



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

Juliana Jannuzzi Duclos do Rêgo

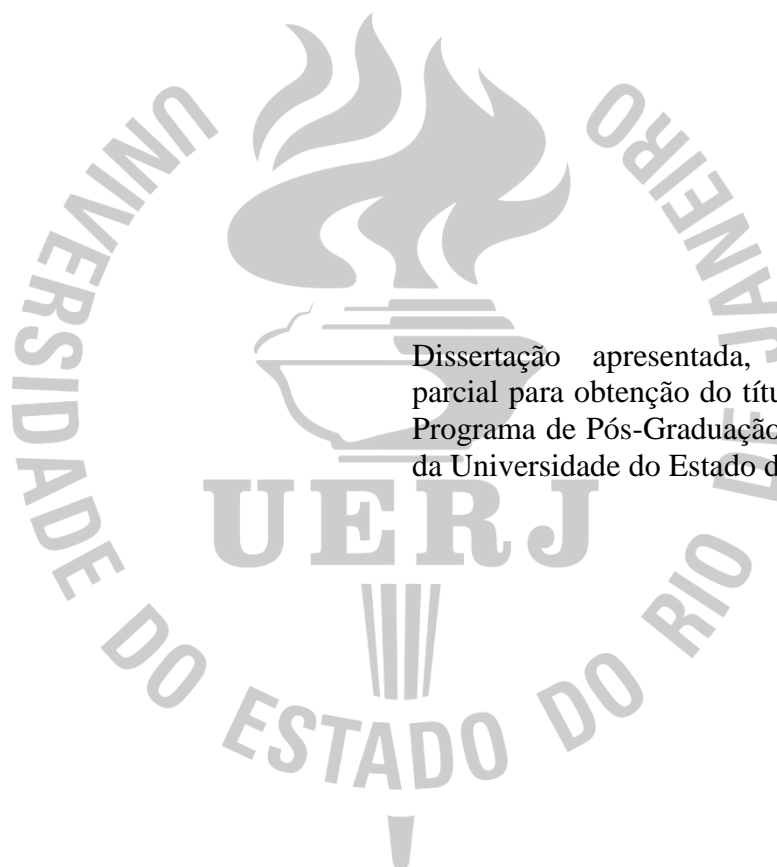
Caracterização da distribuição de alelos de *loci* STR do cromossomo Y com elevada taxa de mutação em uma amostra populacional do Rio de Janeiro

Rio de Janeiro

2015

Juliana Jannuzzi Duclos do Rêgo

Caracterização da distribuição de alelos de *loci* STR do cromossomo Y com elevada taxa de mutação em uma amostra populacional do Rio de Janeiro



Dissertação apresentada, como requisito parcial para obtenção do título de Mestre, 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. Dayse Aparecida da Silva

Rio de Janeiro

2015

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

R343 Rêgo, Juliana Jannuzzi Duclos do.

Caracterização da distribuição de alelos *de loci STR* do cromossomo Y com elevada taxa de mutação em uma amostra populacional do Rio de Janeiro / Juliana Jannuzzi Duclos do Rêgo. – 2015.
109 f.

Orientador: Elizeu Fagundes de Carvalho.
Coorientadora: Dayse Aparecida da Silva.

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

1. Cromossomo Y - Teses. 2. Marcadores genéticos - Teses. 3. Genética de populações - Teses. 4. Mutagênese - Teses. 5. Polimorfismo (Genética) - Teses. I. Carvalho, Elizeu Fagundes de. II. Silva, Dayse Aparecida da. III. Universidade do Estado do Rio de Janeiro. Instituto de Biologia Roberto Alcântara Gomes. IV. Título.

CDU 575.113(815.3)

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

Assinatura

Data

Juliana Jannuzzi Duclos do Rêgo

Caracterização da distribuição de alelos de *loci* STR do cromossomo Y com elevada taxa de mutação em uma amostra populacional do Rio de Janeiro

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

Aprovada em 02 de março de 2015.

Orientador:

Prof. Dr. Elizeu Fagundes de Carvalho
Instituto de Biologia Roberto Alcântara Gomes - UERJ

Coorientadora:

Prof.^a Dra. Dayse Aparecida da Silva
Instituto de Biologia Roberto Alcântara Gomes - UERJ

Banca Examinadora:

Prof. Dr. Rodolpho Mattos Albano
Instituto de Biologia Roberto Alcântara Gomes - UERJ

Prof. Dr. Iuri Drumond Louro
Universidade Federal do Espírito Santo

Prof. Dr. Luís Cristóvão de Moraes Sobrino Pôrto
Instituto de Biologia Roberto Alcântara Gomes - UERJ

Rio de Janeiro

2015

DEDICATÓRIA

Dedico este trabalho:

Primeiramente a Deus, por me guiar em todos os momentos da minha vida.

À minha mãe, à minha irmã e aos meus avós maternos por sempre me incentivarem e acreditarem em mim.

Ao meu grande amor Ruan, pelo companheirismo, amor, incentivo e paciência nos momentos difíceis.

AGRADECIMENTOS

Agradeço primeiramente a Deus que sempre esteve presente em minha vida me guiando e me dando força, sabedoria e coragem para sempre seguir em frente.

Aos meus pais, José Augusto (*in memoriam*) e Maria Elizabeth, por todo amor, apoio e compreensão, e por acreditarem no meu sucesso, mesmo nos momentos nos quais eu mesma duvidava. Por me incentivarem e me mostrarem os melhores caminhos a seguir.

À minha “sistar” Natália, pelas risadas e abraços em todos os momentos! Por ouvir minhas reclamações, por me ajudar nas minhas ilustrações e por sempre estar do meu lado.

À minha vóvis Marlene e ao meu vôvis Chiquinho (*in memoriam*), por todo apoio, amor, incentivo e por serem sempre presentes em todos os momentos da minha vida! Obrigada por todos os chás da tarde, que eram momentos de matar a saudade e de diversão! Obrigada por tudo! Amo vocês “mais do que pra sempre”!

Ao meu marido Ruan por todo o amor, paciência infinita, carinho e compreensão. Obrigada por estar comigo nos momentos mais difíceis, me incentivando a persistir, sempre acreditando que eu conseguiria alcançar meus objetivos, me acalmando nos momentos de estresse, e por estar ao meu lado também nos momentos mais felizes, me fazendo rir. Amo muito você!

Ao Prof. Dr. Elizeu Fagundes de Carvalho, por ter me proporcionado a oportunidade de realizar este projeto e por todo o aprendizado, apoio e orientação, que se revelaram fundamentais para a minha formação.

À Prof.^a Dra. Dayse Aparecida da Silva, pelo apoio e orientação durante o desenvolvimento deste projeto.

À Prof.^a Leonor Gusmão, pela paciência, atenção e disposição em me ajudar, principalmente com estatística.

Aos amigos Anna, Ju, Rômulo e Fê por compartilhar conhecimentos, momentos felizes e estressantes e pela companhia. Muito obrigada!

A Monique, Holly, Silvia, Saulo, Rossana, Fernanda, Márcia pela ajuda técnica e pela atenção.

A Jaque, Flavio, Henrique, Marcos e Ednardo pela ajuda e colaboração em diversas etapas deste trabalho.

À minha família, pela força e por estar sempre presente em minha vida.

A todos os meus amigos por cada momento vivido que estão sempre guardados na minha memória.

Ao CNPq e a FAPERJ pelo apoio financeiro.

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

RESUMO

RÊGO, Juliana Jannuzzi Duclos do. *Caracterização da distribuição de alelos de loci STR do cromossomo Y com elevada taxa de mutação em uma amostra populacional do Rio de Janeiro*. 2015. 109 f. Dissertação (Mestrado em Biociências) – Programa de Pós-Graduação em Biociências, Universidade do Estado do Rio de Janeiro, 2015.

Marcadores genéticos presentes no cromossomo Y, como os microssatélites (Y-STRs) e polimorfismos de único nucleotídeo (Y-SNPs) são utilizados na caracterização de linhagens masculinas, visto que são transmitidos às gerações seguintes sem alterações, a menos que ocorram mutações (Singh *et al.*, 2011; Mitchell & Hammer, 1996; Butler, 2009). Por isso, esses marcadores são amplamente empregados em diversas situações, destacando-se o uso constante dos Y-STRs na genética forense por apresentarem alta capacidade de discriminar linhagens. Recentemente, foram descritos 13 marcadores com taxas de mutação substancialmente superiores àquelas verificadas para *loci* STR do cromossomo Y, denominados “*Rapidly Mutating*” (RM) Y-STRs (Ballantyne *et al.*, 2010; Kayser *et al.*, 2012). Devido às taxas de mutação elevadas, os RM-YSTRs apresentam maior eficiência na discriminação entre indivíduos proximamente relacionados, pertencentes à mesma linhagem patrilínea. O presente trabalho buscou aprofundar o conhecimento acerca das características populacionais e mutacionais dos *loci* RM-YSTRs em amostra do Rio de Janeiro, contribuindo com estudos desta natureza na população brasileira. Realizou-se a análise de 13 *loci* do cromossomo Y em 258 indivíduos do sexo masculino, compondo 129 pares de pais e filhos, nascidos no estado do Rio de Janeiro. O DNA das amostras foi extraído, conforme os protocolos vigentes na rotina do LDD-UERJ. As sequências genéticas de interesse foram amplificadas pela técnica de reação em cadeia da polimerase (PCR) através da realização de três PCR multiplex, cujos produtos de amplificação foram separados por eletroforese em sequenciador automático ABI-3500 (Applied Biosystems). Para os pares pai/filho que apresentaram haplótipos mutados, empregou-se a técnica de sequenciamento para confirmação das mutações. Os *loci* RM-YSTR geraram um poder de discriminação de 1,0 na amostra analisada, o que significa que todos os 129 indivíduos da amostra populacional apresentaram haplótipos diferentes para tais marcadores, com frequências de 0,0077 e diversidade haplotípica igual a 1. Além disso, foram obtidos valores elevados de diversidade gênica para os 13 marcadores. A análise de distância genética e os resultados de AMOVA baseados nos valores de F_{st} demonstraram que os RM-YSTR não indicam subdivisão populacional e traços ancestrais comuns. Tais valores estão associados às elevadas taxas de mutação encontradas, cuja média foi de $2,11 \times 10^{-2}$. Foi possível observar que os *loci* RM-YSTR são muito discriminativos na amostra miscigenada analisada, além de terem maior capacidade de diferenciar indivíduos do que outros conjuntos de marcadores normalmente usados em estudos populacionais e análises forenses. Sendo assim, é possível concluir que os marcadores RM-YSTR são promissores para discriminar indivíduos da mesma linhagem patrilínea, visto que devido às suas elevadas taxas mutacionais e poder de discriminação, são capazes de diferenciar indivíduos de maneira mais eficiente do que os outros conjuntos de STR. Porém, é necessário maior número de estudos para melhor caracterização destes *loci* em diferentes populações.

Palavras-chave: Cromossomo Y. Marcadores genéticos. Microssatélites. RM-YSTR.

ABSTRACT

RÊGO, Juliana Jannuzzi Duclos do. *Characterization of the STR loci alleles's distribution of Y chromosome with high mutation rate in a population sample of Rio de Janeiro*. 2015. 109 f. Dissertação (Mestrado em Biociências) – Programa de Pós-Graduação em Biociências, Universidade do Estado do Rio de Janeiro, 2015.

Genetic markers on Y chromosome, as microsatellites (Y-STRs) and single nucleotide polymorphisms (Y-SNPs) are used for the characterization of male lineages, since they are fully transmitted to next generations unless mutations occurs (Singh *et al.*, 2011; Mitchell & Hammer, 1996; Butler, 2009). Therefore, these markers are widely applied in several situations, highlighting the constant use of Y-STRs in the field of forensic genetics because of their high capacity of discriminate lineages. Recently, 13 rapidly mutating markers were described due to their highly mutation rates in comparison to other common Y-chromosome STRs, being called as “Rapidly Mutating Y-STR (RM-YSTR) (Ballantyne *et al.*, 2010; Kayser *et al.*, 2012). As a result of their high mutation rates, RM-YSTRs display high efficiency in discriminating paternally related males. The present work aimed to deepen the knowledge about population and mutational RM-YSTR loci characteristics in Rio de Janeiro sample, and then, contribute to other studies with this purpose in Brazilian population. Y chromosome 13 STRs analysis was realized in 258 males born in Rio de Janeiro state, grouped in 129 fathers/sons pairs. The extraction of DNA from biological samples was performed according to routine protocols from LDD-UERJ. Target sequences were amplified by three polimerase chain reactions (PCR) and the amplicons were separated through electrophoresis on automated sequencer ABI-3500 (Applied Biosystems). When mutations were detected, they were confirmed by sequencing. Among the investigated sample, RM-YSTR loci showed a discrimination capacity of 1,0 which means that all 129 analyzed individuals have different haplotypes for these markers, displaying frequencies of 0,0077 and haplotype diversity of 1,0. Moreover, high values of genetic diversities were obtained for the 13 markers. Distance genetic analysis and AMOVA values based on F_{st} results did not show population substructure and common ancestral traits. These results are associated with high mutation rates found, with an average rate about $2,11 \times 10^{-2}$. RM-YSTR showed to be very discriminative at this mixed sample, besides proving to be more discriminative than other markers commonly used in population studies and forensic analysis. Thus, it is possible to conclude that RM-YSTR markers are promising to discriminate individuals of the same male strain and due to their high mutation rates and discrimination capacity, they are able to differentiate individuals better than other common markers. Nevertheless, for a better characterization of these loci in different populations more studies are needed.

Keywords: Y chromosome. Genetic markers. Microsatellites. RM-YSTR.

LISTA DE FIGURAS

Figura 1 –	Caracterização do genoma humano.....	16
Figura 2 –	Representação esquemática dos cromossomos sexuais X e Y.....	21
Figura 3 –	Perfil genético do indivíduo 120209-23SP da população do Rio de Janeiro obtido através da tipagem de quatro marcadores RM-YSTR amplificados pelo multiplex RM1.....	44
Figura 4 –	Perfil genético do indivíduo 120209-23SP da população do Rio de Janeiro obtido através da tipagem de quatro marcadores RM-YSTR amplificados pelo multiplex RM2.....	45
Figura 5 –	Perfil genético do indivíduo 120209-23SP da população do Rio de Janeiro obtido através da tipagem de cinco marcadores RM-YSTR amplificados pelo multiplex RM3.....	46
Figura 6A –	Eletroferograma de sequenciamento de alelos do locus DYS627.....	74
Figura 6B –	Eletroferograma de sequenciamento de alelos do locus DYS627.....	75

LISTA DE GRÁFICOS

Gráfico 1 -	Distribuição das frequências haplotípicas do locus DYF387S1 na população do estado do Rio de Janeiro (N=129).....	56
Gráfico 2 -	Distribuição das frequências alélicas do locus DYS570 na população do estado do Rio de Janeiro (N=129).....	57
Gráfico 3 -	Distribuição das frequências alélicas do locus DYS576 na população do estado do Rio de Janeiro (N=129)	57
Gráfico 4 -	Distribuição das frequências haplotípicas do locus DYF399S1 na população do estado do Rio de Janeiro (N=129)	58
Gráfico 5 -	Distribuição das frequências alélicas do locus DYS518 na população do estado do Rio de Janeiro (N=129)	59
Gráfico 6 -	Distribuição das frequências alélicas do locus DYS626 na população do estado do Rio de Janeiro (N=129)	59
Gráfico 7 -	Distribuição das frequências alélicas do locus DYS526a/b na população do estado do Rio de Janeiro (N=129)	60
Gráfico 8 -	Distribuição das frequências alélicas do locus DYS627 na população do estado do Rio de Janeiro (N=129)	61
Gráfico 9 -	Distribuição das frequências alélicas do locus DYS449 na população do estado do Rio de Janeiro (N=129)	61
Gráfico 10 -	Distribuição das frequências haplotípicas do locus DYF403S1a/b na população do estado do Rio de Janeiro (N=129)	62
Gráfico 11 -	Distribuição das frequências haplotípicas do locus DYF404S1 na população do estado do Rio de Janeiro (N=129)	63
Gráfico 12 -	Distribuição das frequências alélicas do locus DYS547 na população do estado do Rio de Janeiro (N=129)	64
Gráfico 13 -	Distribuição das frequências alélicas do locus DYS612 na população do estado do Rio de Janeiro (N=129)	64

LISTA DE TABELAS

Tabela 1 -	Descrição dos marcadores RM-YSTRs.....	26
Tabela 2 -	Loci RM-YSTR com os respectivos pares de primers usados para amplificação por PCR e os fluoróforos empregados.....	32
Tabela 3 -	Soluções de primers utilizadas nas reações multiplexes.....	33
Tabela 4 -	Condições termocíclicas utilizadas para amplificações dos multiplex RM1, RM2 e RM3.....	34
Tabela 5 -	Condições termocíclicas utilizadas para as reações de sequenciamento dos produtos purificados.....	38
Tabela 6 -	Populações utilizadas para análise de distância genética, seus respectivos tamanhos e referências bibliográficas.....	42
Tabela 7 -	Haplótipos RM-YSTR de indivíduos do Rio de Janeiro (Continua).....	48
Tabela 8 -	Diversidade gênica nos 13 loci RM-YSTR.....	65
Tabela 9 -	Valores de distância genética baseados no Fst.....	67
Tabela 10 -	Valores de P para o índice de Fst.....	67
Tabela 11 -	Matriz de significância dos valores de P para o índice de Fst.....	68
Tabela 12 -	Valores de AMOVA baseados no índice de Fst resultantes da comparação entre amostras miscigenadas do continente Americano e amostras da África Subsaariana, Europa e Ásia.....	69
Tabela 13 -	Haplótipos RM-YSTR dos pares de pai/filho nos quais foram observadas mutações(Continua)	70
Tabela 14 -	Características mutacionais dos loci RM-YSTR na população miscigenada do Rio de Janeiro.....	76

LISTA DE ABREVIATURAS E SIGLAS

AMEL	gene da amelogenina
AMOVA	Análise de Variância Molecular
CCD	<i>charge-couple device</i>
DNA	Ácido desoxirribonucléico
dNTP	Nucleoside triphosphate
EDTA	Ácido etilenodiaminotetracético
f	Frequência
Fst	Índice de Fixação
ISFG	<i>International Society of Forensic Genetics</i>
LDD	Laboratório de Diagnósticos por DNA
N	Número
NR1	<i>Non-recombining Region of the Y chromosome</i>
P	Índice de significância
PAR	<i>Pseudoautosomal Regions</i>
PCR	<i>Polimerase Chain Reaction</i>
PCR-RFLP	<i>Restriction Fragment Length Polymorphism</i>
pH	Potencial hidrogênico
RFUs	Unidades Relativas de Fluorescência
RM-YSTRs	<i>Rapidly Mutating Y-STRs</i>
s	segundos
SNP	<i>Single Nucleotide Polymorfism</i>
STR	<i>Short Tandem Repeats</i>
SWGAM	<i>U.S Scientific Working Group on DNA Analysis Methods</i>
TE	Tampão tris, EDTA
Tris	Trihidroximetil aminometano
UERJ	Universidade do Estado do Rio de Janeiro
VNTR	<i>Variable Number of Tandem Repeats</i>
YHRD	<i>Haplotype Reference Databases</i>

LISTA DE SÍMBOLOS

%	Porcentagem
Yp	Braço curto do cromossomo Y
Yq	Braço longo do cromossomo Y
Kb	1000 pares de bases
=	Igual
mL	Mililitro
-	Menos
°C	Grau Celsius
Mb	Megabases
mM	Milimolar
μL	Microlitro
ng	Nanograma
μM	Micromolar
pb	Pares de bases
F	<i>Forward</i>
R	<i>Reverse</i>
g	Gravidade
+	Mais
Vic [®]	Patenteado pela <i>Applied Biosystems</i>
NED [™]	Patenteado pela <i>Applied Biosystems</i>
6-FAM	6-carboxi-fluoresceína
nm	Nanômetros
Σ	Somatório
≤	Menor ou igual
>	Maior
C	Criança
SP	Suposto pai

SUMÁRIO

	INTRODUÇÃO	15
1	OBJETIVOS	28
1.1	Objetivo geral	28
1.2	Objetivos específicos	28
2	MATERIAL E MÉTODOS	29
2.1	Seleção da amostra populacional	29
2.2	Extração de DNA genômico por Chelex	30
2.3	Análise Molecular do Cromossomo Y	30
2.3.1	<u>Amplificação de <i>loci</i> hipervariáveis por sistemas multiplex</u>	30
2.3.1.1	Conjuntos de <i>primers</i> utilizados nas reações multiplex.....	31
2.3.1.2	Reações de amplificação multiplex.....	33
2.3.2	<u>Eletroforese capilar dos produtos da amplificação</u>	34
2.3.2.1	Preparação dos produtos da PCR.....	35
2.3.2.2	Eletroforese e detecção dos produtos de amplificação.....	35
2.3.2.3	Análise dos produtos de amplificação.....	36
2.3.3	<u>Confirmação de mutações</u>	36
2.3.3.1	Amplificação através de PCR monoplex.....	37
2.3.3.2	Eletroforese em gel de agarose.....	37
2.3.3.3	Sequenciamento.....	38
2.4	Análises estatísticas	39
2.4.1	<u>Estimativas das frequências alélicas e haplotípicas</u>	39
2.4.2	<u>Cálculo da diversidade gênica por <i>locus</i> e da diversidade haplotípica</u>	40
2.4.3	<u>Poder de discriminação de haplótipos</u>	40
2.4.4	<u>AMOVA</u>	40
2.4.5	<u>Parâmetros de distância genética</u>	41
2.4.6	<u>Análise mutacional de <i>loci</i> RM-YSTR</u>	42
3	RESULTADOS	43
3.1	Haplótipos do cromossomo Y em indivíduos miscigenados do Rio de Janeiro	43
3.2	Estimativa das Frequências Haplotípicas e Poder de Discriminação	55

3.3	Estimativa das Frequências alélicas dos <i>loci</i> RM-YSTR	55
3.4	Análise da Diversidade Gênica por <i>locus</i> e Diversidade Haplotípica	65
3.5	Análise da distância genética entre populações	66
3.6	Estudo mutacional dos <i>loci</i> RM-YSTR	69
3.6.1	<u>Detecção de mutações nos <i>loci</i> RM-YSTR</u>	69
3.6.2	<u>Análise da taxa de mutação</u>	76
4	DISCUSSÃO	77
4.1	Determinação dos haplótipos e nomenclatura dos RM-YSTRs	77
4.2	Frequência Haplotípica e Poder de Discriminação	81
4.3	Diversidade Gênica e Haplotípica	82
4.3.1	<u>Diversidade Gênica</u>	82
4.3.2	<u>Diversidade Haplotípica</u>	83
4.4	Distância Genética	84
4.5	Estudo Mutacional	85
	CONCLUSÃO	88
	REFERÊNCIAS	89
	ANEXO A - Toward Male Individualization with Rapidly Mutating Y-Chromosomal Short Tandem Repeats	92
	ANEXO B - Male lineage strata of Brazilian population disclosed by the simultaneous analysis of STRs and SNPs	104
	ANEXO C - Parecer do Comitê de Ética em Pesquisa do Hospital Pedro Ernesto	104

INTRODUÇÃO

Genoma Humano e Marcadores genéticos

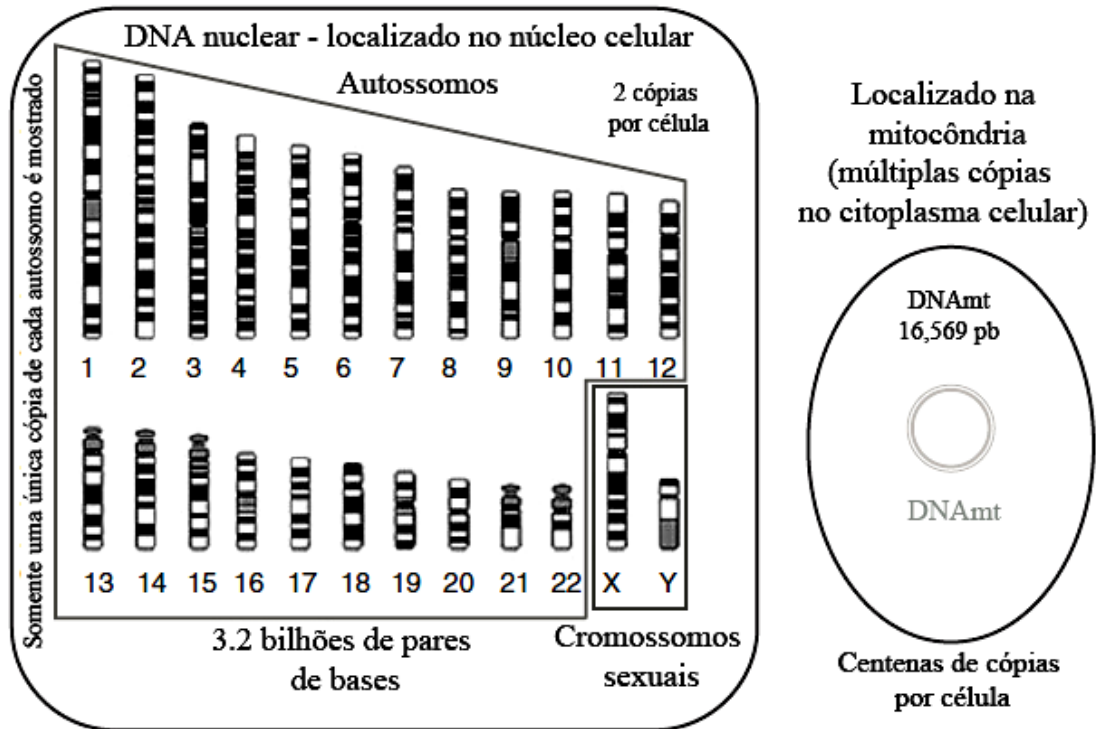
O genoma humano é formado por moléculas de ácido desoxirribonucleico (DNA) nas quais se encontram armazenadas informações genéticas de um indivíduo (An integrated encyclopedia of DNA elements in the human genome, 2012). Cada ser humano possui, aproximadamente, 100 trilhões de células e, cada célula, tem a capacidade de armazenar as moléculas de DNA no seu interior, em estruturas denominadas cromossomos. A maior parte do material genético está presente no núcleo celular e, uma menor parte fica em organelas que se encontram fora do núcleo, denominadas mitocôndrias (Butler, 2009).

Cada célula somática possui dois pares de cromossomos, um de origem materna e outro de origem paterna, sendo assim chamada de célula diplóide. Nos seres humanos, as células diplóides apresentam 22 pares de cromossomos autossômicos homólogos e um par de cromossomos sexuais (Figura 1). Aproximadamente, 98% do DNA humano está contido nos cromossomos autossômicos e no cromossomo sexual X, presente em uma cópia nos homens e duas nas mulheres (Jobling, 2012). Os 2% restantes do genoma humano estão distribuídos em dois segmentos de DNA de herança uniparental, que consistem no DNA mitocondrial e no cromossomo Y (Figura 1).

O DNA mitocondrial consiste em uma molécula de DNA circular, com 16.569 pares de base e é encontrado nas mitocôndrias, que são organelas presentes no citoplasma celular que têm como função principal gerar energia para a célula. Dependendo do tipo celular, o número de mitocôndrias por célula varia, podendo ser encontradas cerca de centenas a milhares de cópias destas organelas. As mitocôndrias apresentam herança matrilinear, sendo transmitidas somente pelas mulheres aos seus filhos, de ambos os sexos (Butler, 2009).

Com relação ao cromossomo Y, cuja transmissão ocorre de pai para filho ao longo das gerações, a herança genética é unicamente patrilinear. Assim, este cromossomo apresenta herança uniparental e não se recombina com o cromossomo X, sendo transmitido integralmente às gerações futuras, a menos que ocorram eventos mutacionais (Jobling, 2012).

Figura 1 - Caracterização do genoma humano



Legenda: O genoma humano é constituído por 23 pares de cromossomos nucleares e um pequeno DNA circular conhecido como DNA mitocondrial. Os cromossomos 1 até o 22 são numerados de acordo com seu tamanho e o par é formado por um cromossomo de origem paterna e um de origem materna. Já o DNA mitocondrial se encontra nas mitocôndrias e é uma herança exclusivamente materna. Fonte: Butler, 2009– *com modificações*.

De uma maneira geral, as moléculas de DNA são transmitidas às gerações futuras, podendo adquirir eventuais mutações, apresentando assim, variações entre indivíduos (Jobling, 2012). Estima-se que os seres humanos apresentem 99,7% de semelhança nos seus genomas, podendo ser diferenciados a nível individual pelos 0,3% restantes que apresentam alterações (Butler, 2011). As diferenças nas sequências de DNA de dois ou mais indivíduos são decorrentes de polimorfismos genéticos e consistem na base para a identificação humana através do DNA. O padrão de polimorfismos caracteriza o perfil genético de cada indivíduo, fato extremamente relevante em questões forenses, sendo observado em estudos populacionais, que remetem à história demográfica das populações, permitindo traçar movimentos migratórios, expansões e colonizações (Rodenbusch, 2008; Jobling, 2012).

O emprego do DNA como ferramenta na identificação humana foi primeiramente sugerido em 1985 por Alec Jeffreys e colaboradores, os quais afirmaram que certas regiões do

genoma humano eram equivalentes a “impressões digitais” do DNA, devido a sua habilidade de individualizar as pessoas com um padrão multibandas, tal qual um código de barras. Além disso, este grupo sugeriu que, através da análise das sequências dessas regiões genômicas, também seria possível rastrear padrões de herança genética. Posteriormente, tais regiões passaram a ser amplamente utilizadas como marcadores genéticos, sendo que os mais utilizados em identificação humana encontram-se localizados em sítios não-codificantes intergênicos ou em áreas intragênicas, mais especificamente nos íntrons (Butler, 2009). À localização de marcadores no genoma deu-se o nome de *locus* (no plural, *loci*). Desde a descoberta de Alec Jeffreys e colaboradores, em 1985, milhares de *loci* vem sendo caracterizados e utilizados com diversos objetivos, de acordo com o padrão de variação destes marcadores.

Marcadores genéticos empregados na identificação humana por DNA apresentam diferenças na sequência e no comprimento de segmentos de nucleotídeos entre indivíduos. A variação de sequências polimórficas pode ser atribuída a alterações do tipo troca de base, denominada *Single Nucleotide Polymorphism* (SNP ou Polimorfismos de Base Única) ou a diferenças no tamanho do gene, causadas pela alteração do número de sequências repetitivas consecutivas (*in tandem*). Tais sequências são classificadas de duas maneiras, de acordo com o número de pares de bases que formam a estrutura repetitiva: *Variable Number of Tandem Repeats* (VNTR) ou Números Variáveis de Repetições *in Tandem* e *Short Tandem Repeats* (STR) ou Repetições Curtas *in Tandem* (Rodenbusch, 2008; Butler, 2009).

De acordo com as frequências com que esses marcadores variam na população (taxas mutacionais), com o entendimento das propriedades mutacionais de cada *locus* e com o conhecimento da distribuição populacional dos marcadores genéticos, eles são candidatos para diversas aplicações, incluindo estudos de movimentos migratórios dos povos através dos séculos, estudos antropológicos com reconstrução da origem remota dos indivíduos da população atual, além de aplicações na ciência forense e na investigação de vínculos genéticos.

Single Nucleotide Polymorfism (SNP)

Os marcadores de DNA mais frequentemente utilizados são os SNPs. Eles estão presentes ao longo do genoma e apresentam baixas taxas de mutação, de aproximadamente 10^{-8} por nucleotídeo por geração (Jobling, 2012). Eles consistem em mutações pontuais de única base que ocorrem ao acaso e são geralmente marcadores binários, ou seja, há normalmente quatro possibilidades de base para aquele *locus*. Além disso, são passados ao longo das gerações e sua baixa taxa de mutação indica que a probabilidade de encontrar um SNP que sofreu mutação independentemente em dois genomas é pequena, sendo assim, os SNPs são capazes de associar indivíduos por descendência, caracterizando linhagens (Jobling & Tyler-Smith, 1995; Buttlar, 2003; Bird, 2012; Jobling, 2012).

Os SNPs apresentam como vantagens em relação a outros tipos de marcadores o fato de serem facilmente observados e genotipados, além do fato de que teorias da genética de populações conseguem prever a longo prazo a dinâmica destes marcadores, devido a todo o conhecimento sobre o código genético e sinais regulatórios, torna-se moderadamente simples prever as consequências funcionais de uma troca de base (Armour, 2006).

Sendo assim, os marcadores SNPs são melhores empregados em estudos evolutivos, em caracterização de linhagens e em estudos clínicos.

Variable Number of Tandem Repeats (VNTR)

Os marcadores do tipo VNTR, também conhecidos como minissatélites, compreendem várias unidades de repetições sucessivas *in tandem* e cada unidade de repetição é formada por 8 a 80 pares de base (pb). Os alelos gerados por este marcador diferem entre si pelo número exato de repetições assim como pelo comprimento, em pares de base, do *locus* do minissatélite, sendo assim, os alelos podem ser identificados pelo tamanho do seu fragmento.

Os minissatélites representam alguns dos *loci* mais variáveis no genoma humano, apresentando uma taxa de mutação em torno de 10^{-2} e, por isso, foram inicialmente utilizados em estudos de ligação e identificação por DNA, sendo considerados os marcadores mais informativos que contribuíam para as “impressões digitais” do DNA (Jeffreys *et al.*, 1999).

Além disso, os marcadores VNTR também foram considerados um marco no uso da técnica de análise de polimorfismos de comprimento em fragmentos gerados por enzimas de restrição (RFLP - *Restriction Fragment Length Polymorphism*) na rotina de laboratórios.

Devido à sua elevada taxa de mutação e sua capacidade de discriminar indivíduos, os *loci* minissatélites foram amplamente empregados na resolução de casos forenses. Ao mesmo tempo, foram sendo descobertas algumas limitações quanto ao uso de tais marcadores, principalmente no que diz respeito à técnica de RFLP, usada na análise dos minissatélites. A dificuldade é encontrada, principalmente, nas grandes quantidades de DNA necessárias para o bom funcionamento da técnica (>50ng), assim como moléculas de DNA de alto peso molecular (>12kb). Como as amostras encontradas em casos forenses são, geralmente, compostas por moléculas de DNA degradadas e em quantidades limitadas, a metodologia empregada na análise de RFLP não se mostra efetiva, além de ser demorada (Butler, 2009).

