Janeiro) e será usado apenas para fins científicos ligados a este projeto.

5- Participação voluntária

Salientamos que a participação neste projeto é voluntária e que pode, a qualquer momento, decidir desvincular-se ao estudo.

Li o texto acima e estou ciente do conteúdo deste formulário de consentimento. Minhas perguntas foram respondidas. Consinto voluntariamente em participar. Recebi uma cópia deste formulário de consentimento.

Nome do participante: _____

Assinatura: _____ Data: _____

Nome do investigador: _____

Assinatura: _____ Data: _____

Gostaria de ser informado sobre os resultados gerais do projeto aquando da publicação em revistas científicas? () Sim () Não

ANEXO B – Parecer do Comitê de Ética em Pesquisa do Hospital Pedro Ernesto



UNIVERSIDADE DO ESTADO DO RIO DE JANEIRO
HOSPITAL UNIVERSITÁRIO PEDRO ERNESTO
COMITÊ DE ÉTICA EM PESQUISA



Rio de Janeiro, 14 de outubro de 2009

Do: Comitê de Ética em Pesquisa
Profª. Patrícia Maria C. O. Duque
Para: Coord. Prof. Elizeu Fagundes de Carvalho

O Comitê de Ética em Pesquisa do Hospital Universitário Pedro Ernesto, após avaliação, considerou o projeto (2485-CEP/HUPE – CAAE: 0067.0.228.000-09) "REGIÕES POLIMÓRFICAS DO DNA COMO FERRAMENTAS PARA ESTUDOS DE GENÉTICA DE POPULAÇÕES E FORENSE: UTILIZAÇÃO COMO MARCADORES DE ANCESTRALIDADE, MOVIMENTOS MIGRATÓRIOS, ESTUDOS SOBRE ESTRUTURAÇÃO GÊNICA DA POPULAÇÃO BRASILEIRA E INVESTIGAÇÕES CRIMINAIS E DE VÍNCULOS GENÉTICOS" aprovado, encontrando-se este dentro dos padrões éticos da pesquisa em seres humanos, conforme Resolução n.º196 sobre pesquisa envolvendo seres humanos de 10 de outubro de 1996, do Conselho Nacional de Saúde, bem como o consentimento livre e esclarecido.

O pesquisador deverá informar ao Comitê de Ética qualquer acontecimento ocorrido no decorrer da pesquisa.

O Comitê de Ética solicita a V. Sª., que ao término da pesquisa encaminhe a esta comissão um sumário dos resultados do projeto.


Profª. Patrícia Maria C. O. Duque
Membro do Comitê de Ética em Pesquisa
COMITÊ DE ÉTICA EM PESQUISA
HUPE/UERJ