Short Tandem Repeats (STR)

Os marcadores STR ou microssatélites são sequências repetitivas *in tandem*, compostas de 1 a 7 pares de base, localizadas ao longo do genoma e se caracterizam por serem marcadores altamente polimórficos, apresentando taxa de mutação na ordem de 10^{-3} (Kayser *et al.*, 2004; Rodenbusch, 2008; Butler, 2009; Ballantyne *et al.*, 2010; Kayser *et al.*, 2012; Bull *et al.*, 1999). As sequências repetitivas que compõem o STR são nomeadas de acordo com o comprimento da unidade de repetição. Sendo assim, uma sequência composta pela repetição de três nucleotídeos é denominada unidade repetitiva trinucleotídica.

Além de apresentarem variação no comprimento da unidade de repetição, tais sequências também variam quanto ao número de repetições e ao padrão que as unidades repetitivas se organizam. Visto isso, os microssatélites são divididos em categorias de acordo com o padrão das unidades repetitivas. Os marcadores STRs simples são formados por unidades repetitivas idênticas no comprimento e na sequência, já os compostos são formados por duas ou mais repetições simples adjacentes, porém divergentes entre si. Existem ainda os STRs complexos, os quais possuem vários blocos de unidades repetitivas, que variam em comprimento e em sequências que se intercalam com as unidades repetitivas. Além destes, também há os STRs caracterizados por repetições complexas hipervariáveis, que dão origem a alelos não-

consensuais. Estes alelos apresentam diferenças no tamanho e na sequência, conferindo a estes marcadores uma dificuldade de genotipagem (Butler, 2009). Além disso, é importante ressaltar que nem todos os alelos STRs são formados por unidades de repetições completas. Tais alelos são denominados alelos microvariantes, visto que contém unidades de repetições incompletas (Butler, 2011).

Uma grande gama de microssatélites tem sido descoberta no genoma humano ao longo dos anos e, acredita-se que eles constituam, aproximadamente, 3% do total das sequências de DNA (Bull *et al.*, 1999; Rodenbusch, 2008; Butler, 2009). Ainda não há muito conhecimento sobre alguns *loci* STR, principalmente no que diz respeito às funções que eles possam desempenhar no genoma, assim como os padrões de variação de suas sequências e suas propriedades mutacionais. Entretanto, um grande número de marcadores microssatélites tem sido caracterizado por laboratórios comerciais e de pesquisa.

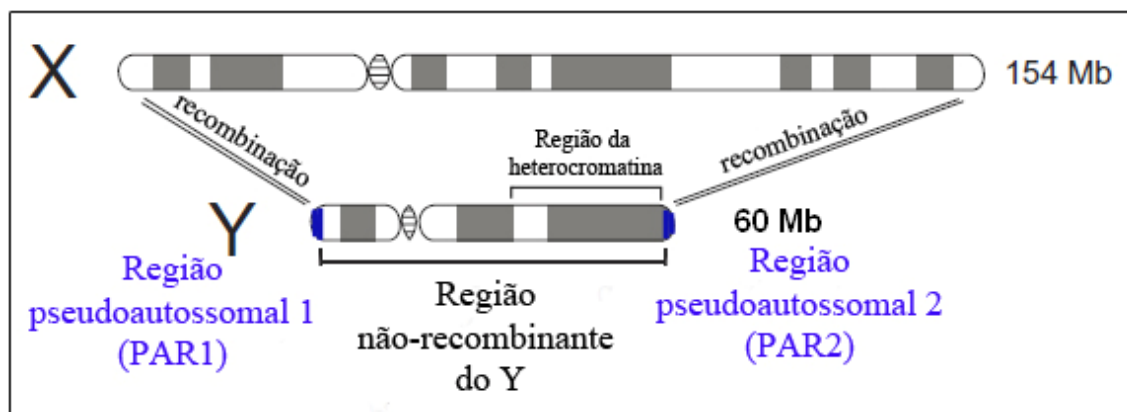
Como o número de repetições que compõem os STRs pode ser extremamente variável entre os indivíduos, eles têm sido muito utilizados em mapeamentos genéticos, em estudos populacionais e, principalmente, na área forense. Devido à sua elevada taxa mutacional e ao fato de serem facilmente detectados através da técnica da reação em cadeia da polimerase (PCR), o uso destes marcadores se tornou muito popular e se revelou uma poderosa ferramenta para a identificação humana. Tendo isto em vista, os STRs localizados em cromossomos autossômicos são o primeiro foco de uma análise. Entretanto, em análises de vínculo genético por herança paterna quando o suposto pai é falecido ou não encontrado, assim como em casos de investigação de suspeitos de violência sexual, o uso dos *loci* STR do cromossomo Y demonstra ser uma ferramenta complementar de grande utilidade (Kayser & Sajantila, 2001; Rodenbusch, 2008; Kayser *et al.*, 2012).

O cromossomo Y

O cromossomo Y é o menor cromossomo humano, apresentando um tamanho de, aproximadamente, 60 milhões de pares de base. Ele apresenta em sua extensão a região de determinação sexual (SRY, do inglês *Sex-determining Region*), que está localizada na posição de 2,56 megabases (Mb), e o gene da amelogenina, também conhecido como gene *AMEL Y*, localizado na região de 6,70Mb do Y (Butler, 2003; Singh *et al.*, 2011; Butler, 2011).

Nas porções distais dos braços curto (Yp) e longo deste cromossomo estão localizadas regiões conhecidas como *pseudoautosomal regions* (PAR1 e PAR2) ou regiões pseudoautosossomais, que estão sujeitas à ocorrência de *crossing-over* (recombinação) com as regiões homólogas do X durante a meiose (Figura 2). Entretanto, em aproximadamente 95% da extensão do cromossomo Y não há recombinação, sendo assim denominada região não-recombinante do Y (NRY) (Ali & Hasnian, 2002). A NRY é constituída por duas porções, denominadas de eucromatina e heterocromatina. A eucromatina apresenta cerca de 23Mb de extensão, ocupando o braço pequeno (Yp) e o centrômero do cromossomo Y, além de uma pequena parte do braço longo do Y (Yq). Esta porção é conhecida por ser uma zona codificante, na qual se encontram genes responsáveis por importantes funções biológicas do organismo masculino. Enquanto que a heterocromatina está presente no braço longo do Y e possui uma abrangência de, aproximadamente, 40 Mb (Figura 2). É formada por sequências altamente repetitivas, o que dificulta a obtenção de sequenciamentos confiáveis e, além disso, não é transcrita (Quintana-Murci *et al.*, 2001; Singh *et al.*, 2011; Bachtrog, 2013).

Figura 2 - Representação esquemática dos cromossomos sexuais X e Y



Legenda: As extremidades do cromossomo Y (PAR1 e PAR2) sofrem *crossing-over* com as extremidades do cromossomo X, enquanto que o restante do Y não se recombina com segmentos do X, sendo conhecido como região não-recombinante. Fonte: Butler, 2011, *com modificações*.

A região não-recombinante sempre se apresenta no estado haploide e é encontrada somente em pessoas do sexo masculino, sendo integralmente transmitida a todos os indivíduos de uma mesma linhagem patrilínea a menos que ocorram eventos mutacionais ao longo do tempo. Devido ao seu estado haplóide e ao fato de não haver recombinação, as sequências de DNA do cromossomo Y passam a apresentar variações por causa do acúmulo

de mutações de novo e dos rearranjos intracromossômicos ao longo do processo evolutivo (Ali & Hasnian, 2002). Em sua maioria, tais variações são constituídas pela duplicação e acumulação de sequências repetitivas *in tandem*, além da ocorrência de inversões. Estes acontecimentos podem ter moldado a estrutura molecular do cromossomo Y humano ao longo da evolução (Quintana-Murci *et al.*, 2001; Kayser & Sajantila, 2001; Ali & Hasnian, 2002; Gusmão & Carracedo, 2003; Kittler *et al.*, 2003; Butler, 2009).

Como só há uma cópia do cromossomo Y, em comparação com as duas cópias que os cromossomos autossômicos apresentam, o número de cromossomos Y encontrados na população é menor do que o número dos demais cromossomos (Santos *et al.*, 2000). Sendo assim, seu baixo tamanho amostral na população e as práticas culturais específicas de alguns povos, como as culturas em que há a prática da residência patrilocal após o casamento, influenciam a variação da diversidade do cromossomo Y entre os diferentes grupos populacionais (Santos *et al.*, 2000; Kayser *et al.*, 2012). Além disso, a ocorrência de mutações atrelada à falta de recombinação intercromossômica tem como consequência a reorganização seletiva de marcadores, incluindo os polimorfismos, no cromossomo Y. Tais processos contribuem para o aglomeramento geográfico de variantes do Y e o torna uma ferramenta útil para investigações de genética de populações humanas (Santos *et al.*, 2000; Jobling & Tyler, 2000).

O estudo genético do cromossomo Y é importante em diferentes aplicações no que diz respeito à genética humana. Isso inclui a análise de evidências forenses, parentesco por herança paterna, investigações da história das populações, estudo dos movimentos migratórios e pesquisas genealógicas.

Nas últimas décadas, o cromossomo Y vem sendo amplamente empregado no estudo da evolução, migração, vínculos genéticos, dentre outros. A determinação de haplótipos (conjunto de alelos em *loci* adjacentes que são transmitidos integralmente aos descendentes, a menos que ocorram mutações) altamente informativos do cromossomo Y é de extrema importância em estudos forenses e em investigação de parentesco por herança paterna (Quintana-Murci *et al.*, 2001). Existem vantagens e ao mesmo tempo limitações quanto ao seu emprego em análises forenses e de vínculos genéticos. A principal vantagem é que este cromossomo é encontrado somente em homens, o que pode facilitar a solução de crimes, principalmente os casos de estupro, e ainda a identificação de vítimas de sexo masculino em casos de catástrofes, por exemplo, facilitando a identificação de indivíduos através do DNA de homens pertencentes à mesma linhagem patrilínea. Já as limitações se resumem ao fato de que este cromossomo é transmitido integralmente aos indivíduos de mesma linhagem

patrilínea, o que dificulta, principalmente em casos forenses, a diferenciação dos homens de mesma linhagem paterna (Butler, 2011).

Posto isso, há cada vez mais a necessidade do aumento do número de marcadores polimórficos do Y para tais aplicações, especialmente para a individualização de indivíduos dentro da mesma linhagem patrilínea. A porção NRY contém diversos polimorfismos com diferentes taxas de mutação, o que faz com que seja bastante explorada, já que os *loci* desta região podem ser usados em estudos com propósitos diversos, dependendo da característica de tais marcadores.

Marcadores STR do cromossomo Y

O primeiro marcador STR do cromossomo Y foi pioneiramente descrito em 1992 por um grupo de pesquisadores liderado por Lutz Roewer e foi denominado Y-27H39, cuja nomenclatura foi posteriormente estabelecida como DYS19. No decorrer dos anos seguintes, a descoberta de marcadores STR no cromossomo Y sucedeu-se de forma mais lenta do que a descoberta de microssatélites nos autossômicos. Este cenário só começou a mudar no início do século XXI, como uma consequência direta da disponibilidade de informações sobre as sequências de DNA geradas pelo Projeto Genoma Humano e pelo desenvolvimento de novas ferramentas de bioinformática, que possibilitaram em conjunto, a busca por sequências de DNA em bancos de dados. A partir deste momento, sequências do cromossomo Y humano tornaram-se publicamente disponíveis, possibilitando que diversos marcadores STR começassem a ser descobertos e depositados em bancos de dados genômicos (Butler, 2003).

Os marcadores microssatélites do cromossomo Y definem haplótipos dentro dos haplogrupos (grupos de haplótipos que compartilham um ancestral comum), permitindo assim uma maior diferenciação intrapopulacional. As distribuições haplotípicas dos Y-STRs encontradas em diversas populações do mundo têm-se tornado disponíveis através de publicações, ou ainda, por meio de bancos de dados forenses em larga-escala. Dentre estes, o mais utilizado é o banco de dados de haplótipos de referência *online* conhecido como *Haplotype Reference Databases* (YHRD), cujos objetivos principais consistem na produção de informações confiáveis acerca das frequências haplotípicas de Y-STRs e visam disponibilizar dados de populações de diferentes locais ao longo do mundo (www.yhrd.org).

O incremento de tais bancos de dados mostra-se relevante, pois seu uso engloba além das estimativas de frequências de haplótipos para consequentes cálculos de probabilidades, as análises comparativas de populações.

As características do cromossomo Y e de seus marcadores são úteis em diversas situações, como em estudos evolutivos de curta escala de tempo ou em nível microgeográfico e também em investigações forenses e de parentesco, especialmente em situações que o indivíduo, cujo vínculo paterno está sendo investigado, já é falecido. (Kayser *et al.*, 2000; Quintana-Murci *et al.*, 2001; Gusmão & Carracedo, 2003; De Souza Góes *et al.*, 2005; Rodenbusch, 2008; Ballantyne *et al.*, 2010). O emprego destes *loci* microssatélites nestas ocasiões é estritamente dependente do conhecimento e interpretação de suas taxas de mutação e frequências haplotípicas nas diferentes populações. Além disso, eles também podem ser utilizados juntamente com outros conjuntos de *loci* informativos, sejam estes os próprios STRs ou outro tipo de marcador como, por exemplo, os SNPs, com o intuito de obter dados mais completos e consistentes, caso haja tal necessidade (Palha *et al.*, 2012; Bird, 2012).

Tendo isso em vista, em 1997, a Comunidade Forense Européia estabeleceu um conjunto de marcadores Y-STR cujas propriedades mutacionais geravam informações que favoreciam seu emprego em algumas situações que demandavam a identificação de uma pessoa, principalmente em investigações de cunho forense e de linhagem patrilinear. Este grupo de marcadores, o qual ainda é amplamente utilizado pelos laboratórios mundiais, foi denominado “haplótipo mínimo” e é formado pelos *loci* DYS19, DYS389I/II, DYS390, DYS391, DYS392, DYS393 e DYS385a/b, podendo ainda ser acrescido do *locus* YCAIIa/b, que é considerado um marcador opcional, passando assim a ser identificado como “haplótipo estendido” (Buttler, 2003; Gusmão & Carracedo, 2003; Rodenbusch, 2008).

Em 2003, o *U.S Scientific Working Group on DNA Analysis Methods (SWGDM)* sugeriu o uso conjunto dos marcadores do haplótipo mínimo e de dois novos *loci*, denominados DYS438 e DYS439, que, posteriormente, também passaram a ser incorporados em kits comerciais (Buttler, 2003).

Atualmente, a maioria das informações acerca do cromossomo Y ainda é revelada através do uso destes *loci*. Ainda que novos marcadores tenham sido adicionados aos haplótipos dos kits comerciais disponíveis, os *loci* que compõem o haplótipo mínimo, provavelmente, ainda continuarão a ser utilizados por muito tempo.

Devido à necessidade de aumentar a capacidade de discriminação de indivíduos, novos kits comerciais têm surgido, sendo os mais comumente utilizados o PowerPlex®Y (Promega Corporation, Madison, WI, USA) que apresenta 12 *loci* STRs, o Yfiler® (Applied

Biosystems®) que permite a análise de 17 STRs e o Yfiler Plus® (Applied Biosystems®), que engloba 26 STRs. Todos estes kits são formados por STRs do haplótipo mínimo acrescidos de outros STRs do cromossomo Y.

Recentemente, a empresa Promega desenvolveu o kit de amplificação PowerPlex®Y23 (Promega Corporation, Madison, WI, USA). Este kit é composto por 23 *loci* Y-STR, incluindo os marcadores do haplótipo mínimo acrescidos de 14 marcadores: DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS481, DYS533, DYS549, DYS570, DYS576, DYS635 e Y-GATA-H4 (Kovacevic *et al.*, 2013; Calderon *et al.*, 2013; Davis *et al.*, 2013). A maioria dos *loci* que formam o PowerPlex®Y23 apresenta elevada capacidade de discriminação, sendo este um kit muito utilizado na área forense, já que seu alto potencial discriminatório é bastante útil em casos de perfis parciais como os encontrados em cenas de crimes (Thompson *et al.*, 2013).

Marcadores STR com elevada taxa de mutação

Dentre os Y-STRs recentemente descritos, muitos apresentam características adequadas para o emprego na análise forense. As principais peculiaridades que são buscadas em marcadores para seu uso nesta área são o alto grau de polimorfismo, que como consequência vai oferecer uma discriminação mais precisa entre indivíduos, e também sua pouca capacidade de apresentar problemas no momento da sua análise laboratorial, permitindo uma eficiente genotipagem com uma maior facilidade e rapidez (Quintana-Murci *et al.*, 2001; Buttlar, 2003).

A busca por *loci* que apresentem tais particularidades tem contribuído para que muitos estudos sejam realizados e, como consequência, novos potenciais marcadores sejam descobertos. Nesse contexto, recentemente o grupo de Kayser e colaboradores (2012) sugeriu que um conjunto formado por 13 *loci* microssatélites altamente polimórficos, denominados *Rapidly Mutating Y-STRs* (RM-YSTRs), apresentam as características desejáveis para serem aplicados na diferenciação de indivíduos da mesma linhagem. Estes 13 marcadores estão localizados na porção NRY, ao longo dos braços curto e longo do cromossomo Y (Tabela 1), e foram identificados como DYF387S1, DYF399S1, DYF403S1, DYF404S1, DYS449, DYS518, DYS526, DYS547, DYS570, DYS576, DYS612, DYS626 e DYS627. Destes, 9

apresentam-se em cópia única e os outros 4 em múltiplas cópias (Tabela 1). É importante ressaltar que os *loci* de múltiplas cópias ou multialélicos são assim denominados por apresentarem mais de um alelo. Eles são originados pela replicação da mesma região duas ou mais vezes ao longo do cromossomo Y, gerando assim, números variáveis de alelos por *locus*, como pode ser observado nos *loci* DYF387S1, DYF399S1, DYF403S1 e DYF404S1.

Com exceção do *locus* DYS612, que é formado por uma estrutura repetitiva trinucleotídica, todos os demais marcadores que constituem este conjunto são compostos por estruturas repetitivas de 4 pares de bases (Ballantyne *et al.*, 2010; Kayser *et al.*, 2012).

Tabela 1 - Descrição dos marcadores RM-YSTRs

Marcador	Tipo de Repetição	Estrutura Repetitiva	Localização cromossômica	Posição cromossômica
DYF387S1	Tetra, complexo, multi (2)	(AAAG) ₃ (GTAG) ₁ (GAAG) ₄ N ₁₆ (GAAG) _n (AAAG) _o	Yq11.2 Yq11.2	28,0Mb 25,9Mb
DYS570	Tetra, simples	(TTTC) _n	Yq11.2	6,86Mb
DYS576	Tetra, simples	(AAAG) _n	Yq11.2	7,05Mb
DYS518	Tetra, complexo	(AAAG) ₃ (GAAG) ₁ (AAAG) _n (GGAG) ₁ (AAAG) ₄ N ₆ (AAAG) _o	Yq11.221	17,32Mb
DYF399S1	Tetra, complexo, multi (3)	(GAAA) ₃ N ₇₋₈ (GAAA) _n	Yq11.223 Yq11.23 Yq11.2	25,1Mb 26,77Mb 27,2Mb
DYS526a	Tetra, complexo, multi (1+1)	(CCTT) _n	Yq11.2	3,64Mb
DYS526b		(CTTT) _o (CCTT) _p N ₁₁₃ (CCTT) _n		
DYS626	Tetra, complexo	(GAAA) _n N ₂₄ (GAAA) ₃ N ₆	Yq11.223	24,41Mb
DYS627	Tetra, complexo	(AGAG) ₃ (AAAG) _n	Yq11.2	8,65Mb
DYS547	Tetra, complexo	(CCTT) _n T(CTTC) _o N ₅₆ (TTTC) _p N ₁₀ (CCTT) ₄ (TCTC) ₁ (TTTC) _q	Yq11.221	18,87Mb
DYS449	Tetra, complexo	(TTCT) _n N ₅₀ (TTTC) _o	Yq11.2	8,28Mb
DYS612	Tri, complexo	(CCT) ₅ (CTT) ₁ (TCT) ₄ (CCT) ₁ (TCT) _n	Yq11.221	15,75Mb
DYF404S1	Tetra, complexo, multi (2)	(TTTC) _n	Yq11.23 Yq12	25,95Mb 28,0Mb
DYF403S1a	Tetra, complexo, multi (3+1)	(TTCT) _n N ₆ (TTCT) _p	Yq11.2	6,2Mb
DYF403S1b		(TTCT) _n N ₂ (TTCT) _o (TTCC) _p (TTCT) _q	Yq11.2 Yq11.2 Yq11.2	9,65Mb 9,52Mb 6,3Mb

Legenda: Os 13 *loci* se encontram ao longo dos braços curtos e longos do cromossomo Y na região NRY e possuem diferentes estruturas repetitivas. Fonte: Kayser *et al.*, 2012 (A), com modificações.

Os RM-YSTRs apresentam taxas de mutação na ordem de 10^{-2} , superiores em cerca de 10 vezes às taxas de outros YSTRs amplamente utilizados, por este motivo, tais marcadores

apresentam grande capacidade de discriminação, sendo potencialmente capazes de diferenciar indivíduos proximamente relacionados, pertencentes à mesma linhagem patrilinea. Tal perspectiva pode conferir uma vantagem no uso desses *loci* em genética forense, pois se sabe que a maioria dos crimes sexuais é realizada por indivíduos da mesma família da vítima. Eles ainda poderão ser muito úteis em casos de acidentes em massa, possibilitando a identificação de pessoas desaparecidas através da comparação de amostras encontradas no local do acidente com parentes do sexo masculino pertencentes à mesma linhagem paterna. Além disso, devido ao seu potencial elevado de assegurar individualização dentro do mesmo haplogrupo, os RM-YSTRs também podem ser empregados em estudos populacionais que abordam cenários de grupos que tiveram uma diminuição do seu número amostral seguido de uma rápida expansão em sua história recente, ou ainda em casos nos quais há práticas culturais específicas da população, em que os conjuntos de Y-STRs empregados atualmente apresentam uma resolução limitada acerca da diversidade, devido ao pequeno tamanho amostral de cromossomos Y na população em questão (Ballantyne *et al.*, 2010; Kayser *et al.*, 2012).

Os *loci* RM-YSTRs não têm seu uso restrito a tais situações, podendo também ser aplicados com outros objetivos, como por exemplo, no estudo de movimentos migratórios. Ao combinar o uso de microssatélites do cromossomo Y de elevada taxa de mutação com marcadores binários mais estáveis, como os Y-SNPs, podem ser obtidas informações relevantes acerca da estrutura populacional local e continental, corroborando ou não, hipóteses de migrações relativamente recentes na história de uma população (Quintana-Murci *et al.*, 2001; Ballantyne *et al.*, 2010).

Posto isso, verifica-se a necessidade de se aprofundar o conhecimento acerca dos RM-YSTRs e de se analisar suas características em diferentes populações do mundo, pois têm demonstrado serem ferramentas importantes em investigações de parentesco e na área forense. É relevante ressaltar que ainda não há estudos com os marcadores RM-YSTRs na população brasileira, sendo este, portanto, um trabalho pioneiro que visa proporcionar dados inovadores acerca das características mutacionais e populacionais dos *loci* de elevada taxa de mutação e do seu poder de discriminação de indivíduos na população brasileira, que é mundialmente conhecida por seu grau de miscigenação.

1. OBJETIVOS

1.1 Objetivo geral

Estudar a variabilidade e a distribuição populacional de marcadores RM-YSTRs em indivíduos não relacionados geneticamente, do sexo masculino, pertencentes à população do Rio de Janeiro, com a finalidade de gerar informações populacionais e de conferir maior eficiência ao emprego do cromossomo Y em genética forense e em estudos de vínculos genéticos.

1.2 Objetivos específicos

- a) Estimar a frequência dos haplótipos em indivíduos do sexo masculino;
- b) Verificar a ocorrência de mutações entre indivíduos da mesma linhagem patrilínea;
- c) Estimar a frequência das mutações;
- d) Contribuir com informações para o banco de dados de Y-STRs.

2. MATERIAIS E MÉTODOS

Este estudo fez parte de um trabalho de colaboração entre diversos laboratórios mundiais e gerou uma publicação (Anexo A). 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). Durante este trabalho, também foi realizada uma colaboração com outro projeto desenvolvido no LDD-UERJ, gerando uma publicação (Anexo B).

2.1 Seleção da amostra populacional

Foram analisados 258 indivíduos do sexo masculino, constituindo 129 pares de pais e filhos, nascidos no estado do Rio de Janeiro. A amostra populacional foi selecionada a partir de indivíduos que participaram de perícias realizadas no Laboratório de Diagnósticos por DNA da UERJ nos anos de 2011 e 2012, cujos vínculos genéticos de paternidade questionados foram confirmados através da análise de, pelo menos, 15 marcadores STRs autossômicos, com probabilidades de certeza superiores a 99,999%. Os participantes permitiram a utilização de suas amostras biológicas em projetos de pesquisa ao assinar um termo de consentimento livre e esclarecido no ato da coleta da amostra biológica. Além disso, a utilização destas amostras biológicas em projetos de pesquisa foi aprovada pelo Comitê de Ética em Pesquisa do Hospital Pedro Ernesto, parecer CAAE: 0067.0.228.000-09 (Anexo C).

As confirmações dos vínculos de paternidade foram realizadas com a participação das genitoras dos supostos filhos, sendo, portanto, possível definir as heranças materna e paterna de seus filhos, exceto em *loci* onde, eventualmente, estes apresentavam genótipos heterozigotos coincidentes com os de suas genitoras.

As amostras biológicas consistiam em 1 a 3 mL de sangue periférico de cada indivíduo, coletadas em tubos Vacutainer de 5mL contendo EDTA, sendo estas armazenadas a 4°C, até a extração do DNA.

2.2 Extração de DNA genômico por Chelex

A extração de DNA a partir de sangue periférico foi realizada, de acordo com protocolos vigentes na rotina do LDD-UERJ, através do método de Chelex (Walsh *et al.*, 1991).

Para isto, eram adicionados 10 μ L de sangue periférico em um tubo de plástico do tipo eppendorf, de 1,5mL, que continha 1mL de água deionizada (MilliQ) autoclavada. Em seguida, a mistura era homogeneizada e incubada, por 30 minutos, à temperatura ambiente. O tubo era, então, centrifugado a 1200 x g por 5 minutos. Após a centrifugação, descartava-se o sobrenadante e se adicionava 0,5mL de água MilliQ autoclavada ao precipitado, que era homogeneizado e deixado à temperatura ambiente por 30 minutos. Posteriormente, esta preparação era centrifugada nas condições acima citadas. Era descartado o sobrenadante e ao precipitado eram adicionados, sob agitação manual, 300 μ L de solução de Chelex-100 a 5% em água deionizada. A mistura era incubada a 56°C, por 20 minutos e, em seguida a 95°C, por 8 minutos. Após a adição de 1mL de tampão TE-4 (10mM de Tris-Cl pH7.6, 1mM de EDTA pH 8.0) à mistura incubada, a preparação era centrifugada a 1200 x g, por 3 minutos, sendo a solução final, que contém o DNA, armazenada a 4°C.

2.3 Análise Molecular do Cromossomo Y

2.3.1 Amplificação de *loci* hipervariáveis por sistemas multiplex

As sequências gênicas de interesse do cromossomo Y eram amplificadas pela técnica de PCR (*Polimerase Chain Reaction* ou reação em cadeia da polimerase) através da realização de três reações multiplex. Os conjuntos de iniciadores (*primers*) e as condições para a amplificação das sequências de interesse foram descritos por Kayser e colaboradores em 2012 (Kayser *et al.*, 2012).

As reações de PCR multiplex foram padronizadas em termocicladores automáticos, modelo 9600 (*Life Technologies*, Co.). Os multiplexes foram nomeados de RM1, RM2 e

RM3, em referência ao fato de os marcadores em estudo serem designados de “*Rapidly Mutating*”.

2.3.1.1 Conjuntos de *primers* utilizados nas reações multiplex

Para a análise de cada marcador genético, foram usados conjuntos de *primers forward* (F) e *reverse* (R) para a amplificação da sequência nucleotídica de cada *locus*. Os *primers* F de cada marcador foram confeccionados com dNTPs marcados com fluoróforos para que o produto da amplificação pudesse ser analisado conforme protocolo descrito por Kayser e colaboradores (Kayser *et al.*, 2012), com modificações. Os fluoróforos (*Life Technologies*) utilizados neste estudo são rotineiramente empregados pelo laboratório LDD-UERJ e encontravam-se em conformidade com a proposta do trabalho de Kayser *et al.*, 2012. Na Tabela 2, são descritas as sequências no sentido 5'→3' de cada conjunto de *primers*, bem como o fluoróforo utilizado para a marcação do *primer* F de cada um dos conjuntos. Foram utilizados os fluoróforos Vic[®], NED[™] e 6-FAM[™], que emitem fluorescências nas cores verde, amarelo e azul, respectivamente.

Tabela 2 - Loci RM-YSTR com os respectivos pares de primers usados para amplificação por PCR e os fluoróforos empregados

<i>Locus</i>	<i>Seqüência dos primers 5'-3'</i>	<i>Fluoróforos</i>
DYF387S1	F - GCCTGGGTGACAGAGCTAGA	Vic [®]
	R - TCACACTTCTCACACTGTGGC	
DYS570	F - GAACTGTCTACAATGGCTCACG	Ned TM
	R - TGTCTGGTTTCTTGACTATGCTGA	
DYS576	F - GTTGGGCTGAGGAGTTCAATC	6 - Fam TM
	R - CTCCAGGAAATGAGACTGCC	
DYF399S1	F - GGGTTTTCCACCAGTTTGCAT	6 - Fam TM
	R - AGGAATGTCCCAAACATGG	
DYS518	F - GGCAACACAAGTGAAACTGC	Ned TM
	R - ATCACCCATGGTAAGAGCTGA	
DYS526a/b	F - TCTGGTGAAGTATCCAAACC	6 - Fam TM
	R - ACCTTCTGGCGAAGTAACCC	
DYS626	F - GCAAGACCCCATAGCAAAAG	6 - Fam TM
	R - AAACATGTCCCAAATTTCTTCTT	
DYS627	F - CTAGGTGACAGCGCAGGATT	Vic [®]
	R - CTTGCCATTTGCTCATTATCC	
DYF403S1a/b	F - CAAAATTCATGTGGATAATGA	Vic [®]
	R - TAGATGGAATCCTGCTCTGT	
DYF404S1	F - GGCTTAAGAAATTTCAACGCATA	Vic [®]
	R - CTGCAATTGTTCCATCATGG	
DYS449	F - TGAATATTTCCCTTAAGTGTGTGA	6 - Fam TM
	R - GTCCAACCTAGAGTGCAATGG	
DYS547	F - TCCATGTTACTGCAAAATACAC	6 - Fam TM
	R - GACACGTTTATGCTCTGTCA	
DYS612	F - CCCCATGCCAGTAAGAATA	6 - Fam TM
	R - TTTTCTTTGCCTTCCCTCAC	

Para a amplificação de cada marcador, os *primers forward* (F) e *reverse* (R), com concentrações iniciais de 100µM, foram diluídos e misturados em um único tubo, originando uma solução de *primers* F+R a uma concentração final de 10µM. Posteriormente, soluções contendo o par de *primers* de cada marcador foram reunidas, conforme mostrado na Tabela 3, originando os multiplexes RM1, RM2 e RM3.

Tabela 3 - Soluções de primers utilizadas nas reações multiplexes

Mix de <i>primers</i>	<i>Primers F+R</i> (10µM)	Volume para 1 reação de PCR
RM1	DYF387S1	0,3µL
	DYS570	0,2µL
	DYS576	0,5µL
	DYF399S1	0,4µL
	H ₂ O MilliQ	2,6µL
	Volume final do mix	4µL
RM2	DYS518	0,2µL
	DYS526a/b	0,2µL
	DYS626	0,2µL
	DYS627	0,4µL
	H ₂ O MilliQ	3,0µL
	Volume final do mix	4µL
RM3	DYF403S1a/b	0,6µL
	DYF404S1	0,2µL
	DYS449	0,2µL
	DYS547	0,4µL
	DYS612	0,2µL
	H ₂ O MilliQ	2,4µL
	Volume final do mix	4µL

2.3.1.2 Reações de amplificação multiplex

As preparações das reações PCR multiplex eram realizadas em uma capela esterilizada por luz ultravioleta de comprimento de onda de 254nm. Na amplificação por PCR dos marcadores RM-YSTR, foram utilizados tampão QIAGEN[®] multiplex PCR master mix 2x, proveniente do QIAGEN[®] Multiplex PCR kit (HotStar Taq[®] DNA Polimerase, Multiplex PCR Buffer pH 8.7, dNTP mix), mix dos *primers* na concentração de 10µM e DNA genômico, originando uma solução com um volume final de 10µL para cada reação de PCR.

Na realização de cada uma das três PCR multiplex, juntamente com as alíquotas de DNA dos indivíduos da amostra populacional estudada, também era amplificado o DNA da

linhagem celular masculina 9948 (Promega[®]), como controle positivo das reações. Como controle negativo, era preparada uma reação sem o acréscimo de DNA.

As reações de PCR multiplex eram procedidas em termociclador automático modelo 9600 (*Life Technologies*). As condições termocíclicas seguiram o protocolo descrito por Kayser *et al.*, 2012 e são apresentadas na Tabela 4.

Tabela 4 - Condições termocíclicas utilizadas para amplificações dos multiplex RM1, RM2 e RM3

Temperatura (°C)	Tempo	Ciclos
94	10 min	
94	30s	10
65-1/ciclo	30s	
72	60s	
94	30s	25
50	30s	
72	60s	
60	45 min	

2.3.2 Eletroforese capilar dos produtos da amplificação

Após serem submetidos à etapa da PCR, os fragmentos gerados por amplificação eram separados através da eletroforese capilar no sequenciador automático ABI PRISM 3500 (*Life Technologies*), gerando eletroferogramas com os alelos que definiram os haplótipos de cada indivíduo.

2.3.2.1 Preparação dos produtos da PCR

Anteriormente à eletroforese capilar, uma alíquota dos produtos de amplificação era preparada para serem aplicadas na placa que seria colocada no sequenciador automático.

Para isto, era feita uma mistura contendo 9,8 μ L de formamida (HI-DI *Formamide*, *Life Technologies*) e 0,2 μ L de padrão interno de tamanho LIZ 500 *Size Standard* (*Life Technologies*), da qual o volume de 9,0 μ L era colocado em cada poço da placa do sequenciador. À essa mistura eram, então, acrescentados 1 μ L do produto de amplificação de cada amostra de DNA. Em seguida, a placa era mantida à temperatura de 95°C por 5 minutos, visando à desnaturação dos produtos de amplificação, sendo posteriormente mantida em banho de gelo, por 5 minutos, ou até o início da eletroforese capilar. Por fim, a placa era levada ao sequenciador automático para a realização da eletroforese.

2.3.2.2 Eletroforese e detecção dos produtos de amplificação

Para a eletroforese dos produtos amplificados foram utilizados capilares de 50cm de comprimento (*Life Technologies*) preenchidos com o polímero 3500 POP-7™ (*Performance Optimized Polymer - Life Technologies*). Durante a eletroforese, as amostras eram injetadas no capilar eletrocineticamente devido a uma diferença de potencial nas extremidades dos capilares, sendo o polímero automaticamente substituído a cada nova injeção.

O padrão de migração dos produtos de amplificação ao longo dos capilares é dependente de seus tamanhos, com os menores migrando mais rapidamente que os maiores. A detecção dos produtos amplificados é realizada quando estes passam por uma janela existente no capilar, que se encontra posicionada numa região sobre a qual será incidido um *laser*. Com isto, de acordo com as características de cada fluoróforo, os produtos de amplificação emitirão fluorescência em diferentes comprimentos de onda, a fluorescência emitida é captada e convertida em uma câmara do tipo CCD (*charge-couple device*) que identifica essas emissões fluorescentes.

Por fim, os resultados obtidos são representados por um eletroferograma, no qual se observam picos resultantes do processo de emissão, detecção e transformação do sinal

luminoso emitido pelos fluoróforos, que é visualizado em unidades relativas de fluorescência (RFUs).

2.3.2.3 Análise dos produtos de amplificação

A análise dos produtos de amplificação e a tipagem dos alelos foram realizadas através do programa GeneMapper versão 3.5 (*Life Technologies*).

2.3.3 Confirmação de mutações

Após a genotipagem dos indivíduos, os duos de pais e filhos nos quais se observaram alelos mutados foram selecionados para a confirmação de mutações nos marcadores RM-YSTR. Para isto, foi seguido o procedimento padrão indicado na literatura mundial, que consiste na repetição da tipagem das amostras nas quais foram constatadas mutações.

Adicionalmente a tal procedimento, no presente trabalho optou-se por selecionar algumas das amostras que apresentaram mutações e confirmar os alelos alterados através da técnica de sequenciamento. Desta maneira, foi necessário realizar a amplificação dos pares de alelos mutados por meio de uma PCR monoplex antes de prosseguir para a etapa de sequenciamento, de modo que somente os produtos mutados presentes em um dado *locus* fossem amplificados e sequenciados. Devido ao fato de alguns produtos de amplificação apresentarem características estruturais complexas e mais de um alelo, os produtos de amplificação de alguns *loci* foram excluídos desta etapa. Optou-se por confirmar as mutações por meio do sequenciamento dos alelos pertencentes a *loci* monoalélicos.

2.3.3.1 Amplificação através de PCR monoplex

Os alelos mutados foram amplificados pela técnica de PCR monoplex. Os conjuntos de iniciadores (*primers*) e as condições utilizadas para a amplificação das sequências de interesse foram descritas anteriormente neste trabalho (Tabelas 2, 3 e 4).

Faz-se necessário especificar que na reação monoplex utilizou-se tampão QIAGEN® multiplex PCR master mix 2x, proveniente do QIAGEN® Multiplex PCR kit (HotStar Taq® DNA Polimerase, Multiplex PCR Buffer pH 8.7, dNTP mix), o par de iniciadores (*forward* e *reverse*) na concentração de 10µM e DNA genômico, originando uma solução com um volume final de 20µL para cada reação de PCR.

2.3.3.2 Eletroforese em gel de agarose

Para a avaliação inicial dos produtos da PCR monoplex, era realizada eletroforese em gel de agarose a 2%. Para o preparo das amostras, foram adicionados 3 µL do produto de amplificação a 2 µL de *loading buffer* (10mM Tris-Cl pH 8,0, 0,1% azul de bromofenol, 0,1% xileno cianol FF e 5% Ficoll 400). Após o seu preparo, as amostras eram aplicadas no gel e submetidas à eletroforese em uma cuba horizontal (Horizon 58 – Life Technologies), a 110V, por 30 minutos, em tampão de corrida TAE 1x (40mM Tris-base, 20mM ácido acético, 1mM EDTA). Também foi aplicado no gel um marcador de tamanho molecular (Life Technologies), que consiste numa mistura de fragmentos de DNA com comprimentos entre 50 a 800 pb, permitindo inferir, por meio de comparação, os tamanhos dos produtos de amplificação.

Em seguida, o gel era colocado em um transiluminador de luz ultravioleta (modelo FMBIO II Multi View – Hitachi), tornando possível a visualização de cada produto amplificado e a genotipagem das amostras.

2.3.3.3 Sequenciamento

Após a eletroforese, os produtos de amplificação eram purificados em colunas Microspin HRS-300 Columns (*GE HEALTHCARE*), que eram centrifugadas a 5000 x g, por 3 minutos.

Aos produtos de PCR purificados eram adicionados 2 μ L de água MiliQ, 2 μ L de tampão Big Dye Terminator (*Life Technologies*), 2 μ L de Big Dye (*Life Technologies*) e 1 μ L de primer R, na concentração de 10 μ M, não marcado com fluoróforo. As reações eram desenvolvidas no termociclador modelo 9600 (*Life Technologies*), sob as condições descritas na Tabela 5.

Tabela 5 - Condições termocíclicas utilizadas para as reações de sequenciamento dos produtos purificados

Temperatura (°C)	Tempo	Ciclos
96	15s	25
50	1s	
60	2min	

Após a reação de sequenciamento, as amostras eram purificadas em colunas com resina Sephadex G-50, a 6,7% (Sambrook *et. al.*, 1989). Para formação da malha de Sephadex, eram aplicados 750 μ L da solução em uma coluna montada em um tubo de 1,5mL. A preparação era centrifugada a 5000 x g, por 3 minutos. O filtrado era descartado e a coluna centrifugada novamente. O volume total da reação de sequenciamento era adicionado à coluna Sephadex, que era mais uma vez centrifugada a 5000 x g, durante 3 minutos.

Aos produtos de sequenciamento purificados eram acrescentados 9 μ L de solução formada por 8,8 μ L de formamida (Hi-Di Formamide, *Life Technologies*) e 0,2 μ L de Liz-500. Da mistura, 10 μ L eram aplicados em cada poço da placa do sequenciador. Os produtos da reação de sequenciamento eram desnaturados a 95°C, por 5 minutos, mantidos em gelo, por 5 minutos, e submetidos à eletroforese capilar em sequenciador automático ABI Prism 3500, com polímero POP-7™ (*Life Technologies*), e analisadas com o *software Sequencing Analysis 3.7* (*Life Technologies*).

2.4 Análises estatísticas

Foram realizadas estimativas das frequências dos alelos de cada um dos 13 *loci* RM-YSTR e dos haplótipos verificados, assim como da diversidade gênica de cada *locus* e da diversidade haplotípica desse conjunto de marcadores. Foram também estimados o Poder de Discriminação de haplótipos caracterizados por esse conjunto de marcadores na amostra estudada, os valores de F_{st} de distância genética e a Variância Molecular (AMOVA), sendo estes dois últimos parâmetros aferidos através do uso do *software* Arlequin versão 3.05 (Excoffier *et al.*, 2010).

2.4.1 Estimativas das frequências alélicas e haplotípicas

A estimativa da frequência de um alelo em uma amostra populacional expressa o número de vezes que ele é observado em relação ao número total de alelos verificados em um dado *locus* cromossômico. O mesmo pode ser dito da frequência dos haplótipos encontrados numa determinada amostra. Para as estimativas das frequências alélicas e haplotípicas foi utilizada a fórmula:

$$F_{ij} = N_{ij}/N_j$$

Onde:

F_{ij} = Frequência do alelo ou haplótipo i na população j .

N_{ij} = Número de vezes que o alelo ou haplótipo i é observado na população j .

N_j = Número total de alelos ou haplótipos na amostra representativa da população j .

2.4.2 Cálculo da diversidade gênica por *locus* e da diversidade haplotípica

O parâmetro da diversidade gênica por *locus* determina o poder de discriminação de um determinado *locus*. O mesmo se aplica à diversidade haplotípica, que mede o poder de discriminação de um determinado conjunto de marcadores em uma amostra populacional. Ambos os parâmetros foram calculados através da fórmula, representada a seguir:

$$D = (n/n-1)(1-\sum p_i^2)$$

Onde:

n = tamanho da amostra

p_i = frequência alélica ou haplotípica

2.4.3 Poder de discriminação de haplótipos

O poder de discriminação de haplótipos do cromossomo Y determina a probabilidade de dois indivíduos escolhidos ao acaso em uma população apresentarem haplótipos diferentes. Este índice foi calculado pela proporção entre o número de indivíduos com haplótipos únicos e o número total de indivíduos analisados.

2.4.4 AMOVA

A metodologia de análise da variância de dados moleculares, também chamada de AMOVA (Excoffier *et al.*, 1992), consiste na estimativa da diferenciação populacional baseada em informações geradas por perfis moleculares.

No que diz respeito ao cromossomo Y, considera-se o número de unidades de repetições dos *loci* STR analisados dentro de uma população ou entre populações distintas para o cálculo da variação populacional (YHDR.org). A estimativa da diferenciação populacional baseada em haplótipos do cromossomo Y expressa o quão divergente é o haplótipo de um indivíduo em comparação com outro haplótipo qualquer.

Neste estudo, os cálculos estatísticos relacionados à AMOVA foram realizados através do uso do *software* Arlequin 3.5 (Schneider *et al.*, 2000).

2.4.5 Parâmetros de distância genética

A diferenciação genética pode ser caracterizada por alterações na distribuição das frequências alélicas das populações ao longo das gerações, devido, principalmente, a eventos como mutação, migração, deriva genética e seleção natural. Tais eventos são capazes de modificar as probabilidades de sobrevivência e reprodução de indivíduos da população, promovendo, conseqüentemente, a fixação ou extinção de alelos no decorrer do tempo. Como esses eventos não ocorrem de maneira idêntica em todas as populações no mesmo período de tempo verifica-se o estabelecimento de diferentes características genéticas nos diversos grupos populacionais. Alguns parâmetros estatísticos são utilizados para estimar a divergência gênica, também conhecida como distância genética, existente dentro de populações e entre pares de populações.

Um parâmetro populacional específico bastante utilizado é o índice estatístico F de Wright ou F_{st} , que indica uma estimativa da diferenciação genética entre subpopulações e, conseqüentemente, indica também o grau de isolamento das mesmas, visto que em uma população isolada há pouco fluxo gênico e pouca divergência genética. Os valores de F_{st} variam de 0 (não há diferenças genéticas entre as populações, que compartilham igualmente as frequências alélicas) a 1 (há diferenças genéticas entre as populações causadas pela fixação de alelos diferentes por cada uma).

Para o estudo da distância genética foram utilizadas as populações representadas na Tabela 6. Os cálculos foram realizados através do *software* Arlequin 3.5.

Tabela 6 - Populações utilizadas para análise de distância genética, seus respectivos tamanhos e referências bibliográficas

Grupo	População	Número de indivíduos	Referência Bibliográfica
Miscigenados	Rio de Janeiro (Brasil)	129	Dados deste trabalho
	Buenos Aires (Argentina)	97	Ballantyne, <i>et. al.</i> , 2014
	Costa Rica	102	
África Sub-Saariana	Bakangos (Angola)	25	
	Guinea-Bissau	92	
	Xhosa (Sul da África)	116	
Europa	Grã-Bretanha	99	
	Marches (Itália)	100	
	Holanda	112	
Ásia	Kanto (Japão)	147	
	Timor Leste	110	
	Bai (China)	110	

2.4.6 Análise mutacional de loci RM-YSTR

Para a análise das características mutacionais dos 13 *loci* RM-YSTR foram selecionados 258 indivíduos do sexo masculino, constituindo 129 pares de pais e filhos.

Calculou-se a taxa de mutação por *locus* e a taxa de mutação média dessa amostra. A taxa de mutação por *locus* é definida pela proporção entre o número de mutações ocorridas no *locus* e o número de meioses. Enquanto que a taxa de mutação média é estabelecida através da razão entre o somatório das mutações ocorridas em todos os duos pai/filho e o total de meioses.

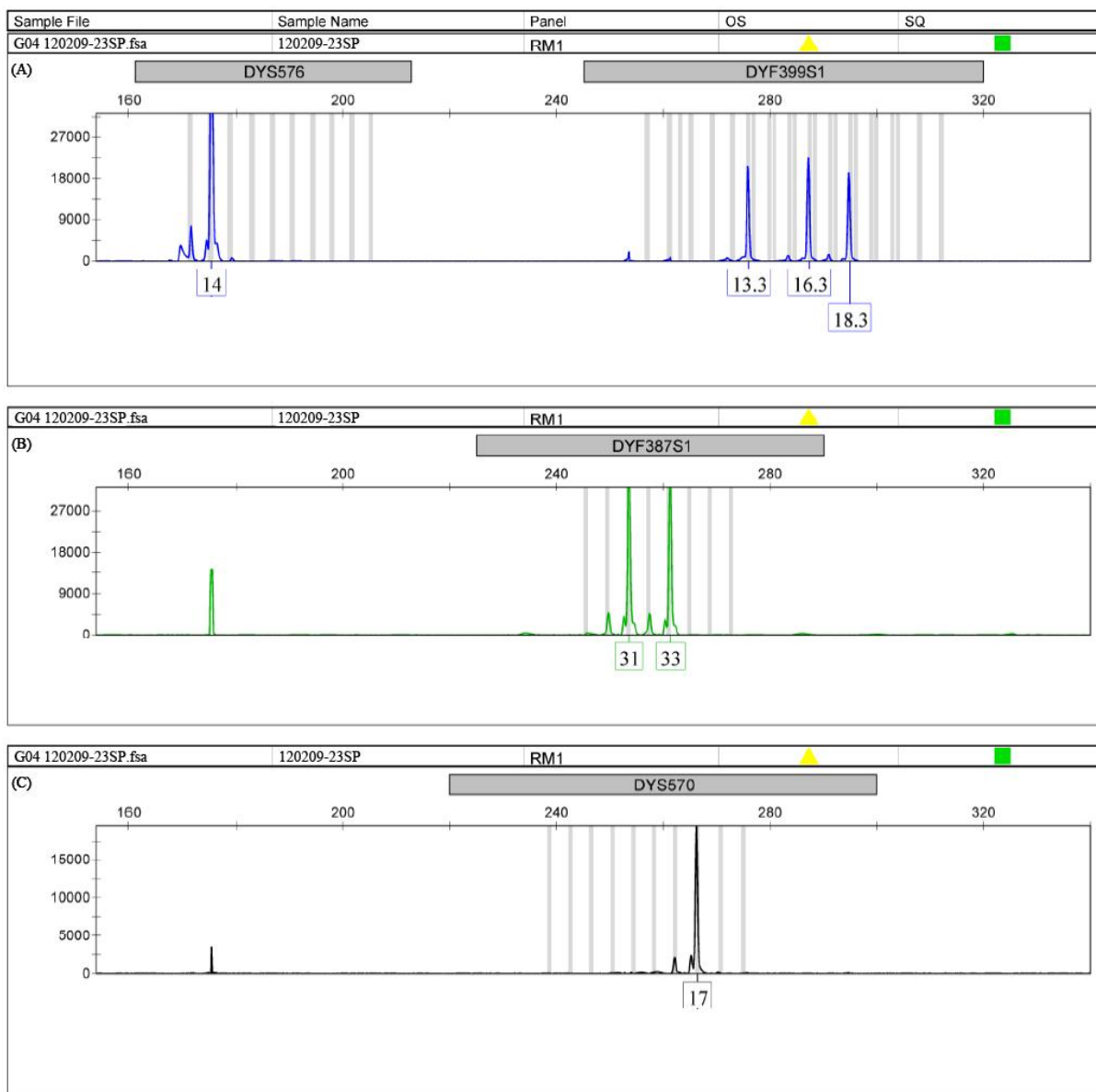
3. RESULTADOS

3.1 Haplótipos do cromossomo Y em indivíduos miscigenados do Rio de Janeiro

Neste estudo, foram analisados 258 indivíduos do sexo masculino nascidos no estado do Rio de Janeiro, constituindo 129 pares de pais e filhos participantes de investigações de paternidade, conforme descrito no item 2.1 de Materiais e Métodos. Os resultados gerados fazem parte de um trabalho de colaboração entre diversos laboratórios mundiais e gerou uma publicação (Anexo A). A metodologia de análise dos 13 *loci* RM-YSTR foi estabelecida com eficácia e consistiu na amplificação por PCR em três reações multiplex (RM1, RM2 e RM3) de alíquotas de DNA de cada indivíduo, na detecção dos produtos de amplificação gerados (alelos) por eletroforese capilar no sequenciador automático ABI PRISM 3500 (*Life Technologies*) e sua posterior genotipagem decorrente da interpretação de eletroferogramas dos alelos com auxílio do software GeneMapper v4.1.

As figuras 3, 4 e 5 apresentam eletroferogramas representativos dos resultados típicos da tipagem de um indivíduo através das PCRs multiplex RM1, RM2 e RM3, respectivamente, reunindo os 13 *loci* RM-YSTR analisados. Os pares de *primers* utilizados nas reações multiplex destes *loci* foram marcados com um dos três fluoróforos: 6-FAM, VIC e NED. Nos eletroferogramas a seguir, o eixo das abscissas (X) representa o tamanho dos produtos de amplificação em pb e os picos representam o número de repetições *in tandem* dos alelos dos *loci* microssatélites, nomeados de acordo com a nomenclatura internacional (Gusmao *et al.*, 2006; Kayser & Ballantyne, 2014; Mulero *et al.*, 2014). O eixo das ordenadas (Y) representa a intensidade relativa de fluorescência (RFU – *relative fluorescence unit*) dos produtos de amplificação gerados para cada *locus*.

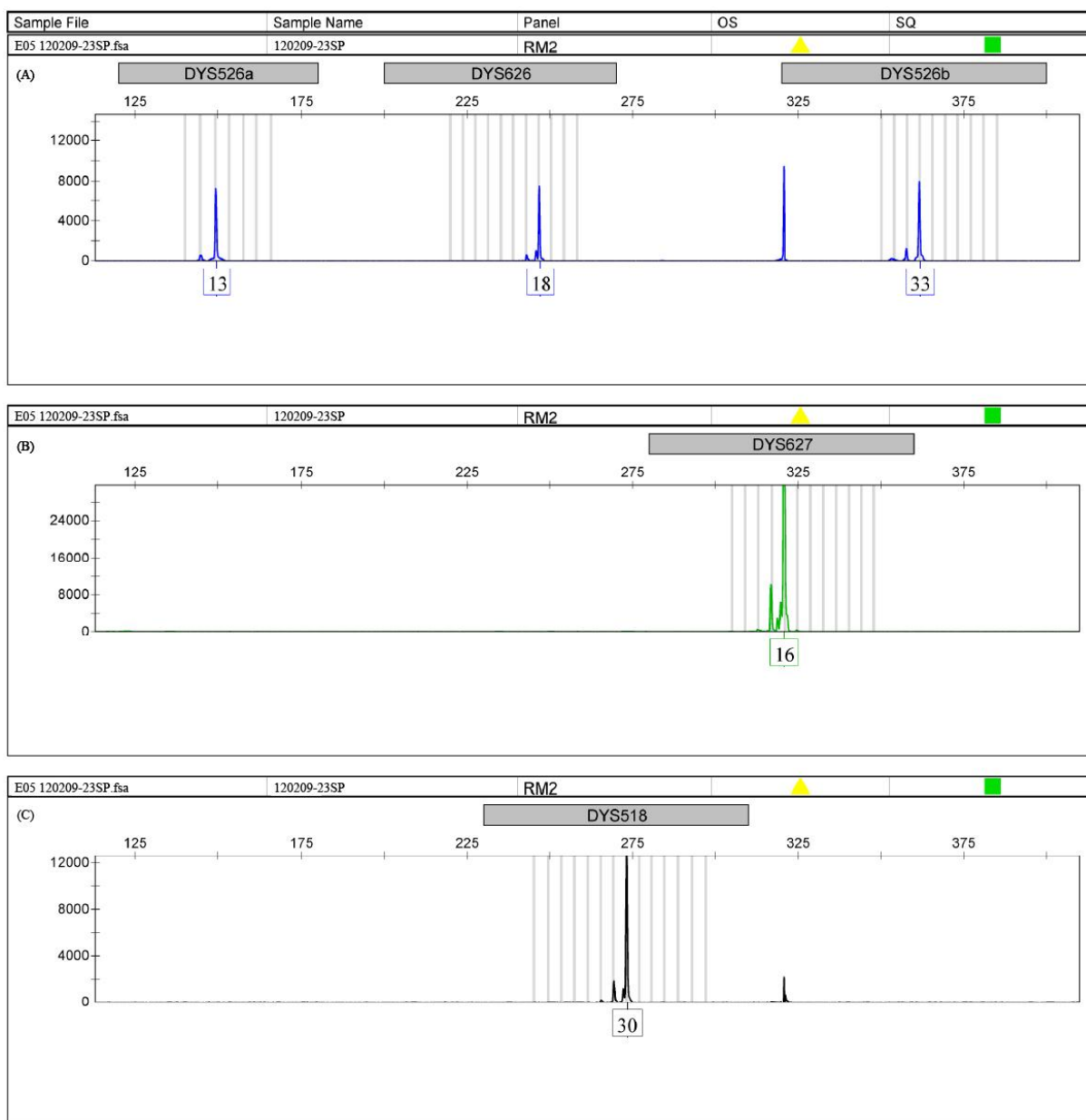
Figura 3 - Perfil genético do indivíduo 120209-23SP da população do Rio de Janeiro obtido através da tipagem de quatro marcadores RM-YSTR amplificados pelo multiplex RM1



Legenda: O multiplex RM1 reúne os marcadores genéticos DYS576, DYF399S1, DYF387S1 e DYS570. Uma alíquota de DNA foi amplificada por PCR multiplex do conjunto de marcadores. Os produtos de amplificação foram separados por eletroforese capilar no sequenciador automático ABI PRISM 3500, os alelos foram detectados pela fluorescência emitida e interpretados no software GeneMapper ID v4.1. O eixo X representa o tamanho dos fragmentos em pb, o eixo Y a intensidade de fluorescência (RFUs) e os picos representam o número de repetições *in tandem* do alelo. (A) *primers* marcados com 6-FAM; (B) *primers* marcados com VIC; (C) *primers* marcados com NED.

Através da análise do eletroferograma mostrado na figura 3, é possível verificar que a genotipagem dos *loci* DYS576 e DYS570 mostrou um único alelo por *locus* enquanto que nos *loci* DYF399S1 e DYF387S1 foram observados 3 e 2 alelos, respectivamente.

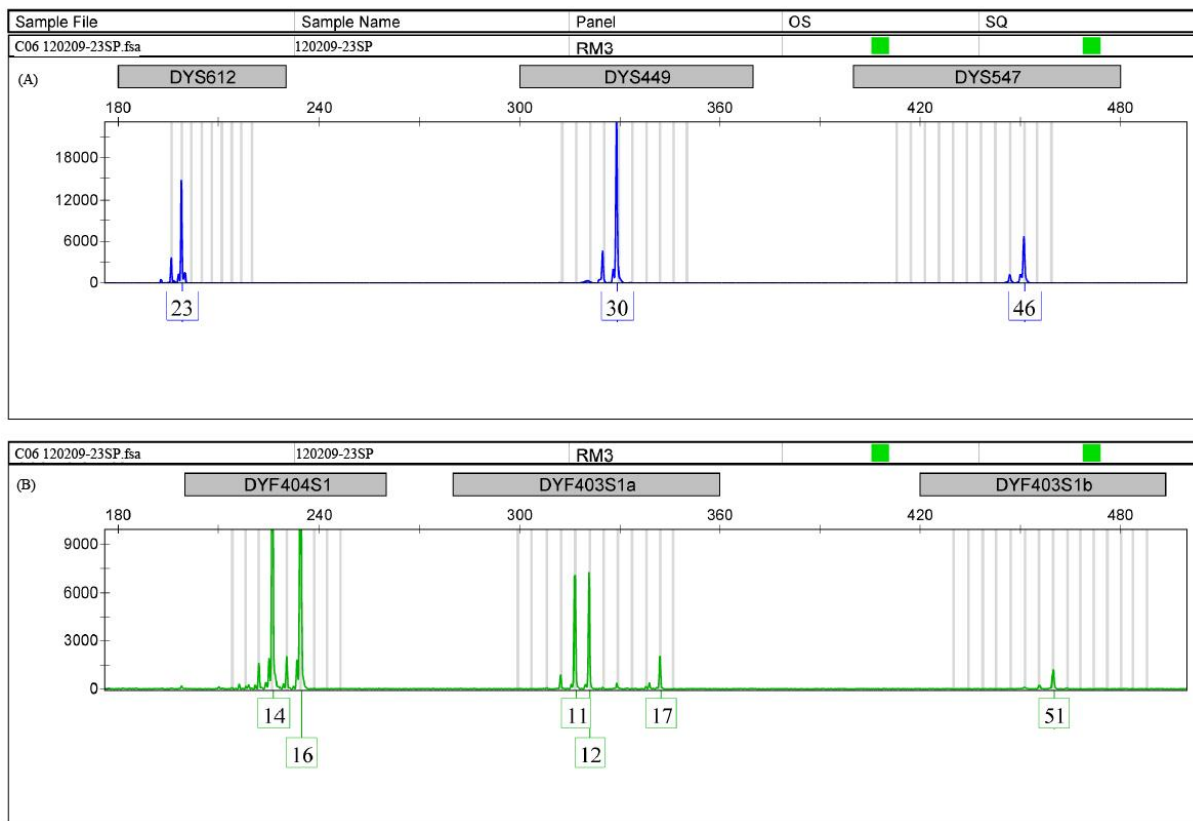
Figura 4 - Perfil genético do indivíduo 120209-23SP da população do Rio de Janeiro obtido através da tipagem de quatro marcadores RM-YSTR amplificados pelo multiplex RM2



Legenda: O multiplex RM2 reúne os marcadores genéticos **DYS526a/b**, **DYS626**, **DYS627** e **DYS518**. Uma alíquota de DNA foi amplificada por PCR multiplex do conjunto de marcadores. Os produtos de amplificação foram separados por eletroforese capilar no sequenciador automático ABI PRISM 3500, os alelos foram detectados pela fluorescência emitida e interpretados no software GeneMapper ID v4.1. O eixo X representa o tamanho dos fragmentos em pb, o eixo Y a intensidade de fluorescência (RFUs) e os picos representam o número de repetições *in tandem* do alelo. (A) *primers* marcados com 6-FAM; (B) *primers* marcados com VIC; (C) *primers* marcados com NED.

Através da análise da figura 4, é possível verificar que em todos os *loci* amplificados pelo multiplex RM2 o indivíduo genotipado apresenta um único alelo.

Figura 5 - Perfil genético do indivíduo 120209-23SP da população do Rio de Janeiro obtido através da tipagem de cinco marcadores RM-YSTR amplificados pelo multiplex RM3



Legenda: O multiplex RM3 reúne os marcadores genéticos DYS612, DYS449, DYS547, DYF404S1 e DYF403a/b. Uma alíquota de DNA foi amplificada por PCR multiplex do conjunto de marcadores. Os produtos de amplificação foram separados por eletroforese capilar no sequenciador automático ABI PRISM 3500, os alelos foram detectados pela fluorescência emitida e interpretados no software GeneMapper ID v4.1. O eixo X representa o tamanho dos fragmentos em pb, o eixo Y a intensidade de fluorescência (RFUs) e os picos representam o número de repetições *in tandem* do alelo. (A) *primers* marcados com 6-FAM; (B) *primers* marcados com VIC.

Através da análise do eletroferograma mostrado na figura 5 é possível verificar que nos *loci* DYS612, DYS449 e DYS547 o indivíduo genotipado apresenta um único alelo enquanto que no *locus* DYF404S1 foram observados 2 alelos. No *locus* DYF403S1a/b foram observados 4 alelos. Este último *locus* se apresenta subdividido em DYF403S1a e DYF403S1b no eletroferograma porque este *locus* possui suas unidades repetitivas separadas por uma sequência de nucleotídeos.

Os alelos, em números de unidades repetitivas, tipados para cada marcador RM-YSTR, representando os haplótipos dos cromossomos Y dos pais, que constituem a amostra

utilizada no estudo populacional, são mostrados na Tabela 7. A ordem dos *loci* na tabela não é a mesma que está nos eletroferogramas das figuras 3, 4 e 5, pois a apresentada na tabela está de acordo com a padronização de apresentação sugerida por Kayser *et al.*, 2012, que tem como objetivo facilitar a análise dos dados e a compilação das informações acerca destes marcadores encontradas em estudos realizados em diferentes países (Ballantyne *et al.*, 2014).

Tabela 7 - Haplótipos RM-YSTR de indivíduos do Rio de Janeiro (Continua)

Código da amostra	Haplótipo	DYF387S1	DYS570	DYS576	DYF399S1	DYS518	DYS526 a/b	DYS626	DYS627	DYF403S1 a/b	DYF404S1	DYS449	DYS547	DYS612	N	f
120124-20SP	1	29	18	19	15.3-18	29	13-36	16	18	10-13-18-51	16	29	43	25	1	0,0077
120124-29SP	2	29-30	17	19	16.3-18-19.3	29	14-37	16	19	13-18-52	15	28	44	25	1	0,0077
120210-20SP	3	29-30	18	19	15.3-20	29	13-36	17	17	9-14-17-48	15-16	31	45	24	1	0,0077
120112-09SP	4	29-31	16	19	15.3-16.3-18	30	14-37	16	17	11-13-18-54	15-16	29	42	26	1	0,0077
120111-07SP	5	29-31	17	18	15.3-18.3-19	29	14-37	17	17	10-13-18-50	13-15	29	43	26	1	0,0077
120210-22SP	6	29-32	16	18	15-15.3-16	26	11-33	18	16	11-16-20-47	14-16	34	37	29	1	0,0077
111117-22SP	7	29-32	17	18	15.3-16.3-18	30	12-33	17	18	8-13-18-50	15-17	29	44	25	1	0,0077
111122-04SP	8	29-32	17	19	15.3-16.3-18	30	12-33	17	18	10-14-18-50	15-16	29	44	25	1	0,0077
120111-29SP	9	29-32	17	20	14.3-15.3-17	29	14-37	17	16	13-15-16-52	13-16	28	39	26	1	0,0077
111121-12SP	10	29-32	18	15	13.3-15.3-19	31	11-31	12	17	11-15-16-52	16-17	34	43	25	1	0,0077
120210-24SP	11	29-32	18	19	15.3-17.3-19	31	13-35	16	20	12-19-50	14-15	29	45	26	1	0,0077
120124-27SP	12	30	16	17	16.3-21	30	13-35	16	19	10-13-18-51	15	28	42	28	1	0,0077
120126-10SP	13	30	17	13	14.3-19-19.3	31	13-34	17	20	11-13-16-50	15-16	29	43	26	1	0,0077
120126-06SP	14	30	17	17	15.3	28	13-35	17	17	12-15-17-52	16	30	43	28	1	0,0077
120124-17SP	15	30	18	18	15.3-18-19.3	31	13-35	18	18	10-12-15-48	15-18	31	42	26	1	0,0077
120126-21SP	16	30	19	18	13.3-16.3-22	28	12-35	16	18	11-14-17-50	15	30	43	26	1	0,0077

Tabela 7- Haplótipos RM-YSTR de indivíduos do Rio de Janeiro (Continuação)

Código da amostra	Haplótipo	DYF387S1	DYS570	DYS576	DYF399S1	DYS518	DYS526 a/b	DYS626	DYS627	DYF403S1 a/b	DYF404S1	DYS449	DYS547	DYS612	N	f
120110-05SP	18	30-31	16	18	15.3-17.3-21	31	13-35	16	17	10-15-17-49	15-16	29	44	27	1	0,0077
100615.28SP	19	30-31	16	20	13.3-14.3-21	28	13-36	18	18	9-16-19-52	15-16	30	44	26	1	0,0077
120117-06SP	20	30-31	17	16	15.3-18-19.3	27	13-36	16	18	9-13-17-48	15-16	29	43	26	1	0,0077
120112-07SP	21	30-31	17	17	13.3-15.3-18	27	13-38	17	17	13-14-16-43	14-16	29	44	26	1	0,0077
120116-01SP	22	30-31	17	17	14.3-17.3-21	29	12-34	16	18	12-13-18-50	15-16	28	45	29	1	0,0077
111116-11SP	23	30-31	17	17	14.3-20	29	14-37	16	18	11-15-16-52	15	28	42	26	1	0,0077
120111-20SP	24	30-31	17	17	15.3-17.3-19	32	13-37	14	19	10-11-15-49	16	29	45	24	1	0,0077
120208-04SP	25	30-31	17	18	13.3-16.3-20	28	13-38	16	18	11-13-17-49	15-16	29	44	27	1	0,0077
120117-04SP	26	30-31	17	18	14.3-16.3-19	31	13-35	17	21	11-14-17-51	16-17	30	43	26	1	0,0077
120126-29SP	27	30-31	17	18	15.3-17.3-21	28	13-35	17	18	10-13-17-49	15-16	30	43	26	1	0,0077
120209-34SP	28	30-31	17	18	15.3-18.3-21	31	13-36	17	21	10-13-17-50	15	30	43	28	1	0,0077
120209-33SP	29	30-31	17	18	16.3-17.3-21	29	13-34	16	21	14-16-17-50	16	29	41	26	1	0,0077
120116-18SP	30	30-31	17	19	13.3-16.3-18	27	13-37	16	18	11-14-16-42	14-16	29	44	26	1	0,0077
120209-06SP	31	30-31	17	19	13.3-17.3-19	27	13-37	16	20	11-13-17-48	13-16	29	43	24	1	0,0077
120116-07SP	35	30-31	18	16	16.3-17.3-19	30	13-35	18	19	10-13-16-49	16-18	29	45	28	1	0,0077
120124-35SP	36	30-31	18	17	13.3-16.3-19	30	13-36	17	18	12-15-17-50	15-16	29	43	25	1	0,0077
120126-05SP	37	30-31	18	17	15.3-16-18.3	30	13-37	17	18	9-10-19-49	15-16	29	45	26	1	0,0077
120210-02SP	38	30-31	18	18	14.3-15.3-20	29	14-37	17	17	11-12-18-51	15-17	29	45	25	1	0,0077
111118-04SP	39	30-31	18	18	15.3-17.3-21	29	13-36	14	24	11-13-18-50	15-16	29	45	26	1	0,0077
111122-08SP	40	30-31	18	20	13.3-16.3-20	30	13-35	17	18	12-13-19-50	15	29	44	26	1	0,0077
120201-23SP	41	30-31	19	17	13.3-17.3-18	29	13-37	16	20	12-14-16-50	15-16	29	44	27	1	0,0077

Tabela 7 - Haplótipos RM-YSTR de indivíduos do Rio de Janeiro (Continuação)

Código da amostra	Haplótipo	DYF387S1	DYS570	DYS576	DYF399S1	DYS518	DYS526 a/b	DYS626	DYS627	DYF403S1 a/b	DYF404S1	DYS449	DYS547	DYS612	N	f
120124-22SP	42	30-31	19	18	14.3-15.3-19	29	14-37	17	17	11-12-18-50	15-17	29	45	25	1	0,0077
111122-16SP	43	30-31	19	18	14.3-17.3-21	29	13-35	16	19	10-13-18-51	16-18	29	44	27	1	0,0077
111122-28SP	44	30-31	19	18	15.3-16-18.3	29	13-35	18	20	11-13-16-47	14-16	31	44	24	1	0,0077
111117-15SP	45	30-31	19	18	15.3-17-17.3	29	13-35	18	20	11-13-16-47	14-16	29	43	24	1	0,0077
120113-24SP	46	30-31	19	18	15.3-18.3-19	28	13-35	16	19	12-15-18-51	15-17	29	43	27	1	0,0077
120126-04SP	47	30-31	19	19	14.3-15.3-18	30	13-35	16	20	12-14-16-49	15-16	30	44	25	1	0,0077
120111-25SP	48	30-31	19	19	14.3-16.3-21	28	14-38	15	18	9-15-19-49	17	30	43	26	1	0,0077
120117-22SP	49	30-31	23	19	12.3-15.3-19	33	11-31	9	14	11-15-48	15-16	31	43	26	1	0,0077
120124-12SP	50	30-31-33-34	20	15	12-13.3-15-15.3-17.3-19	32	11-31	12	19	14-16-18-50	15-16-17	32	44	27	1	0,0077
111118-01SP	51	30-32	18	16	14.3-17.3-19	29	13-36	15	20	12-14-18-49	15-17	30	43	26	1	0,0077
120112-11SP	52	30-32	18	17	13.3-16.3-19	32	11-31	12	18	16-21-49	16-17	34	44	26	1	0,0077
111122-21SP	53	30-32	18	17	15.3-16.3-19	29	13-38	17	18	12-14-19-50	15-16	30	42	27	1	0,0077
111117-03SP	54	30-32	18	17	15.3-16.3-20	29	13-36	16	17	10-13-18-50	15	30	43	25	1	0,0077
111123-14SP	55	30-33	16	19	13.3-15.3-16.3	31	12-34	16	17	11-19-14-50	14-19	30	38	29	1	0,0077
111116-03SP	56	30-33	16	19	14.3-20	29	13-33	16	18	10-12-17-51	15-17	28	43	27	1	0,0077
120125-03SP	57	30-33	19	15	13.3-17.3-18.3	29	11-32	16	16	13-15-52	14-18	28	47	23	1	0,0077
120110-24SP	58	30-33	21	16	16.3	30	13-34	14	19	10-17-19-48	16	28	42	25	1	0,0077
120110-22SP	59	30-34	17	17	11-15.3-19	25	13-33	19	18	8-11-16-49	17-18	29	42	26	1	0,0077
120208-26SP	60	31	16	18	14.3-17.3-20	28	13-36	16	19	12-17-51	15-16	28	43	27	1	0,0077
120112-10SP	61	31	16	19	13.3-17.3-18	29	14-36	16	18	12-15-18-50	15-16	31	44	27	1	0,0077

Tabela 7 - Haplótipos RM-YSTR de indivíduos do Rio de Janeiro (Continuação)

Código da amostra	Haplótipo	DYF387S1	DYS570	DYS576	DYF399S1	DYS518	DYS526 a/b	DYS626	DYS627	DYF403S1 a/b	DYF404S1	DYS449	DYS547	DYS612	N	f
120110-27SP	62	31	17	18	14.3-16.3-21	30	13-36	16	20	11-14-18-51	15-16	30	41	25	1	0,0077
120116-24SP	63	31	17	18	15.3-17.3-19	30	13-35	16	19	13-14-16-50	17	30	42	27	1	0,0077
111123-05SP	64	31	17	19	15.3-17.3-20	29	13-35	18	22	11-14-16-48	15-17	30	42	26	1	0,0077
111116-12SP	65	31	17	19	15.3-19-20	29	13-35	18	19	11-13-15-50	15-16	30	43	28	1	0,0077
120125-19SP	66	31	17	20	15.3-16.3-19	29	13-36	16	19	12-14-16-51	15-17	29	43	26	1	0,0077
120209-16SP	67	31	18	16	14.3-18.3-19	30	13-35	17	19	11-12-16-50	15-17	29	44	27	1	0,0077
111122-33SP	68	31	18	17	14.3-19-19.3	30	12-37	17	20	8-13-17-50	15-16	29	41	25	1	0,0077
120126-12SP	69	31	18	18	14.3-17.3-19	28	13-36	16	21	12-13-18-51	15-16	32	44	28	1	0,0077
120116-22SP	70	31	18	18	15.3-17.3-20	30	13-36	16	19	11-13-18-51	15-16	29	42	25	1	0,0077
120208-18SP	71	31	18	19	13.3-15.3-20	30	13-35	16	19	12-14-17-50	17	30	42	27	1	0,0077
120112-18SP	72	31	18	19	16.3-17-18.3	28	13-35	13	17	10-13-16-49	15	28	42	29	1	0,0077
120117-18SP	73	31	19	18	14.3-16.3-20	30	13-36	15	18	11-12-17-52	15	30	44	26	1	0,0077
120209-02SP	74	31	19	19	13.3-17.3-19	31	13-34	17	17	12-16-50	15	31	45	28	1	0,0077
120111-13SP	75	31	21	18	11.2-17.3-18	36	11-30	13	18	13-15-49	16-17	33	43	25	1	0,0077
120116-14SP	76	31-32	17	18	14.3-18.3-21	29	13-37	14	20	9-12-17-51	16	28	43	26	1	0,0077
120117-12SP	77	31-32	18	17	15.3-18	31	13-35	16	18	13-14-17-51	14-16	35	44	27	1	0,0077
111122-09SP	78	31-32	18	18	14.3-16.3-18.3	30	13-34	18	20	12-17-18-47	15-16	33	43	24	1	0,0077
120208-25SP	79	31-32	19	19	12.3-14.3-17.3	28	12-33	15	18	11-19-20-46	15-16	29	41	20	1	0,0077
120124-31SP	80	31-32-35	17	19	11.3-13.3-16.3-18	30	12-32	19	15	12-13-18-48	12-16	34	41	23	1	0,0077
111117-24SP	81	31-33	18	18	16.3-18.3-20	29	13-35	19	20	10-13-18-51	14-16	30	43	27	1	0,0077

Tabela 7 - Haplótipos RM-YSTR de indivíduos do Rio de Janeiro (Continuação)

Código da amostra	Haplótipo	DYF387S1	DYS570	DYS576	DYF399S1	DYS518	DYS526 a/b	DYS626	DYS627	DYF403S1 a/b	DYF404S1	DYS449	DYS547	DYS612	N	f
120209-23SP	82	31-33	17	14	13.3-16.3-18.3	30	13-33	18	16	11-12-17-51	14-16	30	46	23	1	0,0077
120125-22SP	83	31-33	19	15	14.3-18.3-20.3	30	11-30	17	16	14-15-54	14-17	28	45	23	1	0,0077
111117-02SP	84	31-33	20	17	13.3-17.3-20.3	29	14-34	16	15	13-14-15-52	14-15	31	45	23	1	0,0077
120207-06SP	85	31-34	18	17	13.3-15.3-17	29	12-34	17	20	11-13-15-50	14-16	32	43	24	1	0,0077
111123-17SP	86	31-34	19	15	13.3-16.3	31	13-35	17	17	13-15-17-49	15-16	31	45	23	1	0,0077
120210-12SP	87	31-34	20	16	16.3-19.3	31	12-33	18	16	14-15-49	14-16	30	43	26	1	0,0077
120112-19SP	88	31-34	22	14	12.2-18.3-20.3	28	13-32	17	16	11-13-15-49	15-16	28	43	23	1	0,0077
100922.38SP	89	31-35	17	17	14.3-15-15.1	28	13-35	18	16	13-17-45	14	28	43	26	1	0,0077
120126-19SP	90	31-35	19	16	15.3-16.3-17.3	29	13-32	18	19	14-15-51	15-16	28	44	23	1	0,0077
120110-17SP	91	31-35	20	16	14.3-16.3-17.3	31	12-34	18	19	12-17-19-47	14-17	18	41	27	1	0,0077
120125-11SP	92	31-36	20	19	12.3-16.3	30	13-34	17	18	11-15-45	14-17	30	42	24	1	0,0077
120112-16SP	93	32	17	15	16.3	31	15-37	17	18	10-16-49	14-15	27	39	25	1	0,0077
111116-04SP	94	32	18	17	15.3-16.3-18.3	30	16-37	20	16	10-18-19-47	13-15	28	40	25	1	0,0077
111118-30SP	95	32	19	17	15.3-16.3-17.3	30	11-32	16	17	13-16-19-45	15-17	25	41	26	1	0,0077
111117-19SP	96	32-33	16	17	12.3-17.3-20	31	11-35	16	12	9-13-17-49	16-17	26	41	26	1	0,0077
120113-31SP	97	32-33	17	14	16-17.3-18	27	11-34	17	17	13-16-44	13-15	33	41	23	1	0,0077
120116-23SP	98	32-33	18	17	14.3-16.3	30	15-37	20	16	10-16-20-48	12-15	27	40	28	1	0,0077
120126-35SP	99	32-33	19	17	15.3-17-17.3	29	11-33	16	15	7-11-16-47	15-17	29	47	29	1	0,0077
111121-02SP	100	32-33	20	17	14.3-16.3-18.3	30	13-33	18	19	13-16-50	15-18	27	44	23	1	0,0077
120209-17SP	101	32-33	21	15	13.3-17.3	31	13-35	17	16	13-15-51	15-16	30	45	24	1	0,0077

Tabela 7 - Haplótipos RM-YSTR de indivíduos do Rio de Janeiro (Continuação)

Código da amostra	Haplótipo	DYF387S1	DYS570	DYS576	DYF399S1	DYS518	DYS526 a/b	DYS626	DYS627	DYF403S1 a/b	DYF404S1	DYS449	DYS547	DYS612	N	f
120117-21SP	102	32-33	22	16	14.3-16.3	30	15-36	19	20	10-18-19-48	13-14	28	40	25	1	0,0077
100527.34SP	103	32-33	22	17	15.3-17.3-18.3	33	15-36	20	17	10-16-18-46	12-16	29	41	24	1	0,0077
120125-15SP	104	32-34	18	15	13.3-16.3-17.3	30	13-32	18	15	11-14-15-52	15	33	44	25	1	0,0077
120111-12SP	105	32-34	18	18	14.3-18.3-19	31	14-34	20	15	10-14-18-50	15-16	33	42	26	1	0,0077
120209-12SP	106	32-34	18	19	12.3-18.3-19.3	30	13-33	18	17	12-14-15-50	14-18	26	45	24	1	0,0077
111116-18SP	107	32-34	21	16	13.3-17.3-20.3	30	14-34	18	20	13-14-16-50	14-17	27	42	22	1	0,0077
111122-25SP	108	32-34	21	16	17.3-19.3	30	16-38	19	16	12-16-18-45	14-15	30	41	26	1	0,0077
120208-13SP	109	32-35	16	16	14.3-16.3-20	29	12-34	17	16	11-12-17-48	14-16	30	42	25	1	0,0077
111117-10SP	110	32-35	17	15	11-15.3-18	27	12-33	20	20	14-15-16-43	14	28	44	26	1	0,0077
111122-22SP	111	32-35	17	16	11-16.3-19	28	12-33	19	18	15-16-43	14	28	44	26	1	0,0077
111122-11SP	112	32-35	18	14	14.3-15.3-16.3	29	13-32	19	16	12-13-16-51	15-20	30	45	25	1	0,0077
120125-07SP	113	32-35	18	18	15-15.1-16.3	27	12-33	16	16	16-18-46	13.2-15	27	45	25	1	0,0077
120124-39SP	114	33	16	16	15.3-16-16.3	30	11-34	17	19	11-17-47	16	31	44	30	1	0,0077
120209-09SP	115	33	17	16	16.3-18-20	28	11-34	20	17	14-15-17-44	15-16	39	40	21	1	0,0077
120209-04SP	116	33	19	16	15.3-16.3-18.3	32	15-36	19	17	10-16-18-47	13-16	30	41	25	1	0,0077
120126-27SP	117	33	19	18	14.3-17.3-19.3	31	13-33	17	19	13-14-17-48	14-16	31	44	21	1	0,0077
120117-10SP	118	33	20	18	11.3-15.3-17	31	10-30	12	16	11-17-46	14	33	43	27	1	0,0077
120116-05SP	119	33-34	16	16	13.3-15.3-17	30	13-34	18	17	11-18-48	14-15	31	41	26	1	0,0077
120124-34SP	120	33-34	17	15	10.3-12.3-13.3-17-18	29	13-35	17	17	11-12-16-54	14-15-18	28	43	26	1	0,0077
111118-05SP	121	33-34	17	18	14.3-15.3-17.3	30	12-31	20	16	12-15-16-51	15-16	31	45	24	1	0,0077

Tabela 7 - Haplótipos RM-YSTR de indivíduos do Rio de Janeiro (Conclusão)

Código da amostra	Haplótipo	DYF387S1	DYS570	DYS576	DYF399S1	DYS518	DYS526 a/b	DYS626	DYS627	DYF403S1 a/b	DYF404S1	DYS449	DYS547	DYS612	N	f
120208-03SP	122	33-34	18	16	14.3-17.3-19.3	29	12-33	21	18	14-15-48	14-15	31	42	27	1	0,0077
120209-11SP	123	33-34	19	16	14.3-15.3-17.3	31	15-34	18	16	10-17-18-46	13-15	29	41	26	1	0,0077
120111-09SP	124	33-34	20	15	14.3-16.3-20.3	29	13-32	16	18	11-15-52	15-16	28	45	23	1	0,0077
111121-13SP	125	33-35	19	15	13.3-17.3-18.3	29	13-32	15	18	13-15-51	15-16	28	44	22	1	0,0077
120209-03SP	126	34	18	18	12-14.3-19	28	12-34	14	18	16-17-18-19-43	14.2-16	28	40	28	1	0,0077
120207-04SP	127	34	19	18	12.3-16.3-18.3-19	27	12-30	13	18	13-18-50	15-16	35	47	27	1	0,0077
120113-08SP	128	34	20	17	12.2-16.3-19	32	13-35	16	17	10-13-16-46	15-16	30	47	25	1	0,0077
111123-16SP	129	34-35	19	16	15.3-18.3	30	12-35	19	18	10-11-15-48	14-16	32	45	26	1	0,0077

Legenda: Nesta tabela, estão listados os códigos das amostras dos pais, juntamente com seus respectivos haplótipos, o número de vezes que cada haplótipo é encontrado nessa amostra (N) e suas frequências (f). Os dados desta tabela foram gerados através da leitura dos alelos detectados pela emissão da fluorescência captada através da eletroforese capilar no sequenciador automático ABI PRISM 3500 após amplificação por PCR multiplex de alíquotas de DNA.

Ao analisar a tabela 7, verificam-se 129 haplótipos distintos nos 129 indivíduos miscigenados, com a tipagem das 13 regiões RM-YSTR do cromossomo Y.

3.2 Estimativa das Frequências Haplotípicas e Poder de Discriminação

A genotipagem dos 129 indivíduos analisados neste estudo se correlacionou com 129 haplótipos distintos, com frequências de 0,0077, conforme foi mostrado na Tabela 7. Sendo Além disso, a análise dos 13 marcadores RM-YSTR revelou um poder de discriminação de 100% ($\pm 1,2190$) na população miscigenada do Rio de Janeiro.

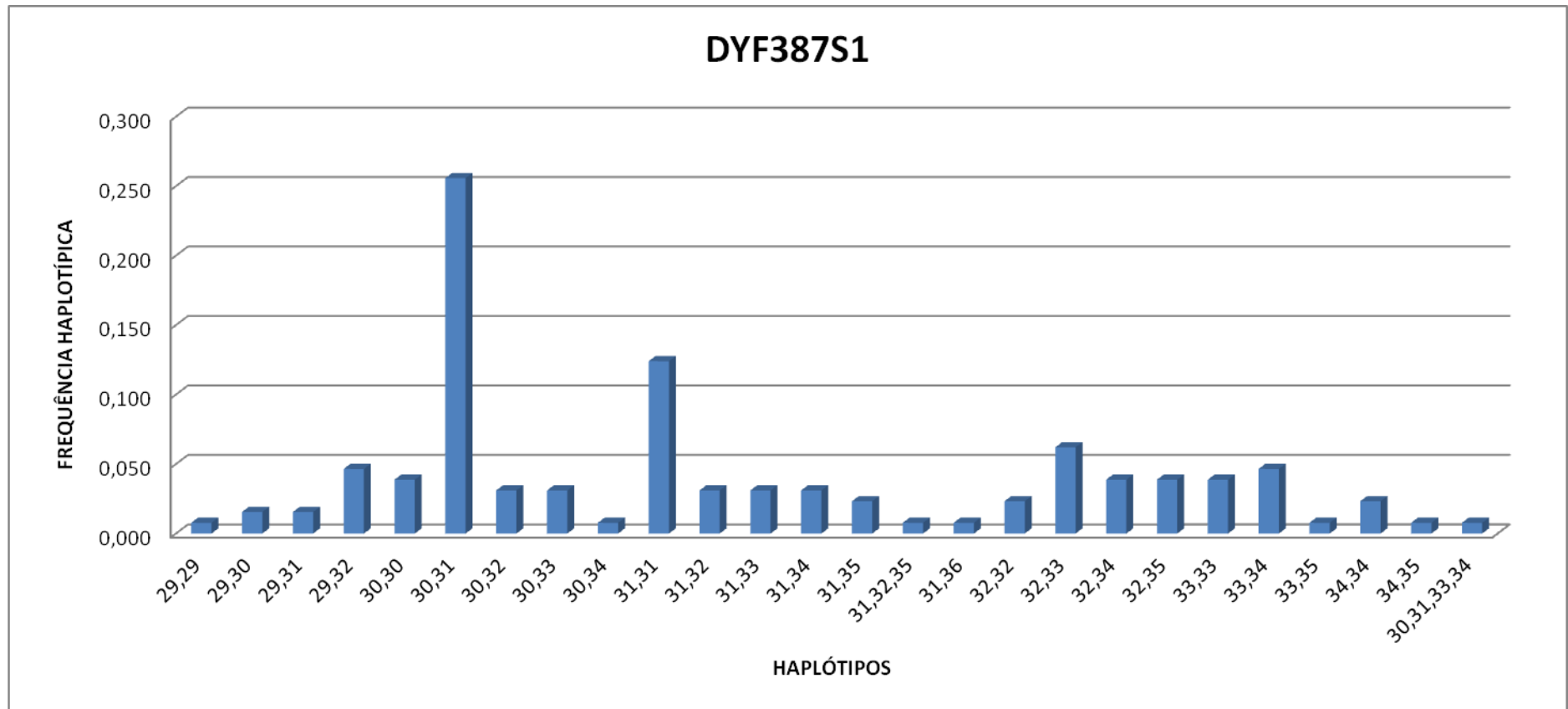
3.3 Estimativa das Frequências alélicas dos *loci* RM-YSTR

Com o intuito de realizar um levantamento acerca das características dos marcadores de rápida mutação na população miscigenada do estado do Rio de Janeiro foram determinadas as frequências alélicas de cada um dos 13 *loci* RM-YSTR. Para os *loci* multialélicos, ao invés de se calcular a frequência alélica, calculou-se a frequência haplotípica de cada um destes *locus*.

As frequências dos diferentes alelos detectados em cada *locus* que compõe o conjunto de marcadores RM-YSTR foram definidas por meio da tipagem de 129 indivíduos do sexo masculino não relacionados geneticamente.

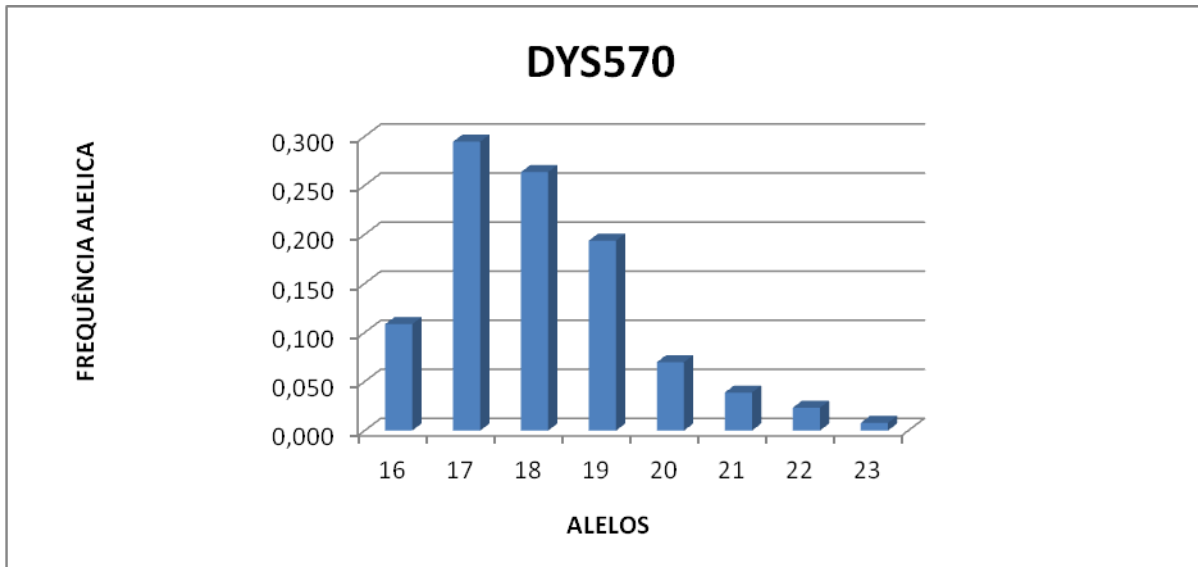
As distribuições das frequências alélicas, ou haplotípicas, encontradas em cada um dos 13 *loci* são apresentadas nos gráficos 1 a 13.

Gráfico 1 - Distribuição das frequências haplotípicas do locus DYF387S1 na população do estado do Rio de Janeiro (N=129)



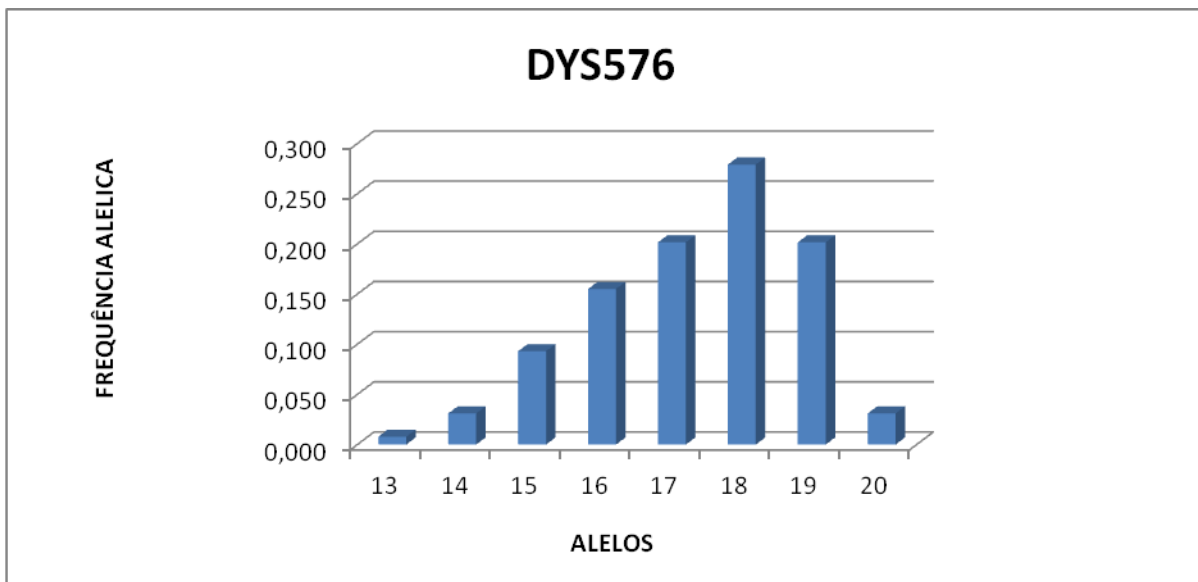
Legenda: Neste gráfico é apresentada a distribuição das frequências haplotípicas referentes ao *locus* DYF387S1 da população do Rio de Janeiro. A estimativa das frequências foi baseada nos dados populacionais deste estudo, gerados após a leitura dos resultados da eletroforese capilar no sequenciador automático ABI PRISM 3500 de produtos de amplificação por PCR multiplex de alíquotas de DNA de 129 indivíduos.

Gráfico 2 - Distribuição das frequências alélicas do locus *DYS570* na população do estado do Rio de Janeiro (N=129)



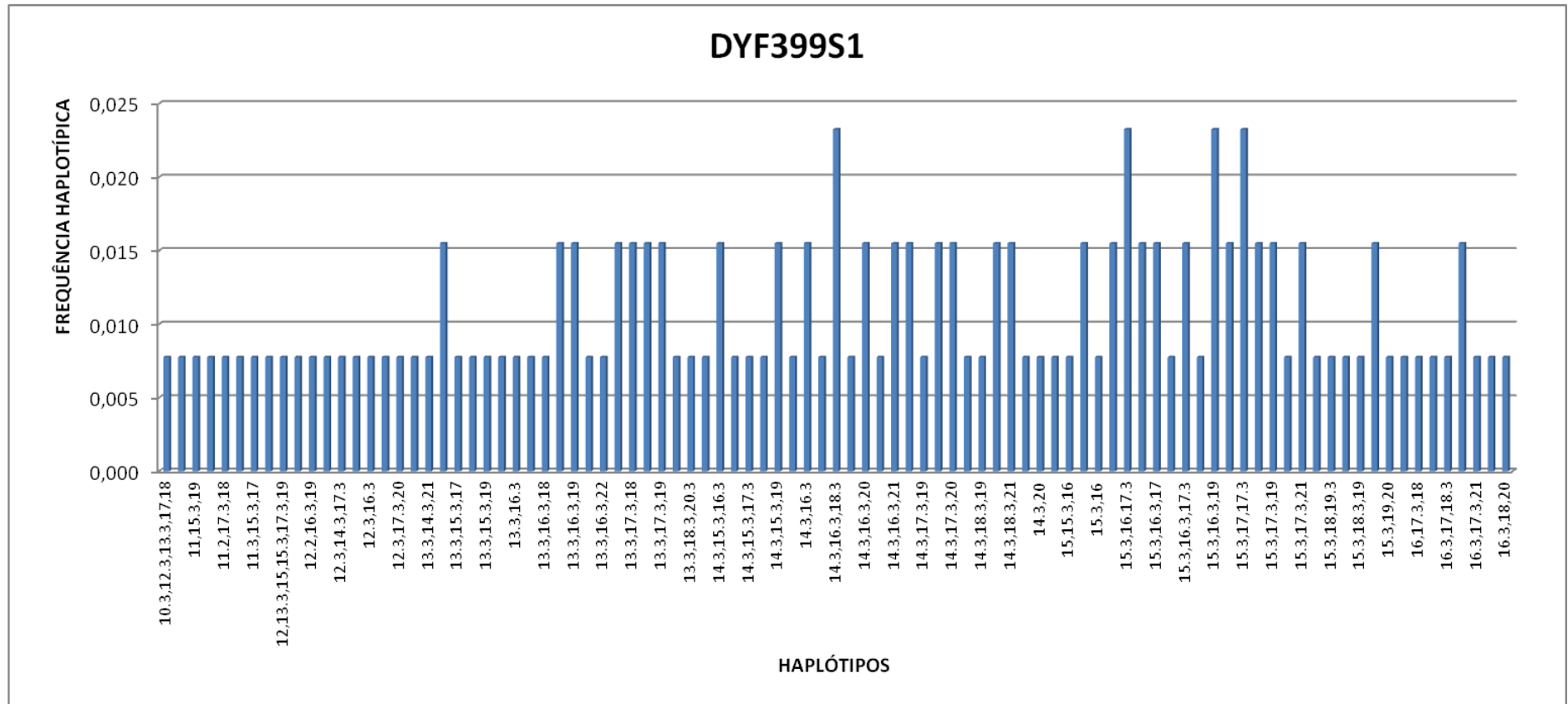
Legenda: Neste gráfico é apresentada a distribuição das frequências alélicas referentes ao *locus* *DYS570* da população do Rio de Janeiro. O cálculo destas frequências foi baseado nos dados populacionais deste estudo, gerados após a leitura dos resultados da eletroforese capilar no sequenciador automático ABI PRISM 3500 de produtos de amplificação por PCR multiplex de alíquotas de DNA de 129 indivíduos.

Gráfico 3 - Distribuição das frequências alélicas do locus *DYS576* na população do estado do Rio de Janeiro (N=129)



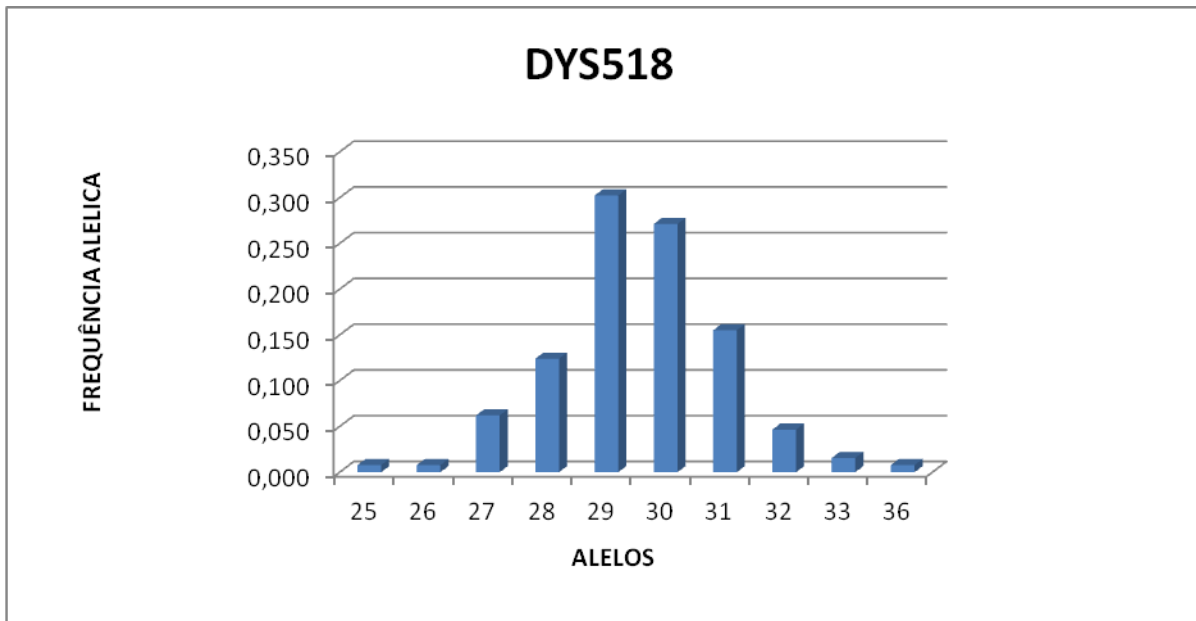
Legenda: Neste gráfico é apresentada a distribuição das frequências alélicas referentes ao *locus* *DYS576* da população do Rio de Janeiro. O cálculo destas frequências foi baseado nos dados populacionais deste estudo, gerados após a leitura dos resultados da eletroforese capilar no sequenciador automático ABI PRISM 3500 de produtos de amplificação por PCR multiplex de alíquotas de DNA de 129 indivíduos.

Gráfico 4 - Distribuição das frequências haplotípicas do locus DYF399S1 na população do estado do Rio de Janeiro (N=129)



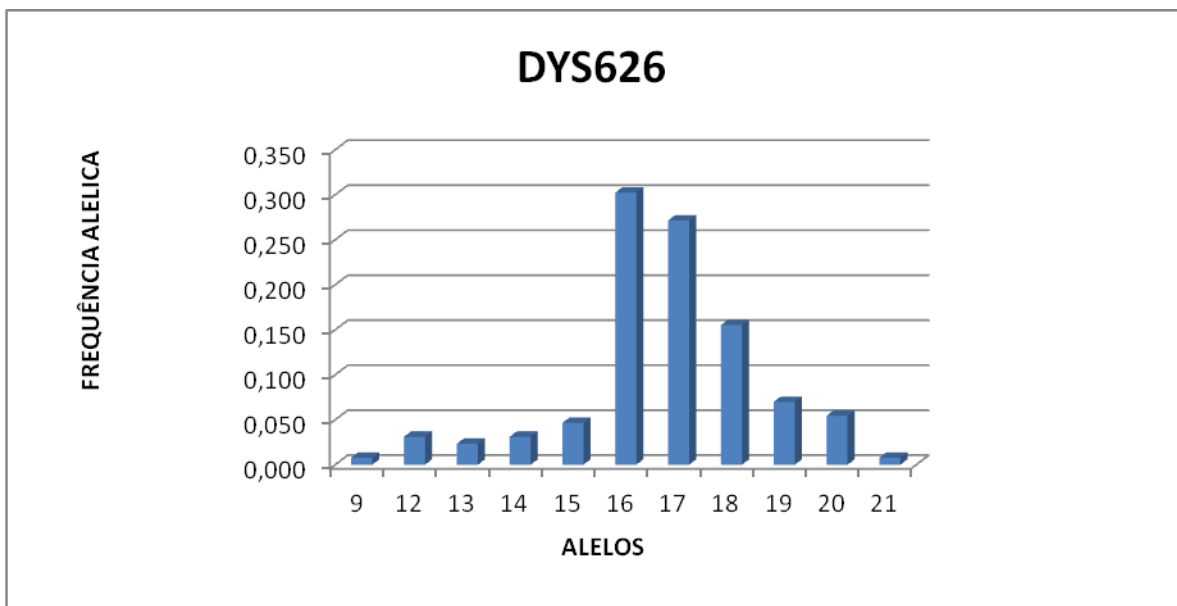
Legenda: Neste gráfico é apresentada a distribuição das frequências haplotípicas referentes ao *locus* DYF399S1 da população do Rio de Janeiro. O cálculo destas frequências foi baseado nos dados populacionais deste estudo, gerados após a leitura dos resultados da eletroforese capilar no sequenciador automático ABI PRISM 3500 de produtos de amplificação por PCR multiplex de alíquotas de DNA de 129 indivíduos.

Gráfico 5 - Distribuição das frequências alélicas do locus DYS518 na população do estado do Rio de Janeiro (N=129)

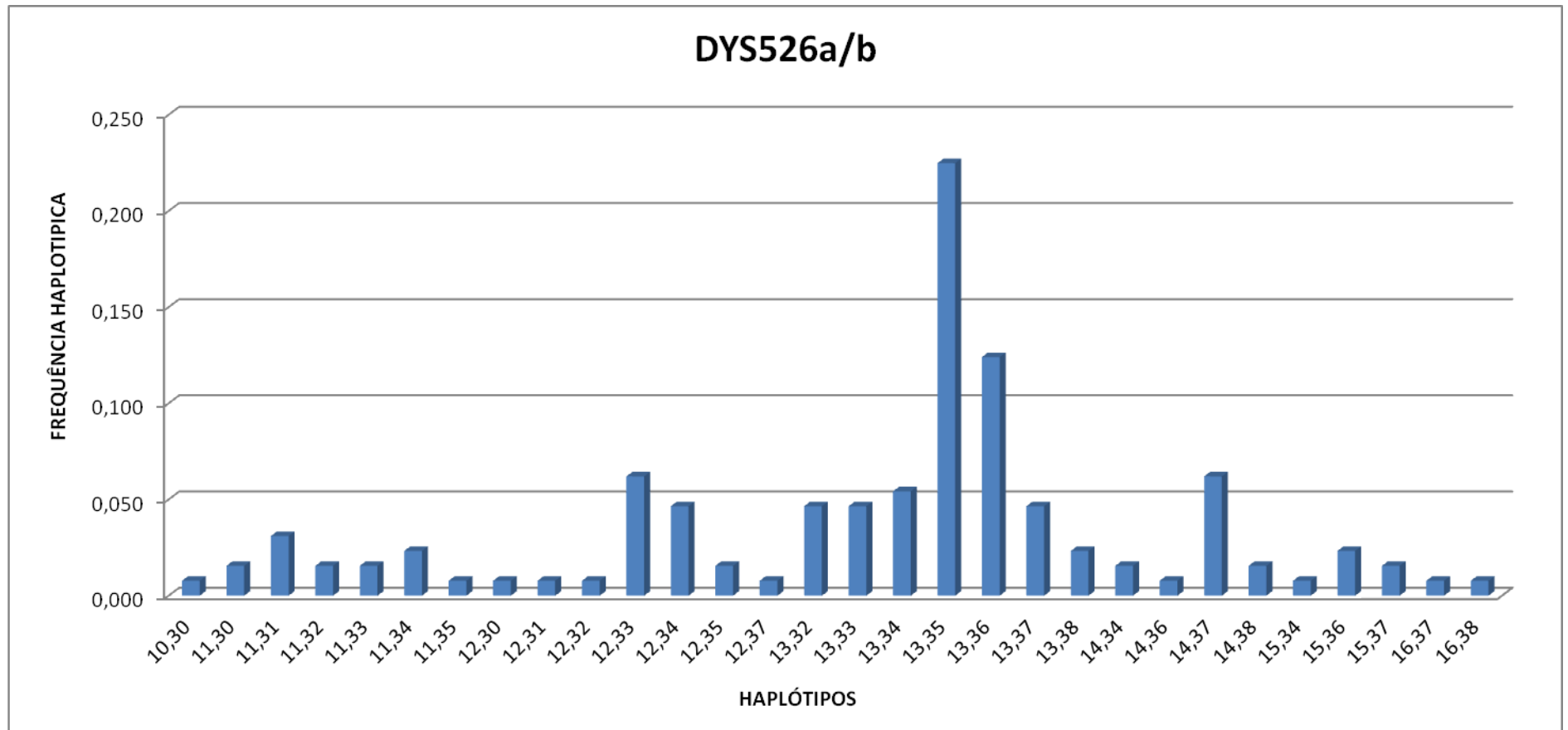


Legenda: Neste gráfico é apresentada a distribuição das frequências alélicas referentes ao *locus* DYS518 da população do Rio de Janeiro. O cálculo destas frequências foi baseado nos dados populacionais deste estudo, gerados após a leitura dos resultados da eletroforese capilar no sequenciador automático ABI PRISM 3500 de produtos de amplificação por PCR multiplex de alíquotas de DNA de 129 indivíduos.

Gráfico 6 - Distribuição das frequências alélicas do locus DYS626 na população do estado do Rio de Janeiro (N=129)

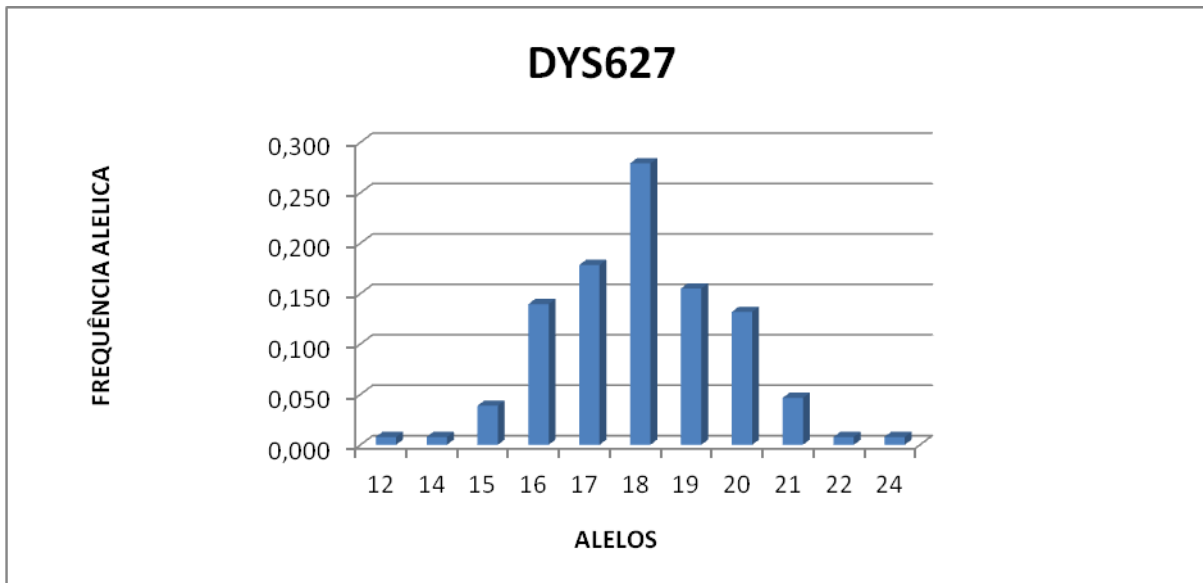


Legenda: Neste gráfico é apresentada a distribuição das frequências alélicas referentes ao *locus* DYS626 da população do Rio de Janeiro. O cálculo destas frequências foi baseado nos dados populacionais deste estudo, gerados após a leitura dos resultados da eletroforese capilar no sequenciador automático ABI PRISM 3500 de produtos de amplificação por PCR multiplex de alíquotas de DNA de 129 indivíduos.

Gráfico 7 - Distribuição das frequências alélicas do locus *DYS526a/b* na população do estado do Rio de Janeiro (N=129)

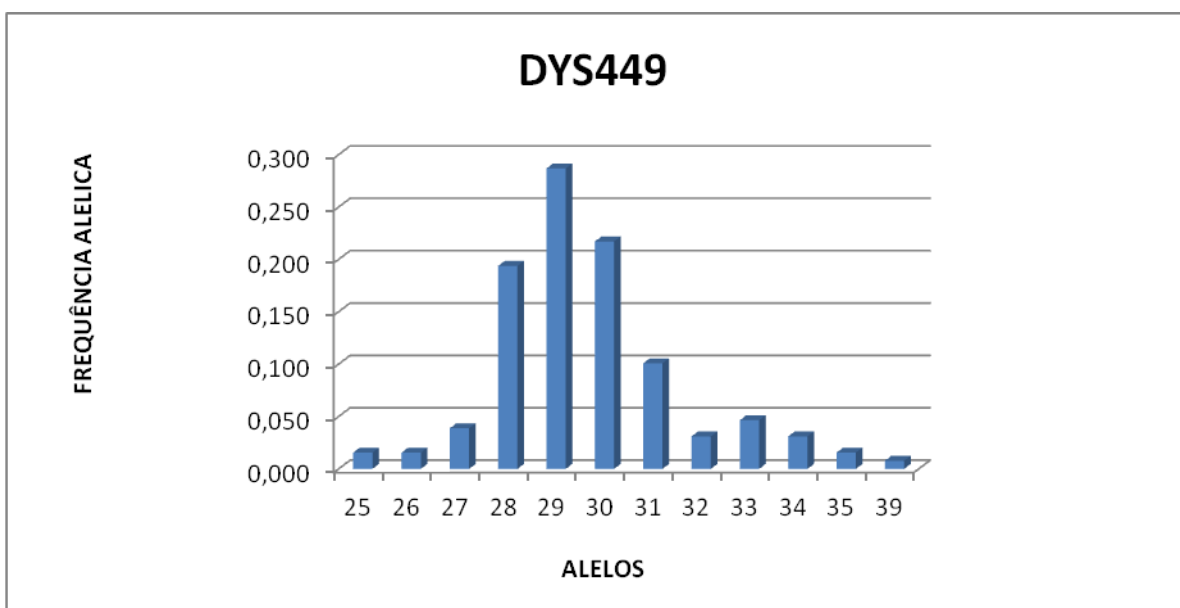
Legenda: Neste gráfico é apresentada a distribuição das frequências haplotípicas referentes ao *locus* *DYS526a/b* da população do Rio de Janeiro. O cálculo destas frequências foi baseado nos dados populacionais deste estudo, gerados após a leitura dos resultados da eletroforese capilar no sequenciador automático ABI PRISM 3500 de produtos de amplificação por PCR multiplex de alíquotas de DNA de 129 indivíduos.

Gráfico 8 - Distribuição das frequências alélicas do locus *DYS627* na população do estado do Rio de Janeiro (N=129)



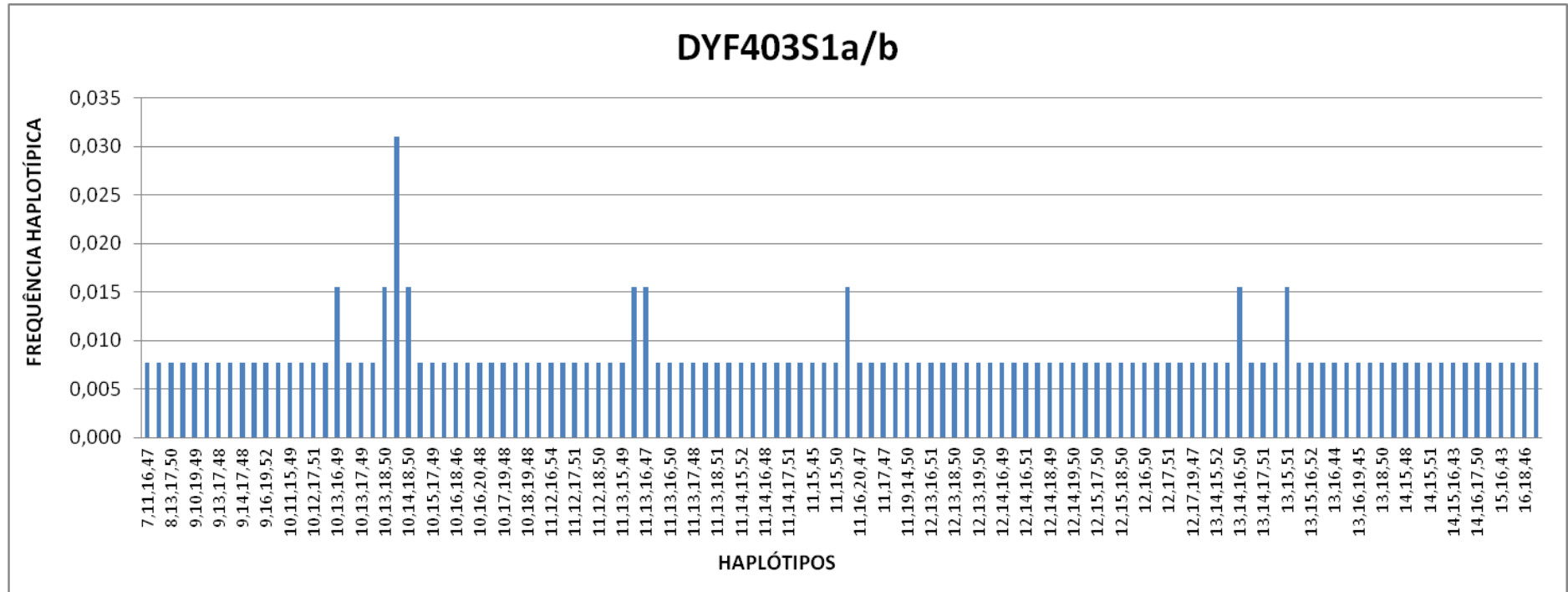
Legenda: Neste gráfico é apresentada a distribuição das frequências alélicas referentes ao *locus* *DYS627* da população do Rio de Janeiro. O cálculo destas frequências foi baseado nos dados populacionais deste estudo, gerados após a leitura dos resultados da eletroforese capilar no sequenciador automático ABI PRISM 3500 de produtos de amplificação por PCR multiplex de alíquotas de DNA de 129 indivíduos.

Gráfico 9 - Distribuição das frequências alélicas do locus *DYS449* na população do estado do Rio de Janeiro (N=129)



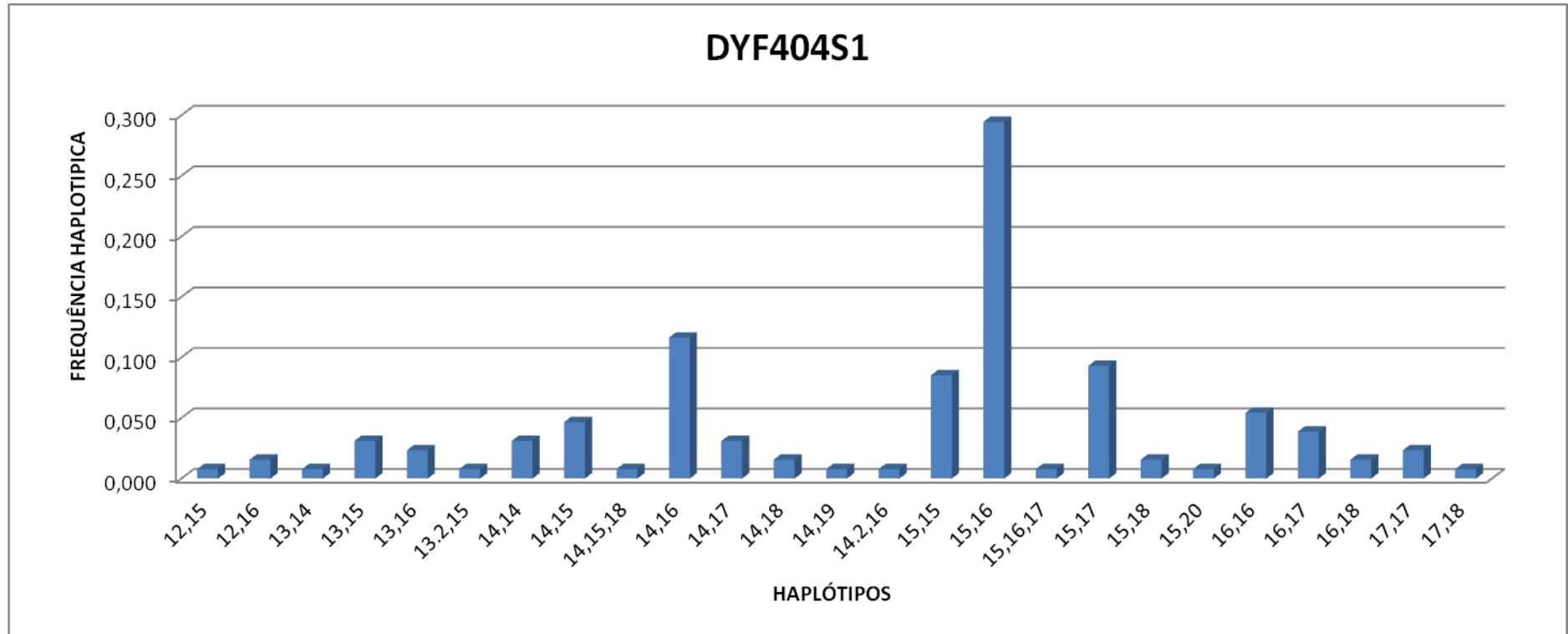
Legenda: Neste gráfico é apresentada a distribuição das frequências alélicas referentes ao *locus* *DYS449* da população do Rio de Janeiro. O cálculo destas frequências foi baseado nos dados populacionais deste estudo, gerados após a leitura dos resultados da eletroforese capilar no sequenciador automático ABI PRISM 3500 de produtos de amplificação por PCR multiplex de alíquotas de DNA de 129 indivíduos.

Gráfico 10 - Distribuição das frequências haplotípicas do locus DYF403S1a/b na população do estado do Rio de Janeiro (N=129)



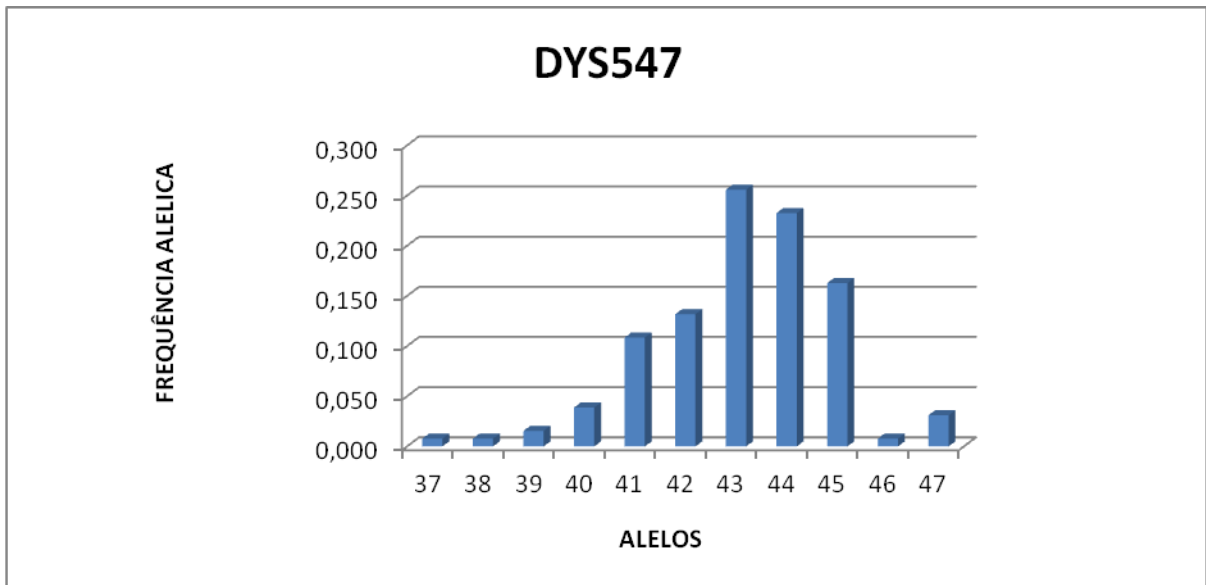
Legenda: Neste gráfico é apresentada a distribuição das frequências haplotípicas referentes ao locus DYF403S1a/b da população do Rio de Janeiro. O cálculo destas frequências foi baseado nos dados populacionais deste estudo, gerados após a leitura dos resultados da eletroforese capilar no sequenciador automático ABI PRISM 3500 de produtos de amplificação por PCR multiplex de alíquotas de DNA de 129 indivíduos.

Gráfico 11 - Distribuição das frequências haplotípicas do locus DYF404S1 na população do estado do Rio de Janeiro (N=129)



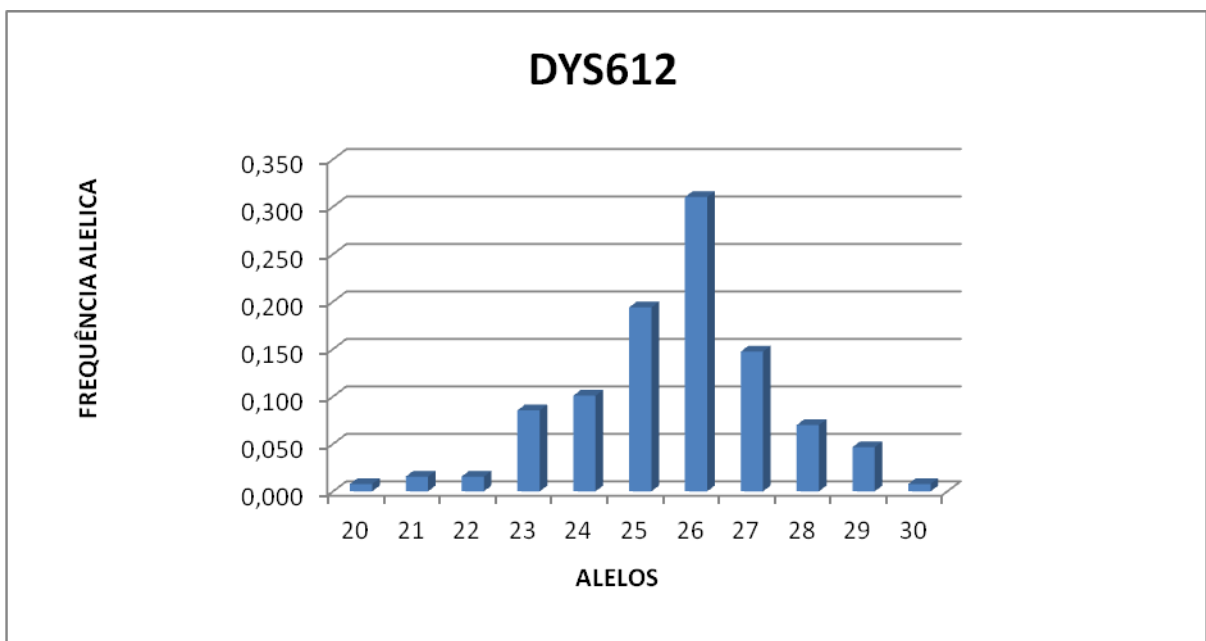
Legenda: Neste gráfico é apresentada a distribuição das frequências haplotípicas referentes ao *locus* DYF404S1 da população do Rio de Janeiro. O cálculo destas frequências foi baseado nos dados populacionais deste estudo, gerados após a leitura dos resultados da eletroforese capilar no sequenciador automático ABI PRISM 3500 de produtos de amplificação por PCR multiplex de alíquotas de DNA de 129 indivíduos.

Gráfico 12 - Distribuição das frequências alélicas do locus DYS547 na população do estado do Rio de Janeiro (N=129)



Legenda: Neste gráfico é apresentada a distribuição das frequências alélicas referentes ao *locus* DYS547 da população do Rio de Janeiro. O cálculo destas frequências foi baseado nos dados populacionais deste estudo, gerados após a leitura dos resultados da eletroforese capilar no sequenciador automático ABI PRISM 3500 de produtos de amplificação por PCR multiplex de alíquotas de DNA de 129 indivíduos.

Gráfico 13 - Distribuição das frequências alélicas do locus DYS612 na população do estado do Rio de Janeiro (N=129)



Legenda: Neste gráfico é apresentada a distribuição das frequências alélicas referentes ao *locus* DYS612 da população do Rio de Janeiro. O cálculo destas frequências foi baseado nos dados populacionais deste estudo, gerados após a leitura dos resultados da eletroforese capilar no sequenciador automático ABI PRISM 3500 de produtos de amplificação por PCR multiplex de alíquotas de DNA de 129 indivíduos.

3.4 Análise da Diversidade Gênica por *locus* e Diversidade Haplotípica

O cálculo da diversidade gênica foi realizado para cada um dos *loci* considerando as frequências alélicas ou haplotípicas, no caso de marcadores multialélicos, verificadas na amostra da população miscigenada do Rio de Janeiro analisada por este estudo. Este parâmetro corresponde à probabilidade de dois alelos escolhidos ao acaso na população serem diferentes e permite a comparação da diversidade entre e dentro das populações, visto que está relacionada com o grau de polimorfismo de cada *locus*.

Conforme pode ser observado na Tabela 8, todos os *loci* analisados apresentam valores elevados de diversidade gênica. O *locus* DYF403S1a/b se destaca por apresentar a maior diversidade gênica (0,9983). O *locus* DYF399S1 apresenta a segunda maior diversidade gênica (0,9950), superior àquelas dos *loci* DYS526a/b e DYF387S1, cujas diversidades gênicas foram 0,9170 e 0,9046, respectivamente. Os demais marcadores, apresentaram valores de diversidade gênica entre 0,7936 e 0,8785.

Tabela 8 – Diversidade gênica nos 13 loci RM-YSTR

<i>Locus</i>	Diversidade Gênica
DYF387S1	0,9046
DYS570	0,7919
DYS576	0,8125
DYF399S1	0,9950
DYS518	0,7953
DYS526a/b	0,9170
DYS626	0,8046
DYS627	0,8320
DYF403S1a/b	0,9983
DYF404S1	0,8785
DYS449	0,8229
DYS547	0,8283
DYS612	0,8259

Legenda: Após a tipagem dos alelos detectados por eletroforese capilar no sequenciador automático ABI PRISM 3500 dos produtos de amplificação por PCR multiplex de alíquotas de DNA dos indivíduos da amostra populacional do Rio de Janeiro, a diversidade gênica por *locus* foi calculada conforme descrito em Materiais e Métodos.

O cálculo da diversidade haplotípica do conjunto de marcadores RM-YSTR foi feito com base nas frequências haplotípicas encontradas na população estudada, conforme dados previamente apresentados na Tabela 7. A diversidade haplotípica verificada pela genotipagem de homens da população miscigenada do Rio de Janeiro nos 13 *loci* RM-YSTR foi de 1,00 ($\pm 1,2190$), o que demonstra que este conjunto de marcadores tem elevada capacidade de discriminar indivíduos na população em estudo.

3.5 Análise da distância genética entre populações

Com o intuito de averiguar o nível de diversidade da amostra analisada nesse estudo e estimar o poder de diferenciação, também conhecido como distância molecular, entre tal população e outras de diversos países, comparou-se a amostra de indivíduos miscigenados do Rio de Janeiro com outras populações apresentadas na Tabela 6 do subitem 2.4.5 de Materiais e Métodos. Além disso, foi calculada a variância de dados moleculares, também chamada de AMOVA, por meio da qual é possível estimar o nível da diferenciação populacional a partir de informações geradas por perfis genéticos.

Tanto o cálculo do nível de diversidade (F_{st}) quanto o de distância molecular entre populações (AMOVA) foram realizados no *software* Arlequin 3.5 (Schneider *et al.*, 2000). Para tais estimativas, o *software* utiliza o conjunto de alelos que forma o haplótipo de cada indivíduo da amostra, contabilizando o número de mutações e as considerando como diferenças. Desta maneira, ele compara o número de diferenças entre os grupos regionais, entre haplótipos em diferentes populações de um mesmo grupo e entre haplótipos dentro de uma mesma população.

Nas Tabelas 9, 10 e 11 são mostrados, respectivamente, os valores distância genética baseados no índice F_{st} , seus valores de P e a matriz de significância resultante da comparação entre as amostras populacionais citadas na Tabela 6.

Tabela 9 - Valores de distância genética baseados no Fst

	Bakangos	Guinea Bissau	Xhosa	Gra-Bretanha	Marches	Holanda	Kanto	Timor Leste	Bai	RJ	Buenos Aires	Costa Rica
Bakangos	0											
Guinea Bissau	0.01517	0										
Xhosa	0.04912	0.03533	0									
Gra-Bretanha	0.04552	0.04871	0.08325	0								
Marches	0.03447	0.03558	0.06689	0.00675	0							
Holanda	0.04524	0.04970	0.07889	0.00541	0.00985	0						
Kanto	0.05037	0.04139	0.07713	0.03818	0.03238	0.04752	0					
Timor Leste	0.02721	0.02703	0.04875	0.03883	0.03065	0.04132	0.02877	0				
Bai	0.04765	0.04394	0.07606	0.04756	0.04095	0.05420	0.02971	0.03033	0			
RJ	0.03712	0.02735	0.05409	0.01319	0.01287	0.01808	0.02858	0.02198	0.03273	0		
Buenos Aires	0.03765	0.03524	0.06080	0.01072	0.00978	0.01674	0.02254	0.02358	0.03264	0.00166	0	
Costa Rica	0.04027	0.03003	0.05592	0.01259	0.01083	0.01576	0.02358	0.02295	0.02995	0.00138	0,00089	0

Os valores de Fst variam de 0 (não há diferenças genéticas entre as populações) a 1 (há diferenças genéticas entre as populações), indicando o nível de distanciamento genético entre as populações.

Tabela 10 – Valores de P para o índice de Fst

	Bakangos	Guinea Bissau	Xhosa	Gra-Bretanha	Marches	Holanda	Kanto	Timor Leste	Bai	RJ	Buenos Aires	Costa Rica
Bakangos	*											
Guinea Bissau	0,00049±0,0002	*										
Xhosa	0,00000	0,00000	*									
Gra-Bretanha	0,00000	0,00000	0,00000	*								
Marches	0,00000	0,00000	0,00000	0,00000	*							
Holanda	0,00000	0,00000	0,00000	0,00235±0,0005	0,00000	*						
Kanto	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	*					
Timor Leste	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	*				
Bai	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	*			
RJ	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	*		
Buenos Aires	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,10989±0,0032	*	
Costa Rica	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,14381±0,0040	0,25586±0,0045	*

Os valores são significativos para $P \leq 0,004$.

Tabela 11 - Matriz de significância dos valores de P para o índice de Fst

	Bakangos	Guinea Bissau	Xhosa	Gra-Bretanha	Marches	Holanda	Kanto	Timor Leste	Bai	RJ	Buenos Aires	Costa Rica
Bakangos		+	+	+	+	+	+	+	+	+	+	+
Guinea Bissau	+		+	+	+	+	+	+	+	+	+	+
Xhosa	+	+		+	+	+	+	+	+	+	+	+
Gra-Bretanha	+	+	+		+	+	+	+	+	+	+	+
Marches	+	+	+	+		+	+	+	+	+	+	+
Holanda	+	+	+	+	+		+	+	+	+	+	+
Kanto	+	+	+	+	+	+		+	+	+	+	+
Timor Leste	+	+	+	+	+	+	+		+	+	+	+
Bai	+	+	+	+	+	+	+	+		+	+	+
RJ	+	+	+	+	+	+	+	+	+		-	-
Buenos Aires	+	+	+	+	+	+	+	+	+	-		-
Costa Rica	+	+	+	+	+	+	+	+	+	-	-	

Os valores de significância para 10200 permutações. (+) indica $P \leq 0,004$ e (-) indica $P > 0,004$.

Na Tabela 12 são mostrados os valores de AMOVA baseados no índice de F_{st} , gerados pelo cálculo da estimativa da diferenciação populacional baseada em informações geradas por perfis moleculares, entre a amostra populacional do Rio de Janeiro e as amostras provenientes da América, África Sub-saariana, Europa e Ásia (Ballantyne *et al.*, 2014).

Tabela 12 - Valores de AMOVA baseados no índice de F_{st} resultantes da comparação entre amostras miscigenadas do continente Americano e amostras da África Subsaariana, Europa e Ásia

Fonte de Variação	Soma dos quadrados	Variância	Porcentagem de variação
Entre grupos regionais	161,619	0,12252	2,11
Entre populações do mesmo grupo regional	123,305	0,09802	1,69
Dentro de populações	6850,356	5,58301	96,20
Total	7135,280	5,80355	

3.6 Estudo mutacional dos *loci* RM-YSTR

Objetivando analisar as características mutacionais dos marcadores *rapidly mutating* na amostra populacional de indivíduos miscigenados do Rio de Janeiro, verificaram-se as taxas de mutação desses 13 *loci* em 129 pares de pais e filhos participantes de investigações de paternidade realizadas no Laboratório de Diagnósticos por DNA da UERJ, cujos vínculos genéticos de paternidade haviam sido confirmados, com probabilidade de certeza superior a 99,999%, por meio do uso de marcadores genéticos dos cromossomos autossômicos.

3.6.1 Detecção de mutações nos *loci* RM-YSTR

Para os 129 pares de pai/filho analisados neste estudo, foram encontradas 30 mutações em 11 *loci* RM-YSTR, sendo os *loci* *DYS518* e *DYS626* os únicos nos quais não se observaram mutações (Tabela 13).

Tabela 13 - Haplótipos RM-YSTR dos pares de pai/filho nos quais foram observadas mutações (Continua)

Código da amostra	DYF387S1	DYS570	DYS576	DYF399S1	DYS518	DYS526 a/b	DYS626	DYS627	DYF403S1 a/b	DYF404S1	DYS449	DYS547	DYS612
120124-29SP	29-30	17	19	16.3-18-19.3	29	14-37	16	19	13-18-52	15	28	44	25
120124-29C	29-30	17	19	16.3-18-19.3	29	14-37	16	19	13-18-52	15	28	45	25
120112-09SP	29-31	16	19	15.3-16.3-18	30	14-37	16	17	11-13-18-54	15-16	29	42	26
120112-09C	29-31	16	18	15.3-16.3-18	30	14-37	16	18	11-13-18-54	15-16	29	42	26
120210-22SP	29-32	16	18	15-15.3-16	26	11-33	18	16	11-16-20-47	14-16	34	37	29
120210-22C	29-32	16	18	15-15.3-16	26	11-33	18	16	11-16-19-47	14-16	34	37	29
111122-04SP	29-32	17	19	15.3-16.3-18	30	12-33	17	18	10-14-18-50	15-16	29	44	25
111122-04C	29-32	17	19	15.3-16.3-19	30	12-33	17	18	10-14-18-50	15-16	29	44	25
120210-24SP	29-32	18	19	15.3-17.3-19	31	13-35	16	20	12-19-50	14-15	29	45	26
120210-24 C	29-32	18	19	14.3-17.3-19	31	13-35	16	20	12-19-50	14-15	29	45	26
120111-20SP	30-31	17	17	15.3-17.3-19	32	13-37	14	19	10-11-15-49	16	29	45	24
120111-20 C	30-31	18	17	15.3-17.3-19	32	13-37	14	19	10-11-15-49	16	29	45	24
120208-04SP	30-31	17	18	13.3-16.3-20	28	13-38	16	18	11-13-17-49	15-16	29	44	27
120208-04 C	30-31	17	19	13.3-16.3-20	28	13-38	16	18	11-13-17-49	15-16	29	44	27
120209-06SP	30-31	17	19	13.3-17.3-19	27	13-37	16	20	11-13-17-48	13-16	29	43	24
120209-06 C	30-31	17	19	13.3-17.3-19	27	13-36	16	20	11-13-17-48	13-16	29	43	24

Tabela 13 - Haplótipos RM-YSTR dos pares de pai/filho nos quais foram observadas mutações (Continuação)

Código da amostra	DYF387S1	DYS570	DYS576	DYF399S1	DYS518	DYS526 a/b	DYS626	DYS627	DYF403S1 a/b	DYF404S1	DYS449	DYS547	DYS612
111117-15SP	30-31	19	18	15.3-17-17.3	29	13-35	18	20	11-13-16-47	14-16	29	43	24
111117-15 C	30-31	19	18	15.3- 16 -17.3	29	13-35	18	20	11-13-16-47	14-16	29	43	24
120124-12SP	30-31-33-34	20	15	12-13.3-15-15.3-17.3-19	32	11-31	12	19	14-16-18-50	15-16-17	32	44	27
120124-12 C	30-31-33-34	20	15	12-13.3-15-15.3- 16.3 -17.3-19	32	11-31	12	19	14-16-18-50	15-16-17	32	44	27
111118-01SP	30-32	18	16	14.3-17.3-19	29	13-36	15	20	12-14-18-49	15-17	30	43	26
111118-01 C	30-32	18	16	14.3-17.3-19	29	13-36	15	20	13 -14-18-49	15-18	30	43	26
120112-11SP	30-32	18	17	13.3-16.3-19	32	11-31	12	18	16-21-49	16-17	34	44	26
120112-11 C	30-32	18	17	13.3-16.3-19	32	11-31	12	18	16-21-49	16-17	34	44	27
111122-21SP	30-32	18	17	15.3-16.3-19	29	13-38	17	18	12-14-19-50	15-16	30	42	27
111122-21 C	30-32	18	17	15.3- 17.3 -19	29	13-38	17	18	12-14-19-50	15-16	30	42	27
111117-03SP	30-32	18	17	15.3-16.3-20	29	13-36	16	17	10-13-18-50	15	30	43	25
111117-03 C	30-32	18	17	15.3-16.3-20	29	13-36	16	17	10-13-18- 51	15	30	43	25
120110-22SP	30-34	17	17	11-15.3-19	25	13-33	19	18	8-11-16-49	17-18	29	42	26
120110-22 C	30-34	17	17	11-15.3-19	25	13-33	19	18	8-11- 17 -49	17-18	29	42	26
111116-12SP	31	17	19	15.3-19-20	29	13-35	18	19	11-13-15-50	15-16	30	43	28
111116-12 C	31	17	19	15.3- 20	29	13-35	18	19	11-13-15-50	15-16	30	43	28
120209-02SP	31	19	19	13.3-17.3-19	31	13-34	17	17	12-16-50	15	31	45	28
120209-02 C	31	19	19	14.3 -17.3-19	31	13-34	17	17	12-16-50	15	31	45	28
120210-12SP	31-34	20	16	16.3-19.3	31	12-33	18	16	14-15-49	14-16	30	43	26
120210-12 C	31-34	20	16	16.3- 18.3	31	12-33	18	16	14-15-49	14-16	30	43	26

Tabela 13 - Haplótipos RM-YSTR dos pares de pai/filho nos quais foram observadas mutações (Conclusão)

Código da amostra	DYF387S1	DYS570	DYS576	DYF399S1	DYS518	DYS526 a/b	DYS626	DYS627	DYF403S1 a/b	DYF404S1	DYS449	DYS547	DYS612
120112-16SP	32	17	15	16.3	31	15-37	17	18	10-16-49	14-15	27	39	25
120112-16 C	32	17	15	16.3	31	15-37	17	18	10-16-49	14-15	26	39	25
111118-30SP	32	19	17	15.3-16.3-17.3	30	11-32	16	17	13-16-19-45	15-17	25	41	26
111118-30 C	32	19	17	15.3-16.3-17.3	30	11-32	16	17	13-16-20-45	15-17	25	41	26
111116-18SP	32-34	21	16	13.3-17.3-20.3	30	14-34	18	20	13-14-16-50	14-17	27	42	22
111116-18 C	32-34	21	16	13.3-18.3-20.3	30	14-34	18	20	13-14-16-50	14-17	27	42	22
111122-25SP	32-34	21	16	17.3-19.3	30	16-38	19	16	12-16-18-45	14-15	30	41	26
111122-25 C	32-33	21	16	17.3-19.3	30	16-38	19	16	12-16-18-45	14-15	30	41	26
120125-07SP	32-35	18	18	15-15.1-16.3	27	12-33	16	16	16-18-46	13.2-15	27	45	25
120125-07 C	32-35	18	18	15-15.1-16.3	27	12-34	16	16	16-18-46	13.2-15	27	45	25
120124-39SP	33	16	16	15.3-16-16.3	30	11-34	17	19	11-17-47	16	31	44	30
120124-39 C	33	16	16	15.3-16	30	11-34	17	19	11-17-47	16	31	44	30
120209-04SP	33	19	16	15.3-16.3-18.3	32	15-36	19	17	10-16-18-47	13-16	30	41	25
120209-04 C	33	19	16	15.3-16.3-18.3	32	15-36	19	17	10-16-17-47	13-16	30	41	25
111121-13SP	33-35	19	15	13.3-17.3-18.3	29	13-32	15	18	13-15-51	15-16	28	44	22
111121-13 C	34-35	19	15	13.3-17.3-18.3	29	13-32	15	18	13-15-51	15-16	28	44	22
120209-03SP	34	18	18	12-14.3-19	28	12-34	14	18	16-17-18-19-43	14.2-16	28	40	28
120209-03 C	34	18	18	12-14.3-19	28	12-34	14	18	16-17-18-19-43	14.2-16	28	40	27
120207-04SP	34	19	18	12.3-16.3-18.3-19	27	12-30	13	18	13-18-50	15-16	35	47	27
120207-04 C	34	19	18	12.3-16.3-18.3-19	27	12-30	13	18	13-18-50	15-16	34	47	27

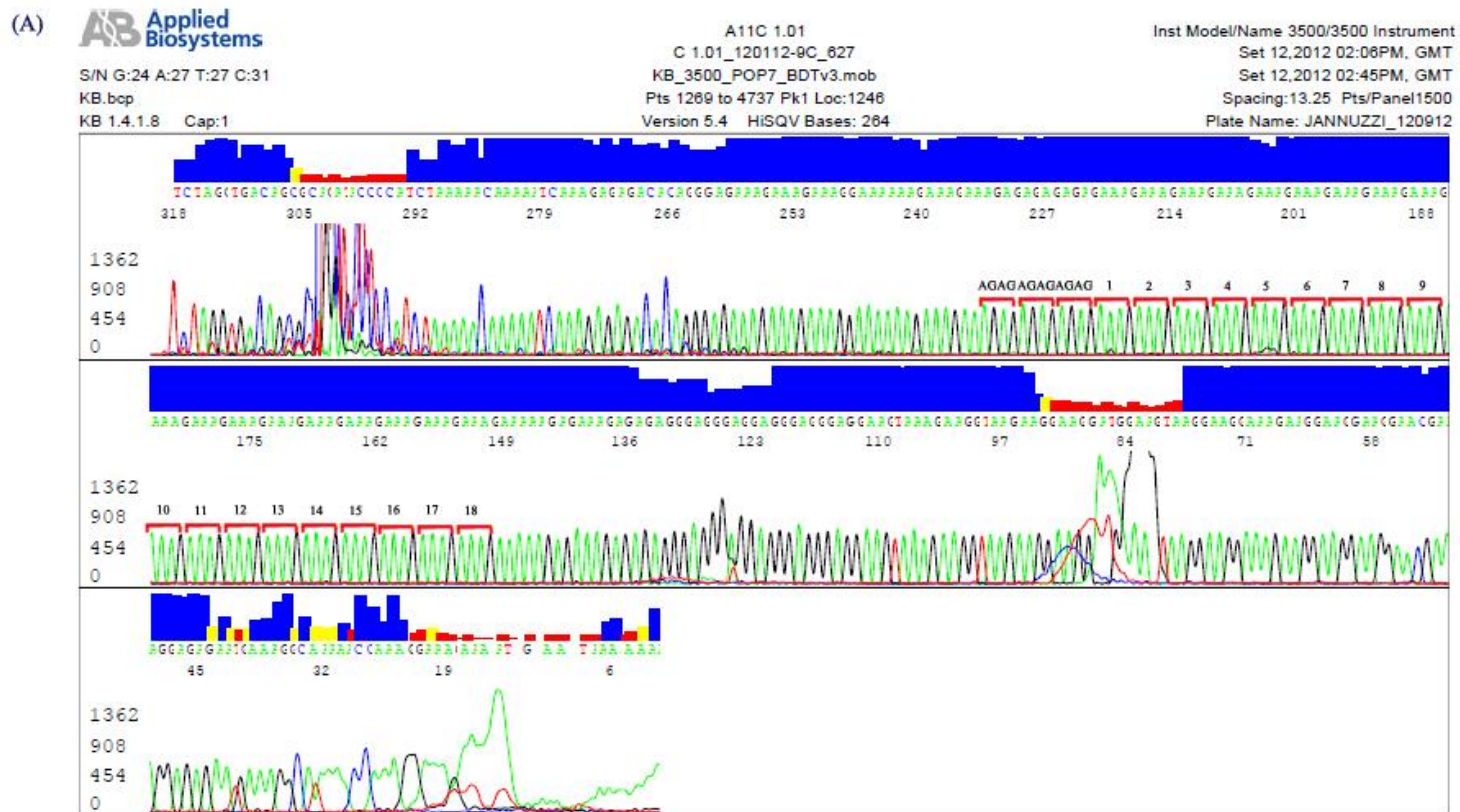
Legenda: Na tabela, são apresentados os pares pai/filho nos quais se verificaram mutações, cujos alelos mutados estão destacados em vermelho. Estes dados foram gerados através da leitura dos alelos detectados pela emissão da fluorescência captada através da eletroforese capilar no sequenciador automático ABI PRISM 3500 após amplificação por PCR multiplex de alíquotas de DNA.

Após a observação de mutações em 28 pares de pai/filho, repetiu-se a tipagem destas amostras e foi realizada a confirmação das 30 mutações, constituindo 12 perdas e 18 ganhos de uma unidade de repetição entre o alelo do pai e o do filho. Adicionalmente, algumas destas amostras foram selecionadas para confirmação da mutação através da técnica de sequenciamento.

Sendo assim, alguns *loci* monoalélicos foram selecionados para o sequenciamento dos alelos, conforme descrito no subitem 2.3.3 de Materiais e Métodos. Os produtos finais do sequenciamento foram detectados por meio de eletroforese capilar em sequenciador automático ABI Prism 3500. Com o sequenciamento dos duos de pai/filho, as mutações foram confirmadas.

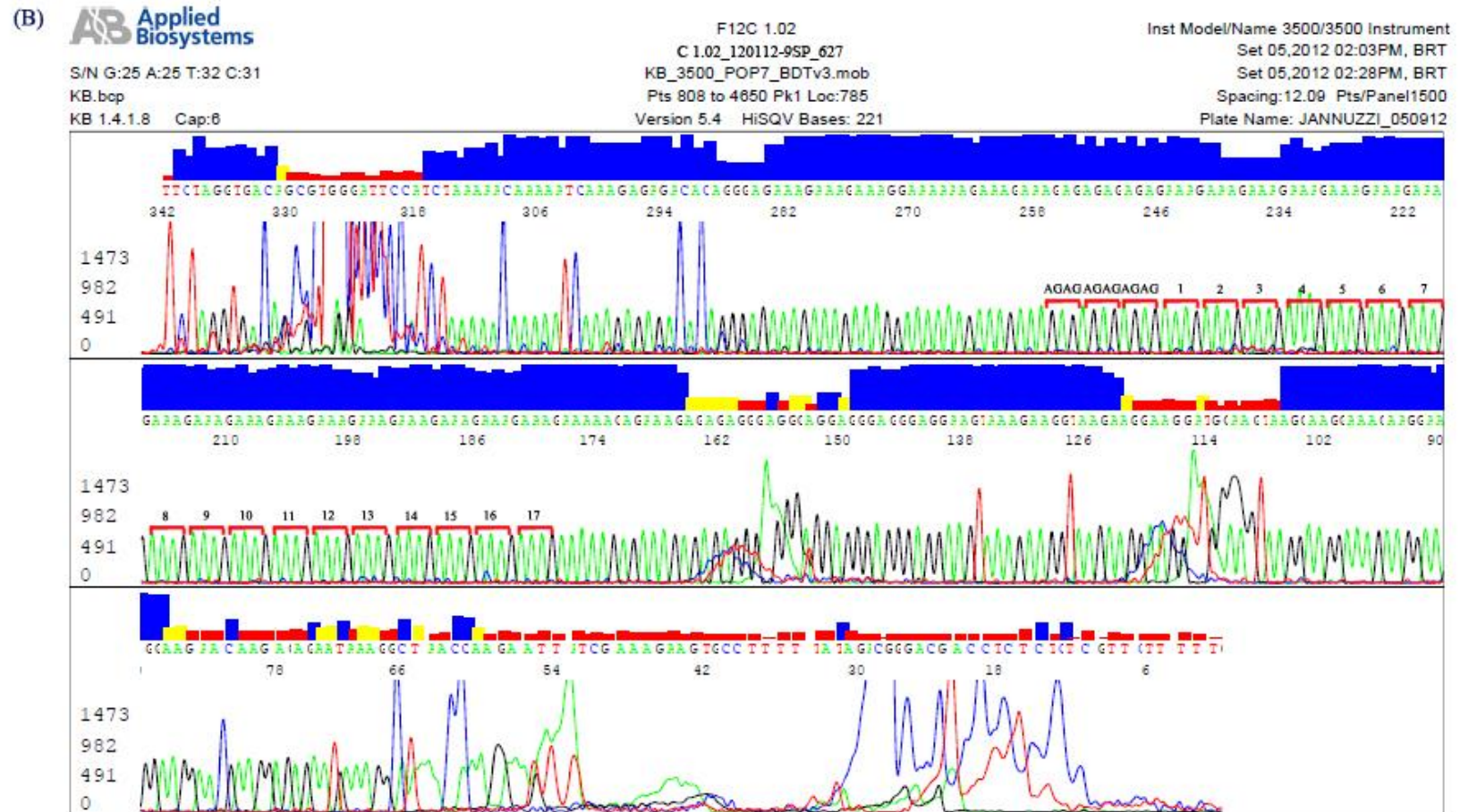
A figura 6 exemplifica um dos eletroferogramas de sequenciamento realizados para o duo 120112-9 referente ao marcador DYS627.

Figura 6A – Eletroferograma de sequenciamento de alelos do locus DYS627



Legenda: Sequenciamento automático do duo 120112-9 na região DYS627 para confirmação de mutação. O sequenciamento foi realizado com o *primer* R, gerando a sequência anti-senso. (A) Eletroferograma do filho do duo 120112-9 evidenciando o alelo mutado (alelo 18), que possui uma unidade de repetição a mais que o pai. (B) Eletroferograma do pai do duo 120112-9 evidenciando o alelo 17, que possui uma unidade de repetição a menos que o filho.

Figura 6B – Eletroferograma de sequenciamento de alelos do locus DYS627



Legenda: Sequenciamento automático do duo 120112-9 na região DYS627 para confirmação de mutação. O sequenciamento foi realizado com o *primer* R, gerando a sequência anti-senso. (A) Eletroferograma do filho do duo 120112-9 evidenciando o alelo mutado (alelo 18), que possui uma unidade de repetição a mais que o pai. (B) Eletroferograma do pai do duo 120112-9 evidenciando o alelo 17, que possui uma unidade de repetição a menos que o filho.

3.6.2 Análise da taxa de mutação

Para a análise mutacional dos marcadores RM-YSTR foram calculadas as taxas de mutação por *locus* e a taxa de mutação média geradas pelos dados dos indivíduos mutados dessa amostra. A tabela 14 mostra os *loci* nos quais se observaram mutações, o número de meioses ocorridas e os resultados dos cálculos das taxas de mutação média e por *locus* encontradas na população miscigenada analisada por este estudo.

Tabela 14 – Características mutacionais dos loci RM-YSTR na população miscigenada do Rio de Janeiro

<i>Locus</i>	Número de mutações	Número de meioses	Taxa de mutação	
DYF387S1	2	129	0,015503876	$1,55 \times 10^{-2}$
DYS570	1	129	0,007751938	$0,77 \times 10^{-2}$
DYS576	2	129	0,015503876	$1,55 \times 10^{-2}$
DYF399S1	10	129	0,07751938	$7,75 \times 10^{-2}$
DYS526a/b	2 (0 +2)	129	0,015503876	$1,55 \times 10^{-2}$
DYS627	1	129	0,007751938	$0,77 \times 10^{-2}$
DYS547	1	129	0,007751938	$0,77 \times 10^{-2}$
DYS449	2	129	0,015503876	$1,55 \times 10^{-2}$
DYS612	2	129	0,015503876	$1,55 \times 10^{-2}$
DYF404S1	1	129	0,007751938	$0,77 \times 10^{-2}$
DYF403S1a/b	6 (5 +1)	129	0,046511628	$4,65 \times 10^{-2}$
Média	30	129	0,021141649	$2,11 \times 10^{-2}$

Legenda: Dados observados na análise dos duos, cujos genótipos foram obtidos através da leitura dos alelos detectados pela emissão da fluorescência captada através da eletroforese capilar no sequenciador automático ABI PRISM 3500 após amplificação por PCR multiplex de alíquotas de DNA.

4. DISCUSSÃO

4.1 Determinação dos haplótipos e nomenclatura dos RM-YSTRs

Os 13 *loci* RM-YSTR analisados neste estudo foram amplificados por meio de três sistemas multiplex denominados RM1, RM2 e RM3 (Kayser *et al.*, 2012). Dentre estes microssatélites, 9 são conhecidos por serem monoalélicos, apresentando apenas um alelo em cada indivíduo analisado e 4 são multialélicos, podendo apresentar 2 ou mais alelos. Sabe-se que os marcadores monoalélicos são comuns no cromossomo Y, caracterizando seu estado haploide, e que as sequências de DNA presentes neste cromossoma podem apresentar variações devido a mutações e a rearranjos estruturais ao longo do tempo, o que pode justificar a existência de *loci* multialélicos no Y. Segundo a literatura disponível, os STRs multialélicos são frequentemente encontrados nas regiões palíndromas deste cromossoma. Devido aos tipos de sequências genéticas presentes em tais regiões, houve o favorecimento de eventos mutacionais, que provocaram a replicação de alguns *loci* e a diferenciação do número de repetições dos alelos ao longo do processo evolutivo (Skaletsky *et al.*, 2003; Rozen *et al.*, 2003; Kayser *et al.*, 2004). Tal fato pode contribuir para o esclarecimento acerca das características dos *loci* RM-YSTR, principalmente com relação ao número de alelos que eles apresentam.

Os perfis genéticos típicos deste conjunto de marcadores encontrados na amostra do Rio de Janeiro analisada no presente trabalho foram mostrados na seção 3.1 de Resultados e são discutidos a seguir.

Na figura 3 é possível visualizar o perfil genético característico para os marcadores amplificados com o multiplex RM1, em que há dois *loci* multialélicos (DYF399S1 e DYF387S1) e dois *loci* monoalélicos (DYS570 e DYS576). Na amostra analisada, o *locus* DYS570 é constituído por alelos com 16 a 23 repetições, cujos comprimentos variam de 240 a 280 pares de bases, sendo o alelo 17 frequentemente encontrado nos indivíduos tipados para este marcador, presente em 30% da amostra. Para este *locus*, o alelo 17 é o mais frequente a nível mundial, principalmente em populações do continente americano caracterizadas por apresentarem elevado grau de miscigenação e em migrantes europeus. Este alelo é mais raro em populações isoladas da Europa, África e América do Sul (Ballantyne *et al.*, 2014).

O *locus* DYS576 é formado por alelos que variam de 13 a 20 repetições, na faixa de 170 a 205pb, apresentando o alelo 18 em 28% dos indivíduos. Na amostra analisada neste

estudo, para ambos os *loci* foram encontrados apenas um alelo em cada indivíduo, porém alguns trabalhos já observaram indivíduos com dois alelos em cada um destes marcadores (Davis *et al.*, 2013; Purps *et al.*, 2014). Para este *locus*, o alelo 18 também é o mais frequente a nível mundial, principalmente na região norte da África e em populações miscigenadas do continente americano, apresentando baixas frequências em países da África Sub-saariana (Ballantyne *et al.*, 2014).

No que diz respeito aos *loci* multialélicos, o *locus* DYF387S1 é caracterizado por apresentar um ou dois alelos, com aproximadamente 240 a 280pb. Para este marcador, observou-se que o haplótipo formado pelos alelos 30,31 é o mais frequente na amostra do Rio de Janeiro, presente em 25% dos indivíduos, o que também é observado em outras populações (Ballantyne *et al.*, 2014). Além disso, nesta amostra se observou que um indivíduo apresentava o haplótipo 30,31,33,34 que foi possivelmente originado após eventos mutacionais raros que acarretaram em quatro cópias alélicas com diferentes números de unidades repetitivas.

O *locus* DYF399S1 é conhecido por ser complexo e multialélico apresentando, geralmente, quatro alelos com diferentes números de repetições, além de cópias alélicas microvariantes. Tais características também são corroboradas em nossa amostra, na qual é possível encontrar indivíduos com um ou até cinco alelos, variando entre 260 e 315pb. Devido a tais características este *locus* é extremamente polimórfico e um marcador com potencial para ser utilizado principalmente na área forense, pois, é capaz de discriminar indivíduos de uma população. Segundo o gráfico 4, os haplótipos gerados por este marcador são bem distribuídos na amostra analisada, com frequências muito parecidas, o que indica que este *locus* é capaz de discriminar indivíduos de uma população com uma maior eficiência do que outros marcadores. Esse padrão de distribuição de frequências também é observado a nível global, corroborando o potencial de discriminação deste marcador em diferentes populações (Ballantyne *et al.*, 2014).

Na figura 4 é possível visualizar o perfil genético característico para os marcadores amplificados com a multiplex RM2, que é formada basicamente por marcadores monoalélicos, com exceção do *locus* DYS526a/b, que apresenta dois alelos. Na amostra analisada, o *locus* DYS518 é constituído por alelos com 25 a 36 repetições, com comprimentos entre 240 e 290pb, apresentando o alelo 29 como o mais frequente na amostra de indivíduos do Rio de Janeiro, presente em 30% dos participantes. Este alelo também é encontrado com frequência elevada na população mundial (www.yhrd.org; Ballantyne *et al.*, 2014).

O *locus* *DYS526a/b* é subdividido em dois segmentos designados a e b, com um alelo em cada segmento. Os alelos *DYS526a* variam de 10 a 16 repetições (138-166pb) e os alelos *DYS526b* variam de 30 a 38 repetições (345-397pb) em nossa amostra. Apesar desta subdivisão, os alelos do *locus* *DYS526a/b* são transmitidos à geração seguinte na forma de um haplótipo, sendo o 13,35 o mais frequente na amostra analisada e na maioria das populações até então estudadas (Ballantyne *et al.*, 2014). O marcador *DYS626* tem alelos que possuem de 9 a 21 repetições (220-270pb) e o mais frequente em nossa amostra é o alelo 16, que é o segundo alelo mais observado em outras populações. Já para o marcador *DYS 627*, observaram-se alelos com 12 a 24 repetições (300-372pb), sendo o alelo 18 o mais frequentemente verificado. Este alelo também é frequente em alguns países do norte da África e da região centro-sul da Ásia, sendo raramente observado no Japão e no Timor-Leste (www.yhrd.org; Ballantyne *et al.*, 2014).

Na figura 5 é mostrado o perfil genético característico para os marcadores amplificados com a multiplex RM3, que possui três marcadores monoalélicos (*DYS612*, *DYS449* e *DYS547*) e dois multialélicos (*DYF403a/b* e *DYF404S1*). O *locus* *DYF403a/b* é considerado extremamente complexo, pois é formado por quatro *loci*, sendo subdividido em *DYF403a* e *DYF403b*. A porção *DYF403a* é constituída por três *loci*, agrupados por possuírem as mesmas características estruturais (300-350pb) e a *DYF403b* possui um *loci* com estrutura repetitiva diferente da apresentada pelos outros três (430-490pb), além de uma baixa emissão de fluorescência em todos os eletroferogramas analisados neste estudo, em conformidade com o que é descrito pela literatura (Kayser *et al.*, 2012; Ballantyne *et al.*, 2014). Na amostra analisada em nosso estudo, é possível observar que os alelos deste marcador possuem frequências bem distribuídas e quase semelhantes em todos os indivíduos estudados, indicando que este *locus* é capaz de diferenciar indivíduos com eficiência. Tal característica também é observada para este *locus* a nível mundial (Ballantyne *et al.*, 2014).

O *locus* *DYF404S1* também é caracterizado por ser complexo e multialélico, apresentando normalmente dois alelos em cada indivíduo (Kayser *et al.*, 2012). Neste estudo também foram encontrados indivíduos com um ou três alelos e com microvariantes, conforme é mostrado no gráfico 11 da seção 4.3 de Resultados. É possível inferir que este *locus* também é um potencial marcador para discriminação de indivíduos pertencentes à mesma população, conforme a distribuição das frequências haplotípicas mostradas no gráfico 11. Além disso, nota-se que os alelos 14, 15,16 ou 17 se destacam com frequências entre 5 e 10% e, quando tais alelos estão juntos no mesmo haplótipo, como é o caso dos haplótipos 14,16; 15,16 e 15,17 as frequências são ainda maiores, podendo alcançar um valor de quase 30%. Isto

também é observado em populações de diversos países, podendo corroborar o potencial elevado de individualização que este marcador apresenta (Ballantyne *et al.*, 2014).

Os *loci* DYS449 e DYS547 são formados por estruturas repetitivas complexas e são monoalélicos. Os alelos do marcador DYS449 possuem de 25 a 39 repetições, em que se observou o alelo 29 como o mais frequente na amostra analisada, assim como na maioria das populações analisadas de outros países, como, por exemplo, em indivíduos da Índia, Argentina, Nova Zelândia e Itália. Em relação ao marcador DYS547, foram observados alelos com 37 a 47 repetições, com destaque para o alelo 43 que foi encontrado em 25% dos indivíduos. O alelo 43 é o mais encontrado a nível global em diferentes amostras populacionais (Ballantyne *et al.*, 2014).

O *locus* DYS612 é o único marcador RM-YSTR que possui sua estrutura repetitiva formada por três nucleotídeos (Kayser *et al.*, 2012; Ballantyne *et al.*, 2014). Possui alelos que variam de 20 a 30 repetições (190-220pb) e o mais frequente em nossa amostra é o alelo 26 que é encontrado em 30% dos indivíduos. Segundo a literatura recente, este alelo é o mais frequentemente observado na população mundial (Ballantyne *et al.*, 2014).

Ao longo do tempo, conforme novos estudos relacionados aos *loci* RM-YSTRs surgiram, a nomenclatura de seus alelos foi bastante questionada e avaliada pelos diferentes grupos (Kayser *et al.*, 2004; Ballantyne *et al.*, 2010; Kayser *et al.*, 2012).

Durante a realização deste trabalho que, conforme dito anteriormente faz parte de um estudo colaborativo entre laboratórios mundiais (Anexo A), a nomenclatura dos marcadores RM-YSTRs passou por algumas atualizações e foi estabelecida em conformidade com o que determina o guia publicado pela *International Society of Forensic Genetics – ISFG* (Gusmao *et al.*, 2006) e com as variações alélicas encontradas pelos laboratórios participantes do estudo. A nomenclatura final instituída já foi inclusive utilizada pela Promega no kit Powerplex Y23 e pela Thermo Fisher em seu kit Yfiler[®] Plus, nos quais estão incluídos alguns *loci* RM-YSTR.

4.2 Frequência Haplotípica e Poder de Discriminação

Como as regiões STRs do cromossoma Y são transmitidas em conjunto, as frequências haplotípicas observadas para um determinado grupo de marcadores são relacionadas com o seu poder de discriminação.

Devido à natureza altamente polimórfica dos marcadores RM-YSTR, espera-se que estes *loci* apresentem um elevado poder de discriminar indivíduos, e conseqüentemente, baixas frequências haplotípicas.

Tal fato é corroborado em nosso estudo, no qual foi obtido um poder de discriminação de 1,0 ($\pm 1,2190$), o que significa que todos os indivíduos analisados na amostra do Rio de Janeiro para os marcadores RM-YSTR apresentaram haplótipos diferentes, com frequências de 0,0077. Estes resultados estão de acordo com os valores encontrados nos dois estudos de referência para estes *loci*, nos quais foram observados valores próximos ou iguais a 1,0 de poder de discriminação em diferentes populações (Kayser *et al.*, 2012; Ballantyne *et al.*, 2014).

Recentemente, em uma colaboração do LDD-UERJ com laboratórios de diversos países, realizou-se o estudo populacional e mutacional dos marcadores que fazem parte do kit comercial PPY23 em diferentes amostras populacionais, inclusive na amostra do Rio de Janeiro avaliada no presente estudo. Ao comparar os resultados obtidos pela tipagem dessa amostra com os marcadores PPY23 e RM-YSTR, foi possível observar que os *loci rapidly mutating* apresentaram maior capacidade de diferenciar indivíduos do que o conjunto de *loci* PPY23, visto que o maior valor de poder de discriminação encontrado com estes marcadores foi 0,97 (Purps *et al.*, 2014).

Ao avaliar a capacidade de discriminar indivíduos apresentada pelos conjuntos de marcadores amplamente utilizados pela comunidade forense, como o Haplótipo Mínimo, Yfiler e PPY23, observa-se que o poder de discriminação é crescente quando se aumenta o número de marcadores e quando tais *loci* apresentam altas taxas mutacionais. Por isso, o Haplótipo mínimo é considerado o grupo de marcadores que apresenta menor capacidade de discriminar indivíduos em uma população, seguido pelo Yfiler e PPY23. Os RM-YSTRs têm demonstrado ser um conjunto de *loci* mais promissor do que os marcadores do kit PPY23, visto que possui um poder de discriminação maior, conseguindo haplótipos únicos em diversas amostras analisadas, inclusive nos indivíduos avaliados em nosso estudo (Ballantyne *et al.*, 2014; Purps *et al.*, 2014).

4.3 Diversidade Gênica e Haplotípica

4.3.1 Diversidade Gênica

O parâmetro de diversidade gênica por *locus*, parâmetro equivalente à heterozigosidade em autossomos, foi estimado com base na variação alélica observada e utilizado para definir em que proporção cada *locus* contribui para a diferenciação da amostra. Foi observado um elevado nível de diversidade gênica (média de 0,8621) com valores a partir de 0,7936, indicando que estes marcadores possuem características que conferem um alto grau de polimorfismo, que contribui para seu elevado poder de discriminação.

Na amostra de indivíduos do Rio de Janeiro, o *locus* DYF403S1a/b foi o mais polimórfico e discriminativo, seguido pelos *loci* DYF399S1, DYS526a/b e DYF387S1, respectivamente. Embora os demais *loci* também tenham apresentado elevados índices de diversidade gênica, eles se apresentaram menos polimórficos do que os citados.

O valor médio de diversidade gênica encontrado nesta mesma amostra ao utilizar os marcadores do kit PPY23 é de 0,6730. O menor valor de diversidade gênica para os *loci* PPY23 na amostra do Rio de Janeiro foi de 0,5040 e o maior foi de 0,9099 para o *locus* DYS385ab, sendo este um valor muito discrepante da média de diversidade gênica encontrada para o restante dos marcadores deste kit. Em comparação com os marcadores RM-YSTR, somente três *loci* do kit PPY23 (DYS385ab, DYS576, DYS570), incluindo dois marcadores RM-YSTR, apresentaram valores de diversidade gênica maiores ou iguais ao valor mínimo observado para os 13 *rapidly mutating*. Sendo assim, o conjunto de marcadores com maior grau de polimorfismo e capacidade de discriminar indivíduos nesta amostra é o RM-YSTR, devido aos elevados valores de diversidade gênica por *locus*.

Em um trabalho publicado em 2012 (Palha *et al.*, 2012), comparou-se a eficiência de conjuntos de marcadores com crescentes números de STRs ao avaliar as diversidades gênicas apresentadas por eles em diferentes populações do Brasil. Os conjuntos de marcadores avaliados nesse trabalho consistiam no Haplótipo Mínimo (9 *loci*), um conjunto de 17 *loci* incluídos no kit Yfiler, um conjunto de 14 *loci* e outros 23 marcadores propostos por esse grupo. Os valores de diversidade gênica encontrados nas diferentes populações do Brasil para os conjuntos de *loci* utilizados foram similares (variando entre 0,64 e 0,70), e apesar de

elevados, não se aproximam da média de diversidade gênica encontrada para os RM-YSTRs na população alvo de nosso estudo, que é de 0,86.

A justificativa dos marcadores RM-YSTRs possuírem maiores valores de diversidade gênica pode estar relacionada ao fato deste conjunto reunir *loci* multialélicos, com grande número de alelos de frequências bem distribuídas e com valores de diversidade gênica superiores aos observados nos *loci* monoalélicos, o que eleva de maneira considerável a diversidade gênica encontrada para estes *loci* e, conseqüentemente, também aumenta o poder de discriminação destes marcadores em relação aos demais.

4.3.2 Diversidade Haplotípica

A diversidade haplotípica correspondente a um conjunto de marcadores está associada às frequências dos haplótipos. Como as frequências dos haplótipos encontradas na amostra analisada neste estudo são baixas e estão bem distribuídas, verificou-se que a diversidade haplotípica encontrada após emprego dos *loci* RM-YSTR foi de 1,0 ($\pm 1,2190$), demonstrando que tais *loci* são muito eficientes em discriminar indivíduos.

O resultado encontrado em nosso trabalho corrobora dados da literatura, nos quais os valores de diversidade haplotípica verificados em populações ao usar os 13 *loci* RM-YSTR são muito próximos ou iguais a 1,0 (Kayser *et al.*, 2012; Ballantyne *et al.*, 2014).

Quando comparados com outros marcadores, os RM-YSTR demonstram ser mais eficientes em diferenciar indivíduos, com elevados valores de diversidade haplotípica e poder de discriminação. Em comparação com os marcadores Yfiler, os *loci rapidly mutating* são mais informativos, sendo possível observar um aumento do número de haplótipos encontrados e uma queda no número de haplótipos compartilhados entre os indivíduos do mesmo grupo regional ou da mesma população e também entre indivíduos de populações diferentes. Mesmo ao retirar da análise os *loci* multialélicos, há um menor compartilhamento de haplótipos RM-YSTR de cópia única do que se observa nos marcadores Yfiler sem o *locus* *DYS385ab* (Kayser *et al.*, 2012; Ballantyne *et al.*, 2014).

No trabalho de Purps e colaboradores (Purps *et al.*, 2014) a tipagem dos indivíduos com os marcadores PPY23 gerou 92,9% de haplótipos únicos, o que é considerado um valor alto, porém ainda baixo quando comparado com os haplótipos únicos gerados pelos RM-YSTR (100%).

4.4 Distância Genética

A amostra de indivíduos do Rio de Janeiro foi comparada com outras 12 populações distribuídas pelos continentes americano, africano, europeu e asiático, conforme mostrado na tabela 6 da seção 2.4.5 de Materiais e Métodos. Foram avaliados o nível de diversidade da amostra alvo deste estudo e o seu grau de diferenciação em relação às outras populações, através da estimativa dos índices de F_{st} e AMOVA por meio do *software* Arlequin 3.5 (Excoffier & Lischer, 2010). É importante ressaltar que há um parâmetro análogo ao F_{st} , que também é amplamente utilizado para estimar a diversidade, conhecido como índice R_{st} . Este considera as taxas de mutação dos *loci* microssatélites e é caracterizado por estimar as diferenças entre haplótipos formados por *loci* monoalélicos (Schneider *et al.*, 2000). Por isso, optou-se por utilizar o índice F_{st} para a estimativa da diversidade neste estudo, tendo em vista que os haplótipos verificados são formados por alguns *loci* multialélicos.

De acordo com as tabelas 9, 10 e 11, é possível observar que as 12 amostras avaliadas são diferentes entre si, com exceção das amostras populacionais de Buenos Aires e Costa Rica, além disso, também se observa que, em todas elas, as frequências dos alelos para os marcadores RM-YSTR são similares, indicando que a maior parte da variação genética não está entre as amostras populacionais. De fato, os resultados de AMOVA (tabela 12) corroboram os resultados de F_{st} , visto que a maior fonte de variação é observada dentro das populações (cujo percentual de variação é de 94,83%), seguida pela variação entre populações da mesma região (3,19%) e entre grupos regionais (1,98%). Os resultados encontrados neste estudo, são bastante similares aos observados pelo grupo de Ballantyne em seu recente trabalho (Ballantyne *et al.*, 2014), os quais também evidenciaram que a maior diferenciação genética foi constatada entre indivíduos da mesma população. Neste trabalho, o grupo ainda sugere que, devido às elevadas taxas mutacionais dos *loci* RM-YSTR, não é possível perceber nenhum traço ancestral comum entre as populações através da análise dos haplótipos gerados por estes marcadores.

4.5 Estudo mutacional

Marcadores STR do cromossoma Y podem ser utilizados em estudos populacionais, evolutivos e em investigações forenses. O uso destes *loci* microssatélites para tais finalidades depende do conhecimento e da correta interpretação de suas frequências haplotípicas e taxas mutacionais em diferentes populações. O estudo dos microssatélites e de suas taxas mutacionais foi iniciado a década de 90 e ainda se demonstra necessário na atualidade, mesmo que alguns *loci* já tenham sido bastante explorados, como é o caso dos marcadores que constituem o Haplótipo Mínimo e o Yfiler, por exemplo (Buttler, 2003; Palha *et al.*, 2012; Ballantyne *et al.*, 2014).

Em 2010, Ballantyne e colaboradores investigaram um elevado número de STR focando em suas características mutacionais, descrevendo pela primeira vez o conjunto de 13 *loci* RM-YSTR, caracterizados por apresentarem taxas mutacionais superiores às normalmente encontradas em *loci* STR (Ballantyne *et al.*, 2010).

Com o intuito de analisar as características mutacionais destes *loci* na amostra populacional de indivíduos miscigenados do Rio de Janeiro, as taxas mutacionais destes 13 marcadores RM-YSTR foram verificadas em 129 pares de pais e filhos cujos vínculos genéticos de paternidade foram previamente confirmados em investigações de paternidade realizadas no LDD-UERJ.

Nesta amostra populacional foram encontradas 30 mutações distribuídas em 11 dos 13 *loci* analisados, com exceção dos marcadores DYS518 e DYS626. As mutações foram confirmadas através da repetição da tipagem das amostras nas quais foram verificadas, além disso, a confirmação também ocorreu por meio do sequenciamento automático dos alelos mutados.

A taxa de mutação média encontrada na amostra de indivíduos miscigenados analisados neste estudo foi de $2,11 \times 10^{-2}$, considerada elevada para STR do cromossoma Y, que normalmente apresentam taxas de mutação na ordem de 10^{-3} (Ballantyne *et al.*, 2010; Kayser *et al.*, 2012). Tal fato pode ser observado em trabalhos realizados no LDD-UERJ com indivíduos miscigenados e afrodescendentes do Rio de Janeiro, genotipados através dos kits comerciais PowerPlex Y e Y-Plex 12, nos quais foram encontradas taxas de mutações médias de $3,96 \times 10^{-3}$ e de $1,8 \times 10^{-3}$, respectivamente (De Souza Góes *et al.*, 2005; 2006, Domingues & Carvalho, 2006). Estima-se que a probabilidade de observar pelo menos uma mutação em um ou mais *loci* com os marcadores RM-YSTR é de 19,5% e esta probabilidade diminui em

outros conjuntos de marcadores, como é o caso do Yfiler, cuja probabilidade é de 4,7%, e do PPY23 com probabilidade de 9,2% (Ballantyne *et al.*, 2010; Ballantyne *et al.*, 2014).

Embora todos os *loci* RM-YSTR tenham apresentado altas taxas mutacionais, o *locus* DYF399S1 se destacou com a maior taxa de mutação dentre os 13 marcadores ($7,75 \times 10^{-2}$), seguido pelo marcador DYF403a/b com uma taxa de $4,65 \times 10^{-2}$. Os *loci* DYS576, DYS526a/b, DYS449 e DYS612 possuem a terceira maior taxa de mutação, conforme pode ser observado na tabela 13 da seção 4.6.2 de Resultados. Nossos resultados corroboram os dados mutacionais encontrados por Ballantyne e colaboradores em seu recente estudo, já que as taxas de mutação encontradas na amostra do Rio de Janeiro são semelhantes às descritas pelo grupo de Ballantyne (Ballantyne *et al.*, 2014). Este fato confirma mais uma vez que este conjunto de *loci* é eficientemente capaz de discriminar indivíduos e sugere que sua eficiência em fazê-lo é a mesma independente da população ou região em que são aplicados.

Uma mutação em uma região STR pode ser caracterizada por gerar a perda ou o ganho de apenas uma unidade de repetição *in tandem*, sendo comum os trabalhos avaliarem as proporções de perdas e ganhos de unidades repetitivas observadas entre pares de indivíduos de mesma linhagem paterna nos quais ocorreram eventos mutacionais. Em 2010, Ballantyne e colaboradores desenvolveram um estudo bem consistente acerca das características mutacionais dos *loci* RM-YSTR e se constatou que, de uma maneira geral, estes marcadores tinham como característica mutacional predominante um maior número de perdas de unidades repetitivas do que de ganhos (Ballantyne *et al.*, 2010). Tal característica não foi verificada na amostra analisada por nosso estudo, no qual foram observadas 12 perdas e 18 ganhos de unidades repetitivas ao avaliar os duos de pais e filhos.

As excepcionais características mutacionais dos *loci* RM-YSTR e sua elevada capacidade de discriminar indivíduos sugerem que tais marcadores possam ser utilizados para diferenciar indivíduos de mesma linhagem patrilínea, podendo ter aplicações diretas na área forense. A capacidade de estes *loci* diferenciarem duos de pais e filhos foi testada em nossa amostra populacional e os resultados obtidos evidenciaram que, dentre os 129 pares de pais e filhos, os *loci* RM-YSTR foram capazes de diferenciar 28 pares. Outros estudos também mostraram que os *rapidly mutating* conseguem diferenciar homens de mesma linhagem paterna ao comparar duos de pais e filhos, irmãos e tios e primos (Ballantyne *et al.*, 2010; Kayser *et al.*, 2012; Ballantyne *et al.*, 2014). Em comparação com os marcadores Yfiler, os *loci* RM-YSTR têm capacidade de diferenciação seis vezes maior que os Yfiler entre pais e filhos e, cinco vezes maior entre irmãos (Ballantyne *et al.*, 2014). Posto isso, é válido ressaltar que o número de indivíduos avaliados em nosso estudo pode não ser grande o suficiente para

obter capacidade de diferenciação na mesma proporção obtida nos estudos citados, cujo número de amostras era elevado. Sendo assim, a necessidade de estudos na população de indivíduos miscigenados com um número amostral maior se torna relevante para verificar se a proporção de diferenciação seria equivalente à encontrada na literatura.

CONCLUSÃO

Portanto, é possível concluir que:

- a) Os resultados obtidos no presente estudo mostraram que estes marcadores possuem elevadas taxas de mutação na amostra miscigenada analisada;
- b) Os 13 marcadores apresentam um elevado poder de discriminação;
- c) Os *loci* RM-YSTR são capazes de diferenciar pares de pai e filho na amostra estudada, porém, tal capacidade não foi observada para todos os indivíduos aparentados analisados;
- d) Em comparação com outros conjuntos de marcadores normalmente empregados em estudos populacionais e análises forenses, os *loci* RM-YSTR possuem maior capacidade de diferenciar indivíduos, observando-se poucos ou nenhum haplótipo compartilhado na população.

Posto isso, os 13 *loci* RM-YSTR são muito promissores, visto que devido às suas elevadas taxas mutacionais, são capazes de diferenciar indivíduos mais eficientemente do que outros conjuntos de STRs. Apesar disso, sabe-se que o somente o emprego dos RM-YSTR ainda não é suficiente para diferenciar com total precisão indivíduos com parentesco paterno. Sendo assim, mais estudos são necessários para um melhor conhecimento das características desses marcadores em diferentes populações e, usá-los juntamente com outros marcadores STR já conhecidos pode ser uma boa alternativa, que irá tornar os resultados obtidos mais robustos e precisos.

REFERÊNCIAS

Ali S ; Hasnain S E. Molecular dissection of the human Y-chromosome. *Gene* 2002, 283 (1-2): 1-10 p.

Armour J A. Tandemly repeated DNA: why should anyone care? *Mutat Res* 2006, 598 (1-2): 6-14 p.

Bachtrog D. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nat Rev Genet* 2013, 14 (2): 113-124 p.

Ballantyne K N, *et al.* Mutability of Y-chromosomal microsatellites: rates, characteristics, molecular bases, and forensic implications. *Am J Hum Genet* 2010, 87 (3): 341-353 p.

Ballantyne K N, *et al.* Toward male individualization with rapidly mutating y-chromosomal short tandem repeats. *Hum Mutat* 2014, 35 (8): 1021-1032 p.

Bird S C. Towards improvements in the estimation of the coalescent: implications for the most effective use of Y chromosome short tandem repeat mutation rates. *PLoS One* 2012, 7 (10): e48638 p.

Bjerkem M, *et al.* Forensic application of the Maxwell™ 16 Instrument. *Profiles in DNA* 2006, 9 (1): 3-5 p.

Bull L N; Pabon-Pena C R ; Freimer N B. Compound microsatellite repeats: practical and theoretical features. *Genome Res* 1999, 9 (9): 830-838 p.

Burgarella C ; Navascues M. Mutation rate estimates for 110 Y-chromosome STRs combining population and father-son pair data. *Eur J Hum Genet* 2011, 19 (1): 70-75 p.

Butler J. Recent developments in Y-short tandem repeat and Y-single nucleotide polymorphism analysis. *analysis* 2003, 15 (91).

Butler J M. *Fundamentals of forensic DNA typing*: Academic Press. 2009

Butler J M. *Advanced Topics in Forensic DNA Typing: Methodology: Methodology*: Elsevier Science. 2011

Calderon S, *et al.* Phylogenetic and forensic studies of the Southeast Florida Hispanic population using the next-generation forensic PowerPlex® Y23 STR marker system. *Legal Medicine* 2013, 15 (6): 289-292 p.

Davis C, *et al.* Prototype PowerPlex® Y23 System: A concordance study. *Forensic Science International: Genetics* 2013, 7 (1): 204-208 p.

De Souza Góes A C, *et al.* Population and mutation analysis of 17 Y-STR loci from Rio de Janeiro (Brazil). *International journal of legal medicine* 2005, 119 (2): 70-76 p.

Domingues P M G D R P ; Carvalho E F D. *Contribuição ao estudo da constituição genética da população brasileira : análise de regiões STR do cromossomo Y em uma amostra populacional de indivíduos afro-descendentes do Rio de Janeiro / Patrícia Mariana Gonçalves da Rocha Porto Domingues*: 2006. 2006

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

Excoffier L; Smouse P E ; Quattro J M. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 1992, 131 (2): 479-491 p.

Gusmao L, *et al.* DNA Commission of the International Society of Forensic Genetics (ISFG): an update of the recommendations on the use of Y-STRs in forensic analysis. *Int J Legal Med* 2006, 120 (4): 191-200 p.

Gusmao L ; Carracedo A. Y chromosome-specific STRs. *Profiles in DNA* 2003, 6 (1): 3-6 p.

Haplotype Reference Databases. Disponível em: <<http://yhrd.org/>>. Acesso em: março, 2015.

An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012, 489 (7414): 57-74 p.

Jeffreys A J, *et al.* Human minisatellites, repeat DNA instability and meiotic recombination. *Electrophoresis* 1999, 20 (8): 1665-1675 p.

Jobling M A. The impact of recent events on human genetic diversity. *Philosophical Transactions of the Royal Society B: Biological Sciences* 2012, 367 (1590): 793-799 p.

Jobling M A ; Tyler-Smith C. Fathers and sons: the Y chromosome and human evolution. *Trends in Genetics* 1995, 11 (11): 449-456 p.

Jobling M A ; Tyler-Smith C. New uses for new haplotypes: the human Y chromosome, disease and selection. *Trends in Genetics* 2000, 16 (8): 356-362 p.

Kayser M, *et al.* A new future of forensic Y-chromosome analysis: rapidly mutating Y-STRs for differentiating male relatives and paternal lineages. *Forensic Sci Int Genet* 2012, 6 (2): 208-218 p.

Kayser M, *et al.* A comprehensive survey of human Y-chromosomal microsatellites. *Am J Hum Genet* 2004, 74 (6): 1183-1197 p.

Kayser M, *et al.* 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* 2000, 66 (5): 1580-1588 p.

Kayser M & Sajantila A. Mutations at Y-STR loci: implications for paternity testing and forensic analysis. *Forensic Sci Int* 2001, 118 (2-3): 116-121 p.

Kittler R, *et al.* Apparent intrachromosomal exchange on the human Y chromosome explained by population history. *Eur J Hum Genet* 2003, 11 (4): 304-314 p.

Kovacevic L, *et al.* Haplotype data for 23 Y-chromosome markers in a reference sample from Bosnia and Herzegovina. *Croat Med J* 2013, 54 (3): 286-290 p.

Mitchell R J ; Hammer M F. Human evolution and the Y chromosome. *Curr Opin Genet Dev* 1996, 6 (6): 737-742 p.

Mulero J, *et al.* Nomenclature update and allele repeat structure for the markers DYS518 and DYS449. *Forensic Science International: Genetics* 2014.

Oliveira A M, *et al.* Male lineage strata of Brazilian population disclosed by the simultaneous analysis of STRs and SNPs. *Forensic Sci Int Genet* 2014, 13): 264-268 p.

Palha T, *et al.* Disclosing the Genetic Structure of Brazil through Analysis of Male Lineages with Highly Discriminating Haplotypes. *PLoS One* 2012, 7 (7): e40007 p.

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

Quintana-Murci L S; Krausz C ; Mcelreavey K. The human Y chromosome: function, evolution and disease. *Forensic science international* 2001, 118 (2-3): 169-181 p.

Sambrook J. F E, Maniatis, T. *Molecular cloning: a laboratory manual*/J. Sambrook, EF Fritsch, T. Maniatis: New York: Cold Spring Harbor Laboratory Press. 1989

Santos F R, *et al.* A polymorphic L1 retroposon insertion in the centromere of the human Y chromosome. *Human Molecular Genetics* 2000, 9 (3): 421-430 p.

Schneider S; Roessli D ; Excoffier L. Arlequin: A software for population genetics data analysis. *User manual version. Genetics and Biometry Laboratory, University of Geneva, Switzerland* 2000.

Singh N P, *et al.* Epigenetic profile of the euchromatic region of human Y chromosome. *Nucleic Acids Res* 2011, 39 (9): 3594-3606 p.

Thompson J M, *et al.* Developmental validation of the PowerPlex® Y23 System: A single multiplex Y-STR analysis system for casework and database samples. *Forensic Science International: Genetics* 2013, 7 (2): 240-250 p.

Underhill P A, *et al.* Y chromosome sequence variation and the history of human populations. *Nat Genet* 2000, 26 (3): 358-361 p.

Vermeulen M, *et al.* Improving global and regional resolution of male lineage differentiation by simple single-copy Y-chromosomal short tandem repeat polymorphisms. *Forensic Sci Int Genet* 2009, 3 (4): 205-213 p.

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* 1991, 10 (4): 506-513 p.

ANEXO A – Toward Male Individualization with Rapidly Mutating Y-Chromosomal Short Tandem Repeats

RESEARCH ARTICLE

Human Mutation

Toward Male Individualization with Rapidly Mutating Y-Chromosomal Short Tandem Repeats



Kaye N. Ballantyne,^{1,2†} Arwin Ralf,^{1†} Rachid Aboukhalid,³ Niaz M. Achakzai,⁴ Maria J. Anjos,⁵ Qasim Ayub,⁶ Jože Balazic,⁷ Jack Ballantyne,^{8,9} David J. Ballard,¹⁰ Burkhard Berger,¹¹ Cecilia Bobillo,¹² Mehdi Bouabdellah,³ Helen Burri,¹³ Tomas Capal,¹⁴ Stefano Caratti,¹⁵ Jorge Cárdenas,¹⁶ François Cartault,¹⁷ Elizeu F. Carvalho,¹⁸ Monica Carvalho,⁵ Baowen Cheng,¹⁹ Michael D. Coble,²⁰ David Comas,²¹ Daniel Corach,¹² Maria E. D'Amato,²² Sean Davison,²² Peter de Knijff,²³ Maria Corazon A. De Ungria,²⁴ Ronny Decorte,²⁵ Tadeusz Dobosz,²⁶ Berit M. Dupuy,²⁷ Samir Elmrghni,²⁸ Mateusz Gliwiński,²⁹ Sara C. Gomes,³⁰ Laurens Grol,³¹ Cordula Haas,¹³ Erin Hanson,⁹ Jürgen Henke,³² Lotte Henke,³² Fabiola Herrera-Rodríguez,³³ Carolyn R. Hill,²⁰ Gunilla Holmlund,^{34,35} Katsuya Honda,³⁶ Uta-Dorothee Immel,³⁷ Shota Inokuchi,³⁸ Mark A. Jobling,³⁹ Mahmoud Kaddura,²⁸ Jong S. Kim,⁴⁰ Soon H. Kim,⁴¹ Wook Kim,⁴² Turi E. King,³⁹ Eva Klausriegler,⁴³ Daniel Kling,²⁷ Lejla Kovačević,⁴⁴ Leda Kovatsi,⁴⁵ Paweł Krajewski,⁴⁶ Sergey Kravchenko,⁴⁷ Maarten H. D. Larmuseau,²⁵ Eun Young Lee,⁴⁸ Ruediger Lessig,³⁷ Ludmila A. Livshits,⁴⁷ Damir Marjanović,⁴⁴ Marek Minarik,⁴⁹ Natsuko Mizuno,³⁸ Helena Moreira,⁵⁰ Niels Morling,⁵¹ Meeta Mukherjee,⁵² Patrick Munier,¹⁷ Javaregowda Nagaraju,^{53†} Franz Neuhuber,⁴³ Shengjie Nie,⁵⁴ Premlephat Nilasitsatporn,⁵⁵ Takeki Nishi,³⁶ Hye H. Oh,⁴⁰ Jill Olofsson,⁵¹ Valerio Onofri,⁵⁶ Jukka U. Palo,⁵⁷ Horolma Pamjav,⁵⁸ Walther Parson,^{11,59} Michal Petlach,⁴⁹ Christopher Phillips,¹⁶ Rafal Ploski,⁶⁰ Samayamantri P. R. Prasad,⁵³ Dragan Primorac,^{59,61,62} Gludhug A. Purnomo,⁶³ Josephine Purps,⁶⁴ Hector Rangel-Villalobos,⁶⁵ Krzysztof Rebała,²⁹ Budsaba Rerkamnuaychoke,⁶⁶ Danel Rey Gonzalez,¹⁶ Carlo Robino,¹⁵ Lutz Roewer,⁶⁴ Alexandra Rosa,^{30,67} Antti Sajantila,^{57,68} Andrea Sala,¹² Jazelyn M. Salvador,²⁴ Paula Sanz,²¹ Cornelia Schmitt,⁶⁹ Anil K. Sharma,⁵² Dayse A. Silva,¹⁸ Kyoung-Jin Shin,⁴⁸ Titia Sijen,³¹ Miriam Sirker,⁶⁹ Daniela Siváková,⁷⁰ Vedrana Škaro,⁷¹ Carlos Solano-Matamoros,⁷² Luis Souto,⁵⁰ Vlastimil Stenzl,¹⁴ Herawati Sudoyo,⁶³ Denise Syndercombe-Court,¹⁰ Adriano Tagliabracci,⁵⁶ Duncan Taylor,⁷³ Andreas Tillmar,^{34,35} Iosif S. Tsybovsky,⁷⁴ Chris Tyler-Smith,⁶ Kristiaan J. van der Gaag,²³ Daniel Vanek,^{75,76} Antónia Völgyi,⁵⁸ Denise Ward,⁷⁷ Patricia Willemse,²³ Eric P. H. Yap,⁷⁸ Rita Y. Y. Yong,⁷⁸ Irena Zupanič Pajnič,⁷ and Manfred Kayser^{1*}

¹Department of Forensic Molecular Biology, Erasmus MC University Medical Centre Rotterdam, Rotterdam, The Netherlands; ²Office of the Chief Forensic Scientist, Victoria Police Forensic Services Department, Macleod, Victoria, Australia; ³Forensic Genetic Unit, Immunology and Biochemistry Laboratory, Faculty of Sciences, Mohammed V Agdal University, Rabat, Morocco; ⁴DNA Typing Laboratory, Centre of Excellence in Molecular Biology, CEMB, University of the Punjab, Lahore, Pakistan; ⁵Forensics Genetic Service, Central Delegation, National Institute of Legal Medicine and Forensic Sciences, I.P., Coimbra, Portugal; ⁶The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, South Cambridgeshire, UK; ⁷Institute of Forensic Medicine, Faculty of Medicine, University of Ljubljana, Ljubljana, Slovenia; ⁸Department of Chemistry, University of Central Florida, Orlando, Florida; ⁹National Center for Forensic Science, University of Central Florida, Orlando, Florida; ¹⁰King's College London, Department of Forensic and Analytical Science, London, UK; ¹¹Institute of Legal Medicine, Innsbruck Medical University, Innsbruck, Austria; ¹²Servicio de Huellas Digitales Genéticas, Facultad de Farmacia y Bioquímica, Universidad de Buenos Aires and CONICET, Buenos Aires, Argentina; ¹³Institute of Legal Medicine, University of Zurich, Zurich, Switzerland; ¹⁴Department of Forensic Genetics, Institute of Criminalistics Prague, Prague, Czech Republic; ¹⁵Department of Public Health Sciences and Pediatrics, University of Turin, Turin, Italy; ¹⁶Forensic Genetics Unit, Institute of Legal Medicine, University of Santiago de Compostela, Santiago de Compostela, Spain; ¹⁷Service de Genetique, Site Centre Hospitalier Felix Guyon, Saint-Denis, Reunion; ¹⁸DNA Laboratory, Department of Ecology, Biology Institute, State University of Rio de Janeiro, Rio de Janeiro, Brazil; ¹⁹Institute of Forensic Science, Yunnan Provincial Department of Public Security, Kunming, China; ²⁰National Institute of Standards and Technology, Applied Genetics Group, Gaithersburg, Maryland; ²¹Institut de Biologia Evolutiva (CSIC-UPF), Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra, Barcelona, Spain; ²²Forensic DNA Lab, Department of Biotechnology, University of the Western Cape, Cape Town, South Africa; ²³Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands; ²⁴DNA Analysis Laboratory, Natural Sciences Research Institute, University of the Philippines, Diliman, Quezon City, Philippines; ²⁵Laboratory of Forensic Genetics and Molecular Archaeology, Department of Imaging and Pathology, KU Leuven, Leuven, Belgium; ²⁶Department of Forensic Medicine, Wrocław Medical University, Wrocław, Poland; ²⁷Department of Family Genetics, The Norwegian Institute of Public Health, Oslo, Norway;

Additional Supporting Information may be found in the online version of this article.

† Deceased.

† These authors contributed equally to this work

* Correspondence to: Manfred Kayser, Department of Forensic Molecular Biology, Erasmus MC University Medical Center Rotterdam, P.O. Box #2040, Rotterdam 3000 CA, The Netherlands. E-mail: m.kayser@erasmusmc.nl

Contract grant sponsors: Netherlands Genomics Initiative (NGI)/Netherlands Organization for Scientific Research (NWO) within the framework of the Forensic Genomics Consortium Netherlands (FGCN); The Wellcome Trust (098051); The Wellcome Trust (087576); Leverhulme Trust (F/00 212/AM); ALPKULTUR and the Austrian Academy of Sciences' "Alpine Research Programme"; Spanish Government (CGL2010-14944/BOS); UBACYT 20020100100744 and PIP CONICET N°1220110100914, Argentina; FWO-Vlaanderen; National Natural Science Foundation of China (31100906 and 81241136); Medical Research Council of South Africa and the National Research Foundation of South Africa; Philippine Council for Industry, Energy and Emerging Technology Research and Development (PCIERD) of the Department of Science and Technology (DOST); Finnish Foundations' Professors Pool, Paulo Foundation.

© 2014 The Authors. **Human Mutation published by Wiley Periodicals, Inc.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

²⁸ Department of Forensic Medicine and Toxicology, Faculty of Medicine, University of Benghazi, Benghazi, Libya; ²⁹ Department of Forensic Medicine, Medical University of Gdansk, Gdansk, Poland; ³⁰ Human Genetics Laboratory, University of Madeira, Funchal, Portugal; ³¹ Department of Human Biological Traces (R&D), Netherlands Forensic Institute, The Hague, The Netherlands; ³² Institut für Blutgruppenforschung LGC GmbH, Cologne, Germany; ³³ Unidad de Genética Forense, Sección de Bioquímica del Departamento de Ciencias Forenses, Poder Judicial, San Joaquín de Flores, Heredia, Costa Rica; ³⁴ Department of Clinical and Experimental Medicine, Faculty of Health Sciences, Linköping University, Linköping, Sweden; ³⁵ Department of Forensic Genetics and Forensic Toxicology, National Board of Forensic Medicine, Linköping, Sweden; ³⁶ Department of Legal Medicine, Faculty of Medicine, University of Tsukuba, Tsukuba City, Japan; ³⁷ Institute of Legal Medicine, Martin-Luther University Halle, Halle, Germany; ³⁸ National Research Institute of Police Science, Kashiwa, Japan; ³⁹ Department of Genetics, University of Leicester, Leicester, UK; ⁴⁰ DNA Forensic Division, Supreme Prosecutors' Office, Seoul, South Korea; ⁴¹ DNA Analysis Laboratory, Eastern District Office, National Forensic Service, Gangwon-do, South Korea; ⁴² Department of Biological Sciences, Dankook University, Cheonan, South Korea; ⁴³ Institute of Legal Medicine, University of Salzburg, Salzburg, Austria; ⁴⁴ Institute for Genetic Engineering and Biotechnology, Sarajevo, Bosnia and Herzegovina; ⁴⁵ Laboratory of Forensic Medicine and Toxicology, School of Medicine, Aristotle University of Thessaloniki, Thessaloniki, Greece; ⁴⁶ Department of Forensic Medicine, Centre for Biostructure, Medical University of Warsaw, Warsaw, Poland; ⁴⁷ Department of Human Genomics, Institute of Molecular Biology and Genetics NASU, Kiev, Ukraine; ⁴⁸ Department of Forensic Medicine, Yonsei University College of Medicine, Seoul, South Korea; ⁴⁹ Genomac Forensic Institute, Prague, Czech Republic; ⁵⁰ Laboratório de Genética Aplicada, Departamento de Biologia, Universidade de Aveiro, Aveiro, Portugal; ⁵¹ Section of Forensic Genetics, Department of Forensic Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark; ⁵² Central Forensic Science Laboratory, Directorate of Forensic Science Services, Ministry of Home Affairs, Government of India, Kolkata, India; ⁵³ Laboratory of DNA Fingerprinting Services, Centre for DNA Fingerprinting and Diagnostics, Hyderabad, India; ⁵⁴ School of Forensic Medicine, Kunming Medical University, Kunming, China; ⁵⁵ DNA and Biology Unit, Central Scientific Crime Detection Division, Office of Forensic Science Police, Royal Thai Police, Bangkok, Thailand; ⁵⁶ Section of Legal Medicine, Department of Biomedical Science and Public Health, Università Politecnica delle Marche, Ancona, Italy; ⁵⁷ Department of Forensic Medicine, Hjelt Institute, University of Helsinki, Helsinki, Finland; ⁵⁸ DNA Laboratory, Institute of Forensic Medicine, Network of Forensic Science Institutes, Budapest, Ministry of Public Administration and Justice, Budapest, Hungary; ⁵⁹ Eberly College of Science, The Pennsylvania State University, University Park, Pennsylvania; ⁶⁰ Department of Medical Genetics, Centre for Biostructure, Medical University of Warsaw, Warsaw, Poland; ⁶¹ The Henry C. Lee College of Criminal Justice and Forensic Sciences, University of New Haven, West Haven, Connecticut; ⁶² University of Split, Medical School, Split and University of Osijek, Medical School, Osijek, Croatia; ⁶³ Eijkman Institute for Molecular Biology, Jakarta, Indonesia; ⁶⁴ Institute of Legal Medicine and Forensic Sciences, Charité – Universitätsmedizin Berlin, Berlin, Germany; ⁶⁵ Instituto de Investigación en Genética Molecular, Universidad de Guadalajara (CUCIenega-UdeG), Jalisco, México; ⁶⁶ Department of Pathology, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, Thailand; ⁶⁷ Medical Sciences Unit, Center of Life Sciences, University of Madeira, Funchal, Portugal; ⁶⁸ Institute of Applied Genetics, Department of Molecular and Medical Genetics, University of North Texas Health Science Center, Fort Worth, Texas; ⁶⁹ Institute of Legal Medicine, Faculty of Medicine, University of Cologne, Cologne, Germany; ⁷⁰ Department of Anthropology, Comenius University in Bratislava, Bratislava, Slovakia; ⁷¹ DNA Laboratory, Genos Ltd., Zagreb, Croatia; ⁷² Facultad de Microbiología, Universidad de Costa Rica, San Pedro de Montes de Oca, San José, Costa Rica; ⁷³ Forensic Science South Australia, Adelaide, Australia and School of Biological Sciences, Flinders University, Adelaide, Australia; ⁷⁴ Scientific and Practical Centre of the State Committee of Forensic Expertises, Minsk, Belarus; ⁷⁵ Forensic DNA Service, Prague, Czech Republic; ⁷⁶ 2nd Medical Faculty, Institute of the Legal Medicine, Charles University, Prague, Czech Republic; ⁷⁷ Forensic Science South Australia, Adelaide, Australia; ⁷⁸ Defence Medical and Environmental Research Institute, DSO National Laboratories, Singapore

Communicated by Peter J. Oefner

Received 25 February 2014; accepted revised manuscript 26 May 2014.

Published online 10 June 2014 in Wiley Online Library (www.wiley.com/humanmutation). DOI: 10.1002/humu.22599

ABSTRACT: Relevant for various areas of human genetics, Y-chromosomal short tandem repeats (Y-STRs) are commonly used for testing close paternal relationships among individuals and populations, and for male lineage identification. However, even the widely used 17-loci Yfiler set cannot resolve individuals and populations completely. Here, 52 centers generated quality-controlled data of 13 rapidly mutating (RM) Y-STRs in 14,644 related and unrelated males from 111 worldwide populations. Strikingly, >99% of the 12,272 unrelated males were completely individualized. Haplotype diversity was extremely high (global: 0.9999985, regional: 0.99836–0.999988). Haplotype sharing between populations was almost absent except for six (0.05%) of the 12,156 haplotypes. Haplotype sharing within populations was generally rare

(0.8% nonunique haplotypes), significantly lower in urban (0.9%) than rural (2.1%) and highest in endogamous groups (14.3%). Analysis of molecular variance revealed 99.98% of variation within populations, 0.018% among populations within groups, and 0.002% among groups. Of the 2,372 newly and 156 previously typed male relative pairs, 29% were differentiated including 27% of the 2,378 father–son pairs. Relative to Yfiler, haplotype diversity was increased in 86% of the populations tested and overall male relative differentiation was raised by 23.5%. Our study demonstrates the value of RM Y-STRs in identifying and separating unrelated and related males and provides a reference database.

Hum Mutat 35:1021–1032, 2014. Published 2014 Wiley Periodicals, Inc.**

KEY WORDS: Y-chromosome; Y-STRs; haplotypes; RM Y-STRs; paternal lineage; forensic

Introduction

Genetic characterization of male individuals and populations by means of Y-chromosome DNA polymorphisms is relevant in various fundamental and applied areas of human genetics such as in evolutionary genetics and population history, for example, for modeling global and regional human evolution, mapping migration patterns across the globe, and tracking cultural and demographic factors such as patrilocality, extrapair paternity, endogamy, and polygyny. Further, Y-chromosome DNA analysis is important in genetic genealogy and for community genetic purposes such as personal ancestry identification, as well as for the identification of male lineages and inferring paternal genetic ancestry for judicial and investigative purposes [Kayser et al., 1997; Underhill et al., 2000; Hammer et al., 2001; Oota et al., 2001; Jobling and Tyler-Smith, 2003; Roewer et al., 2005; Shi et al., 2010]. Similarities at Y-chromosome DNA markers are usually interpreted as indicating shared paternal ancestry of individuals and populations, whereas differences are used to conclude the absence of close paternal relationships. Such interpretations, however, depend in part on the underlying mutation rates of the Y-DNA markers used. Slowly evolving Y-chromosomal single-nucleotide polymorphisms (Y-SNPs), with an average mutation rate of about 3×10^{-8} per nucleotide per generation [Xue et al., 2009; Poznik et al., 2013], are especially suitable for studying distant relationships between male individuals and populations [Underhill et al., 2000; van Oven et al., 2014]. More quickly evolving Y-chromosomal short tandem repeat (Y-STRs) polymorphisms, also referred to as Y-microsatellites, with an average mutation rate of about 10^{-3} per locus per generation [Goedbloed et al., 2009; Ballantyne et al., 2010], have proven useful for testing short to medium timescale paternal relationships [Kayser et al., 2003; Kayser et al., 2005; Roewer et al., 2005; Coble et al., 2009; van Oven et al., 2011] such as enabling male lineages to be connected through common paternal ancestry [Coble et al., 2009], for the same lineages be separated and individualized within that shared ancestral lineage [King and Jobling, 2009], and for the origins of entire population groups be elucidated [Parkin et al., 2007; Rębała et al., 2007]. Forensic usage of Y-STRs has largely focused on identifying paternal lineages using a core set of markers, linking suspects, and crime samples for investigation purposes [Kayser et al., 1997; Roewer, 2009].

Haplotypes generated from conventional Y-STRs, such as the widely used 17 markers included in the commercially available AmpFISTR® Yfiler® PCR Amplification Kit (Life Technologies, San Francisco, CA) (subsequently referred to as Yfiler), suffer from two main limitations: (1) their inability to conclusively resolve some male lineages due to identical haplotypes arising in individuals that are not of common descent because of recurrent mutation, and (2) their inability to differentiate between paternally related males due to the moderately low mutation rate of the loci tested. In general, Y-STRs with much higher mutation rates than those of the conventionally used loci are expected to overcome or at least reduce both limitations. In a previous comprehensive Y-STR mutation rate study where nearly 200 Y-STRs were investigated in almost 2,000 father–son pairs confirmed by autosomal DNA analysis [Ballantyne et al., 2010], 13 Y-STRs with exceptionally high ($>10^{-2}$ per locus per generation) mutation rates were identified and termed rapidly mutating (RM) Y-STRs. Furthermore, in this previous study and a subsequent study [Ballantyne et al., 2012], theoretical and the first

empirical evidence were provided to show that this set of 13 RM Y-STRs is able to achieve an order of magnitude higher male relative differentiation than is available with the commonly used Yfiler set, as well as to drastically improve male lineage differentiation over Yfiler.

However, to fully explore the potential for the RM Y-STR set in differentiating unrelated as well as related males for various purposes, much more data are needed. Therefore, the International RM Y-STR Study Group, a worldwide collaboration between 52 laboratories, was formed. Group members, chosen based on pre-existing practical experience in Y-STR analysis, genotyped under quality-controlled conditions the 13 RM Y-STRs in 14,644 males including 12,272 unrelated males from 111 worldwide populations and 4,744 closely related males. To compare the RM Y-STR set with conventional Y-STRs, Yfiler data were gathered in a subset of 7,784 unrelated men from 65 worldwide populations as well as in most of the male relatives analyzed for RM Y-STRs. The large collection of RM Y-STR haplotypes presented and explored in the present study enables a better understanding of their value to differentiate related and unrelated male individuals and populations, and additionally provides a suitable reference database for future use of RM Y-STRs in forensic, genealogical, anthropological, and population genetic studies.

Materials and Methods

DNA Samples

With a worldwide coverage in mind, contributing laboratories were identified and recruited based on their previous submission of Y-STR population data to the Y-chromosome Haplotype Reference Database (YHRD; www.yhrd.org), hence having proven experience in Y-STR analysis. Each of the 52 contributing laboratories within the International RM Y-STR Study Group genotyped a selection of their own in-house population sample sets, each consisting of seven to 634 individuals (median 100 individuals per population sample) across 111 defined population samples per laboratory. Population sets were to include only unrelated males. Three population samples were notable exceptions—the Biaka Pygmy group and the Bhutanese Lhokpu and Mönpa language groups. These populations, and particular sample sets used here, are known to contain high numbers of male relatives (Biaka), or were specifically selected as having extremely low resolution with Yfiler (Bhutan). Because these samples were ascertained differently to the other 108 populations within the study, they were excluded from the continental groups during analysis to avoid bias. The Aboriginal Australian samples used here were also not ascertained completely randomly, as they were selected based on Y-SNP haplogroups known to be authentic for Aboriginal Australians to avoid admixture effects as described elsewhere [Taylor et al. 2012].

A subset of 7,784 individuals from 65 populations were additionally genotyped with the AmpFISTR Yfiler PCR Amplification Kit (Life Technologies) for the most commonly used panel of 17 Y-STRs. Relative pairs were also analyzed, with 2,339 newly genotyped father–son pairs, 30 brother pairs, and three uncle–nephew pairs, for which relationship had previously been confirmed by autosomal DNA analysis; this dataset of male relatives was supplemented by 156 previously published relative pairs [Ballantyne et al., 2012].

Y-STR genotyping

The organizing laboratory (Department of Forensic Molecular Biology, Erasmus MC University Medical Centre Rotterdam)

provided genotyping protocols, allelic ladders, and tools for allele calling to all participating laboratories, and organized a quality control exercise prior to population data generation (for details see below). The 13 single- or multicopy RM Y-STR markers (DYF387S1, DYF399S1, DYF403S1, DYF404S1, DYS449, DYS518, DYS526, DYS547, DYS570, DYS576, DYS612, DYS626, and DYS627) were amplified in three multiplex PCRs, as described in Supp. Tables S1 and S2. PCR-amplified products were separated and detected using participating laboratories' standard protocols for analyzing STRs—in either ABI310, ABI3100, ABI3130, ABI3500, or ABI3730 Genetic Analyzers with POP-4, POP-6, or POP-7, and size standards ILS-600, ILS-500, or Genescan 500-LIZ. Allele calling was performed with GeneMapper (Life Technologies) or GeneMarker (SoftGenetics, State College, PA) utilizing custom panel and bin sets. The participating laboratories used their own primers and other consumables (except allelic ladders as provided by the host).

The commercially available AmpFISTR® Yfiler® PCR Amplification Kit (Life Technologies) targeting 17 conventional Y-STRs was used and allele calling was performed according to the manufacturer's instructions.

Genotyping Quality Control

To ensure genotyping consistency between the laboratories, all participants received allelic ladders prepared by the organizing laboratory, and six blind control DNA samples previously genotyped by the organizing laboratory. Genotyping of population samples was only allowed after a participating laboratory demonstrated the correct genotyping of these six blind control DNA samples at all 13 RM Y-STRs. If a participant reported erroneous genotypes, screenshots were requested and submitted to the organizing laboratory for an evaluation of the possible cause of error(s). Once the cause of error was identified, three additional blind control DNA samples were provided and genotyped by the participant. If these samples were typed correctly, the laboratory was allowed to type their population samples. In cases of unexpected results such as previously unknown alleles or microvariants, screenshots were requested and submitted to the organizing laboratory for inspection. Independent genotyping, and in some case DNA sequencing, was performed to resolve differences in genotyping of certain samples. Samples with missing data from more than one marker were excluded from data analysis to prevent low quality samples affecting genotype and haplotype distributions. Further, any differences observed between relative pairs were confirmed through duplicate, independent PCR amplifications, and genotyping, and in some cases, DNA sequencing.

Additional quality control was performed for the multicopy markers DYF403S1a+b and DYF399S1, whereby a subset of population sample electropherograms for these markers were sent to the organizing laboratory by all participants for blind confirmatory genotype scoring.

RM Y-STR Nomenclature

The nomenclature of the RM Y-STRs was updated to comply with the guidelines of the International Society of Forensic Genetics—ISFG [Gusmão et al., 2006], and to incorporate new variation observed in repeat structures. As such, all data collated were translated to comply with the updated nomenclature. Supp. Table S3 shows the repeat structure and allele designations used during the project, determined in collaboration with Life Technologies and L. Gusmão (IPATIMUP, Porto, Portugal). The RM Y-STR nomenclature used in the present study is in agreement with that used in the Powerplex® Y23 kit (Promega, Madison, WI) and in the Yfiler Plus Y-STR kit

(Life Technologies) for the RM Y-STR markers included, respectively.

Data Analysis

Haplotype diversities, average number of differing loci between pairs of individuals, pairwise F_{ST} genetic distances, and analysis of molecular variance (AMOVA) were calculated with Arlecore v3.5.1.3 [Excoffier and Lischer, 2010]. Theta (θ) values were calculated as per the reference Weir and Hill (2002). Molecular relationships between samples were analyzed using Network v4.6.1.1., applying the median-joining method. Weighting was applied as described elsewhere [Qamar et al., 2002], although based on updated mutation rates from Ballantyne et al. (2010) and www.yhrd.org. Only single-copy markers were utilized in network construction, due to the inability to assign multicopy genotypes to individual loci. Testing of statistical significance (t -tests) was performed in SPSS v17.0. Theoretical estimates for the rates of relative differentiation for each Y-STR panel were calculated using the formula $P(k < 0) = 1 - P(k = 0) = 1 - e^{-Nm}$ [Ballantyne et al., 2010], where N represents the number of markers and m represents the average mutation rate of the set of markers obtained from the sampling from the posterior distribution.

Results and Discussion

Global Diversity of RM Y-STR Haplotypes and Male Lineage Differentiation

The RM Y-STR set applied here generated exceptionally high haplotype diversity in the worldwide sample set analyzed with 12,156 unique (i.e., not matching any other individual in this dataset) haplotypes observed in 12,272 unrelated samples from 111 worldwide populations (Table 1). The global haplotype diversity estimate was 0.9999985 (sampling variance = 2.85×10^{-8}). Of the 102 haplotypes that were nonunique within this dataset (i.e., matched other individuals in this dataset), 90 were shared between two males, 10 between three males, and two between four males. All nonunique haplotypes were shared between males of the same geographic region, and all but six were from the same sampling population. The six nonunique haplotypes shared between different populations involved males from Argentina (Chubut and Rio Negro regions), Greenland (from two separate samples of Inuit), Czech Republic (from two sample sets), Cologne and Warsaw, and Bhutan (the Lhokpu and Mönpa tribes), respectively. As a result, the proportion of haplotypes shared between populations was exceptionally low at 0.05%, and the proportion of haplotypes shared within populations was slightly higher at 0.8%.

Within each regional geographic group (see Table 1), similar high levels of haplotype diversity were obtained ranging from 0.9999988 and 0.9999996 in Europeans and Migrant Europeans (i.e., European populations sampled in North America and Australia), respectively, down to 0.99836 in Central Asians (although the sample size of the latter group was much lower than that of all other regional groups studied) (Table 1). The average number of differing loci was highest with 17.5 loci in Admixed Native Americans and South Asians, respectively, as well as with 17.4 loci in Native Americans and Europeans, respectively, and was lowest with 15.26 loci in Middle Easterners (Table 1). As the maximal possible number is 21, these results illustrate how strikingly different the haplotypes were within each of the regional groups.

At the population level (Supp. Table S4), similar high levels of haplotype diversity were observed across all populations tested. Of

Table 1. RM Y-STR Haplotype Characteristics in a Global Set of 12,272 Individuals from 111 Populations Summarized for Regional Groups

Group	Individuals	Populations	Haplotypes	Haplotype diversity	θ	Average number of differing loci
Sub-Saharan Africa	303	5	300	0.99993	0.00006	16.5
North Africa	452	4	445	0.99992	0.00010	16.2
Middle East	100	1	100	1.0	0	15.26
Central Asia	86	1	80	0.99836	0.0012	16.18
South Asia	661	8	644	0.99992	0.00009	17.5
East Asia	967	7	964	0.999994	0.000006	17.3
South East Asia	634	6	630	0.99998	0.00002	16.6
Aboriginal Australian	100	1	96	0.99919	0.00069	16.8
Native American	365	10	357	0.99986	0.00015	17.4
Admixed Native American	764	12	758	0.99998	0.00002	17.5
European	5,618	38	5599	0.999988	0.000001	17.4
Migrant Sub-Saharan Africa	663	5	659	0.99998	0.000004	16.3
Migrant European	731	3	730	0.999996	0.000004	17.0
Migrant Asian	649	7	645	0.99998	0.00002	16.7
Bhutan ^a	78	2	56	0.99434	0.00794	13.9
Biaka Pygmy ^a	101	1	94	0.99822	0.00226	17.5
Global	12,272	111	12,156	0.9999985	0.00000238	16.5

^aBhutan and Biaka Pygmies were not assigned to a larger geographic group as the individuals sampled were ascertained differently (Bhutan), or are known to have different population characteristics to all other study populations (Biaka, see *Materials and Methods and DNA Samples*).

the 111 populations, 67 (60%) displayed haplotype diversities of 1.0, meaning that every individual tested per population (sample size ranging from seven to 467 in the various populations) had a different haplotype. The remaining populations had haplotype diversities ranging from 0.972 (Wichi, Salta Province, Argentina) to 0.999994 (Cologne, population 2).

Due to its highly multicopy nature, D4F403S1a+b caused concerns with some study group members regarding genotyping accuracy, so that additional quality control was performed for this marker (as well as for D4F399S1), as described in *Materials and Methods* section. There are four separate loci at the tetranucleotide D4F403S1a+b, three of which overlap in size. Differences in repeat sequences between the copies lead to the presence of partial (0.1, 0.2, and 0.3) alleles, which can challenge interpretation even in single source samples due to a lack of single base resolution with POP-4 during capillary electrophoresis. To determine the effect of removing this potentially problematic marker from the RM Y-STR panel, the entire dataset was additionally analyzed without D4F403S1a+b. The global haplotype diversity slightly decreased from 0.9999985 with the full 13-loci RM Y-STR set to 0.9999981 with the 12-loci set. This translated to an increase in the number of nonunique haplotypes from 116 to 152 in the 12,272 samples—a decrease of zero to five haplotypes per regional population, with an average of 0.32 fewer haplotypes within each of the 111 populations. Populations most affected by the omission of D4F403S1a+b were Aboriginal Australians (from 96 to 92 haplotypes among the 100 men tested), and Bhutan (from 56 to 51 haplotypes among the 78 men tested). However, the vast majority of the populations tested (90 of 111, 81.1%) were not affected by the removal of D4F403S1a+b, with the haplotype diversities remaining the same. Therefore, and because of the extra care in quality control employed for this marker, all following data analyses were performed based on the complete 13-loci RM Y-STR set.

Genetic–Geographic Population Substructure with RM Y-STRs

The θ values for all regional groups and for almost all populations were exceptionally low (Table 1; Supp. Table S4), indicating relatively little population substructure detected with the RM

Y-STR set. Indeed, AMOVA based on F_{ST} values (the inability to accurately assign the alleles of the multicopy markers to specific loci prevented R_{ST} values being calculated) demonstrated that 99.98% of haplotype variation was within populations, extremely high for Y-chromosome markers [Willuweit and Roewer, 2007], whereas 0.02% was among populations within the same regional group, and only 0.002% among regional groups. The global F_{ST} was as low as 0.00017. Between the regional groups, the average pairwise F_{ST} value was only 0.000127, with a maximum value of 0.00058 observed between Aboriginal Australians and Middle Eastern populations. Even when considering individual populations, the highest pairwise F_{ST} value observed was only 0.02815 between the Bhutan Lhokpu and Argentinian Wichi from Salta Province (data not shown). Across all population comparisons, an average pairwise F_{ST} value of 0.000826 was obtained.

To illustrate the magnitude of haplotypic differences between populations and geographic areas, multidimensional scaling (MDS) analysis was performed on Slatkin's linearized F_{ST} values (Fig. 1). The majority of the populations formed a loose cluster, with moderate dispersion across both dimensions. While there were several outlier populations (for example Pakistani Punjabis, Japanese Gunma, and Angola Kimbundos), overall there was no geographic pattern emerging, in line with the AMOVA results. It is notable that populations from the same geographic group did not necessarily cluster together (Fig. 1). The differences in population sample size (Supp. Table S4) did not greatly affect the dispersion of populations—similar distributions were observed when the population sizes were restricted to 100/population (inset, Fig. 1), or 25/population (data not shown). This analysis, together with the AMOVA results, illustrates that no genetic–geographic structuring is detected with this set of RM Y-STR haplotypes, since their high mutation rates have likely removed signals of shared population history, and has driven the high number of unique haplotypes. The low between-population differentiation and lack of substructure could be expected as a result of the high within-population diversities, and the effect of mutation rates on the estimation of θ [Meirmans and Hedrick, 2011]. As a practical consequence, no population substructure correction needs to be applied when using this RM Y-STR set in forensic or other applications, as is usually needed for other DNA marker systems. However, the lack of substructure correction required does not necessarily infer a lack of structure in the distribution of

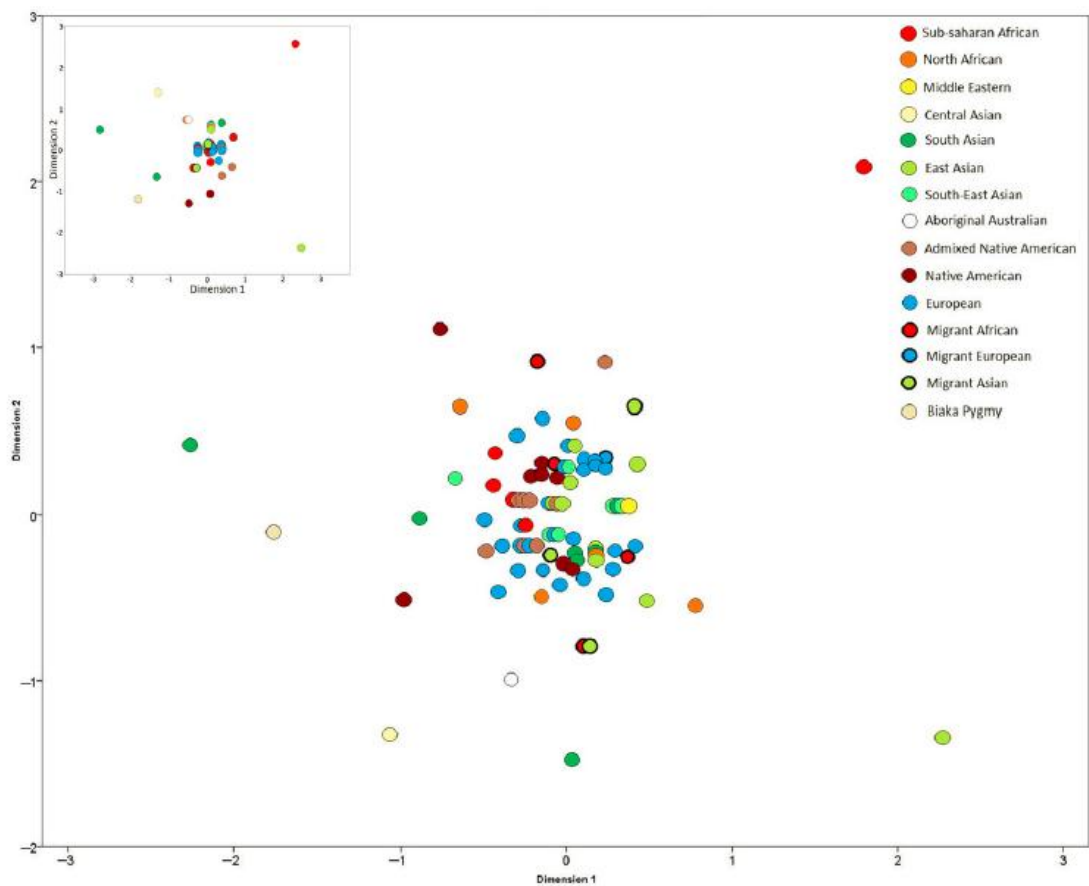


Figure 1. Two-dimensional plot of MDS analysis of Slatkin's linearized F_{ST} values for RM Y-STR haplotypes in a global sample of 12,272 individuals from 111 populations (stress 0.07462). Smaller inset MDS shows the effect of equalized sample size ($N = 100$ individuals or fewer per population, stress = 0.02416).

haplotypes across geographic regions. Nonrandom distributions, caused by shared population histories and common origins, would infer the need to generate and utilize regional or meta-population-specific databases for frequency estimation. As such, the ability for RM Y-STRs to differentiate between male lineage within and between different populations and the relative distribution of haplotypes between regional groups was investigated.

Value of RM Y-STRs in Improving Male Lineage Differentiation

When considering the utility of the RM Y-STR set for future applications in forensic, genealogical or anthropological genetic studies, it is informative to compare its properties to those of conventional Y-STRs, such as the 17 loci included in the Yfiler kit. Hence, RM YSTR and Yfiler haplotypes were compiled for 7,784 individuals across 65 populations as a subset of the global dataset presented here. In every diversity measure applied, the 13 RM YSTR set provided enhanced estimates relative to the 17-loci Yfiler set. Global haplotype diversity

was increased from 0.99995 with Yfiler to 0.999997 with RM Y-STRs; the total number of haplotypes detected was increased from 6,975 to 7,714, and the number of singletons from 6,469 to 7,647. All regional groups showed more haplotypes and higher diversity estimates with RM Y-STRs relative to Yfiler, that is, haplotype diversity increased on average by 0.00226 and the number of singletons on average by 77 (largest increase of 1.56-fold in Aboriginal Australians) (Table 2). At the population level, there were increases in the number of haplotypes and in the haplotype diversity estimates for 56 (86.2%) of the 65 populations with the RM Y-STR set compared with Yfiler, whereas for six (9.2%) populations the same number of haplotypes were obtained because all males were already fully individualized with Yfiler (often in small sample-sized populations), and for three populations (4.6%), one haplotype fewer was detected with the RM Y-STRs than with Yfiler (Supp. Table S5).

There was a significant difference between the RM Y-STR and the Yfiler sets in the level of haplotype sharing between individuals within regional groups, between populations, and within populations (Fig. 2). At the global level, 506 Yfiler haplotypes were shared between 1,318 individuals, compared with only 66 RM

Table 2. Comparison of RM Y-STR and Yfiler Haplotype Characteristics in a Global Set of 7,784 Individuals from 65 Populations Summarized for Regional Groups

Group	Individuals	Populations	Yfiler haplotypes	RM Y-STR haplotypes	Yfiler haplotype diversity	RM Y-STR haplotype diversity	Yfiler average percentage of differing loci	RM Y-STR average percentage of differing loci
North Africa	193	2	173	189	0.99792	0.99973	59%	73%
Central Asia	83	1	67	77	0.99060	0.99824	63%	77%
South Asia	497	6	450	483	0.99951	0.99988	68%	83%
East Asia	633	5	580	630	0.99952	0.99999	64%	82%
South East Asia	200	2	175	198	0.99759	0.99990	66%	81%
Aboriginal Australian	100	1	74	96	0.99152	0.99919	68%	80%
Native American	279	8	233	275	0.99761	0.99988	61%	82%
Admixed Native American	458	6	444	454	0.99986	0.99996	67%	81%
European	4,041	25	3,696	4,025	0.99991	0.999998	66%	82%
Migrant Sub-Saharan Africa	442	3	407	439	0.99909	0.99997	63%	81%
Migrant European	552	3	541	551	0.99993	0.99999	63%	81%
Migrant Asian	205	2	176	203	0.99737	0.99990	66%	83%
Biaka Pygmies	101	1	83	94	0.99505	0.99822	65%	83%
Global	7,784	65	6,975	7,714	0.99995	0.999997	65%	81%

Y-STR haplotypes shared between 70 individuals. For Yfiler at the population level, only three (4.6%) of the 65 populations did not display any matching individuals either within or outside populations; 58 (89.2%) and 52 (80.0%), respectively, showed within and outside population haplotype matches. Within- and outside-population haplotype matches for Yfiler were as high as 62%, for the Wichis—a Native American group from Argentina (although sample size was $N = 13$). Notably, the level of outside-population matches compared with within-population matches was markedly higher for most populations for Yfiler haplotypes, whereas strikingly reduced for RM Y-STRs (Fig. 2). Of the 65 populations, 35 (53.8%) did not show any haplotype matches within and outside populations. Only 29 (44.6%) of the 65 populations displayed haplotype matches within populations, with the highest proportion (11%) observed in Argentinian Wichi from the Salta province. For 49 of the 58 populations that showed within-population Yfiler haplotype matches, no RM Y-STR haplotype matches were observed. The reduction in haplotype matches for RM Y-STRs relative to Yfiler is even more striking when considering outside-population matches. Only four (6.2%) populations displayed outside-population RM Y-STR matches, namely, two Czech population samples (one haplotype), and Cologne and Warsaw (one haplotype). The finding that 61 populations (93.8%) showed no outside population haplotype matches for RM Y-STRs contrasts strongly with those found with Yfiler, for which the corresponding number is 13 (20%).

Figure 3 provides an overview of the population pairs that showed between population haplotype matches for Yfiler (Fig. 3A, blue lines) and RM Y-STRs (Fig. 3B, blue lines). The significant reduction in between-population haplotype sharing ($t_{64} = 8.091$, $P = 2.23 \times 10^{-11}$) demonstrates the power of the RM Y-STR panel in male lineage differentiation. For RM Y-STRs, no haplotype sharing was observed between different populations sampled from the same countries, such as the five Indian, two Japanese, three Korean, three Austrian, two Belgian, three German, two Italian, two Polish, three US American, and two Hungarian populations, except for the two Czech populations. Further, almost no RM Y-STR haplotype sharing was observed between populations from the same geographic region such as the densely sampled continent of Europe except for Cologne and Warsaw. In contrast, many Yfiler haplotypes were shared between populations within countries, and between populations within regions such as Europe (see Fig. 3, insets).

The lack of RM Y-STR haplotype sharing between populations is not limited to the full 13-loci haplotype. Network analysis of 1,000 individuals, selected at random in the same population proportions as the full set of 7,784 samples and based on single-copy Y-STRs, displayed little geographic clustering with RM Y-STR haplotypes (Fig. 4B, perhaps with the exception of North Africans), whereas with Yfiler haplotypes for the same individuals (Fig. 4A), geographic clustering is seen with several groups. For RM Y-STRs, the network was constructed based on nine single-copy Y-STRs (excluding DYS387S1, DYS399S1, DYS403S1, and DYS404S1), and for Yfiler, based on 15 Y-STRs (excluding DYS385a/b). Unlike the Yfiler network, the RM Y-STR network provides almost no clustering of haplotypes according to geographic regions of sampling (except some of the North African haplotypes), which underlines the lack of population substructure as also seen in the MDS and AMOVA of the complete data. Hence, for RM Y-STRs, the need for regional (metapopulation) reference databases for haplotype frequency estimation in forensic and other applications is strongly reduced compared with conventional Y-STRs such as those in the Yfiler kit. Combining the RM Y-STR set and the Yfiler set to generate 30-marker Y-STR haplotypes resulted in the individualization of 25 additional men of the 137 not already individualized with the RM Y-STR set alone. While this does represent a slight improvement on the global scale, from 7,714 to 7,737 haplotypes, it is clear that the vast majority of differentiation was achieved by the RM Y-STR set alone.

Ability of RM Y-STRs to Detect Hidden Paternal Relationships

The observation of RM Y-STR haplotype sharing within, but (almost) not between, populations might indicate that RM Y-STRs allow the detection of unknown (i.e., hidden) paternal relationships in populations. If true, more shared RM Y-STR haplotypes would be expected between putatively unrelated individuals sampled from rural areas, where male relatives often stay in the region (and thus can be sampled), and especially from endogamous populations, than among putatively unrelated individuals sampled from urban areas, where male relatives tend to not stay (and thus cannot be sampled). To test this hypothesis, all populations were designated

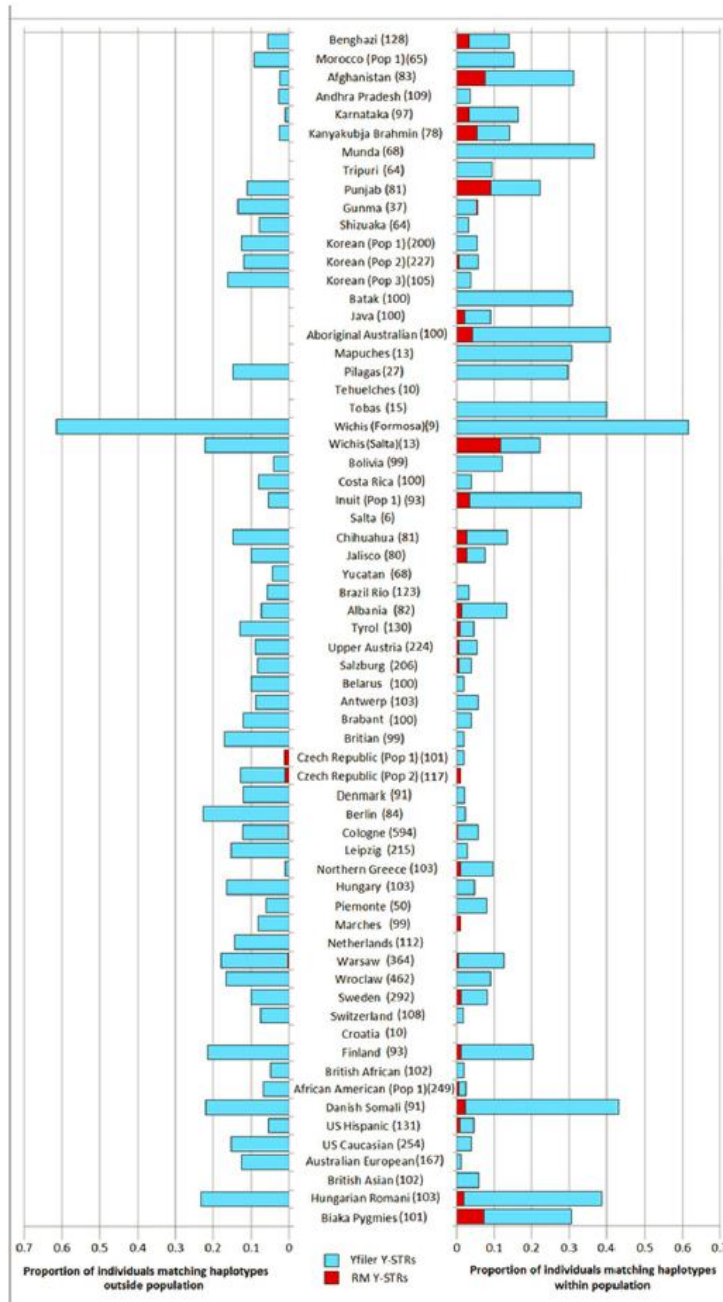


Figure 2. Proportion of individuals with haplotypes shared within populations (right) and between populations (left) for Yfiler (light blue bars) and the RM Y-STR set (dark red bars).

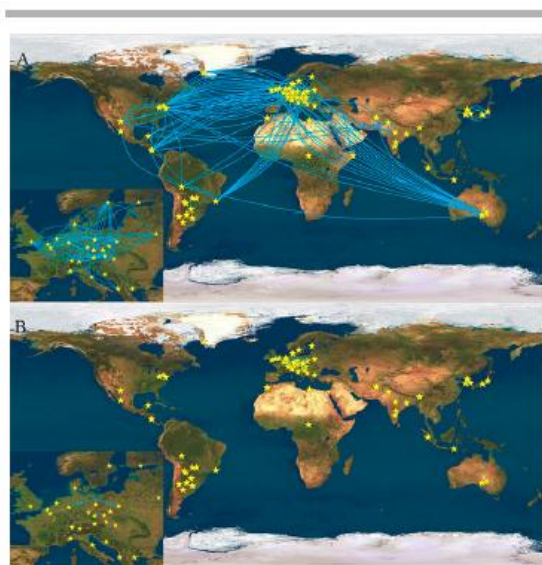


Figure 3. Geographic representation of pairwise between-population haplotype sharing. Blue lines connect population pairs showing shared haplotypes for **A:** Yfiler, and **B:** the RM Y-STRs set. Smaller insets show Europe enlarged.

according to their urbanization status—each set of males was classified as either urban (genetically moderately outbred), rural (more likely to be genetically inbred), mixed (males from both rural and urban populations, in unspecified proportions), or endogamous (populations subscribing to the cultural practice of endogamy—restricting marriage outside of an individual's social class, ethnic group, or tribe). In support of the RM Y-STR set's ability to detect hidden paternal relationships, we indeed see significantly lower hap-

lotype diversity and lower proportions of unique haplotypes in the endogamous group than in the rural group than in the urban group (diversity $t_2 = 7,960$, $P = 1.6 \times 10^{-8}$, shared haplotypes $t_2 = 1.347$, $P = 0.310$) (Table 3). Although the sample sizes differ considerably between the three categories, this does not affect the conclusions as simulations performed with equalized sample numbers ($n = 558$ per category) showed that the RM Y-STRs still result in decreased diversity in endogamous and rural populations compared with urban populations (data not shown; diversity $t_2 = 8936$, $P = 1.3 \times 10^{-8}$, shared haplotypes $t = 1.117$, $P = 0.380$).

On the other hand, these data also demonstrate that even in populations with considerable hidden paternal relationships, such as in endogamous populations, the RM Y-STR set still allows the differentiation of a large number of men. A comparison between Yfiler and the RM Y-STR set for the available populations (Table 4) illustrates the improved value of RM Y-STRs relative to Yfiler in differentiating males when grouping populations into urban, rural, and endogamous. Most tellingly, even in the endogamous group, the haplotype diversity increased and the proportion of shared (i.e., nonunique) haplotypes decreased from 0.99947 and 0.101 with Yfiler to 0.99978 and 0.045 with the RM Y-STR set, respectively, in the rural group from 0.99916 and 0.157 to 0.99997 and 0.008, and even in the urban from 0.99994 and 0.08 to 0.999998 and 0.005.

Value of RM Y-STRs for Male Relative Differentiation

In addition, we tested the value of RM Y-STRs for differentiating male relatives. On a theoretical level, the rate of relative differentiation per meiosis can be calculated using the mutation rates of each locus within a genotyping panel (see *Materials and Methods* section). For the Yfiler set, it is estimated that a mutation at one or more of the 17 Y-STR loci will occur with a probability of 0.047 (95% confidence interval 0.038–0.057) per meiosis, which for the RM Y-STR set is more than fourfold higher at 0.195 (95% CI 0.177–0.21). Notably, for the recently released PowerPlex Y-23 kit (Promega), which targets all 17 Yfiler markers together with two of the 13 RM Y-STR

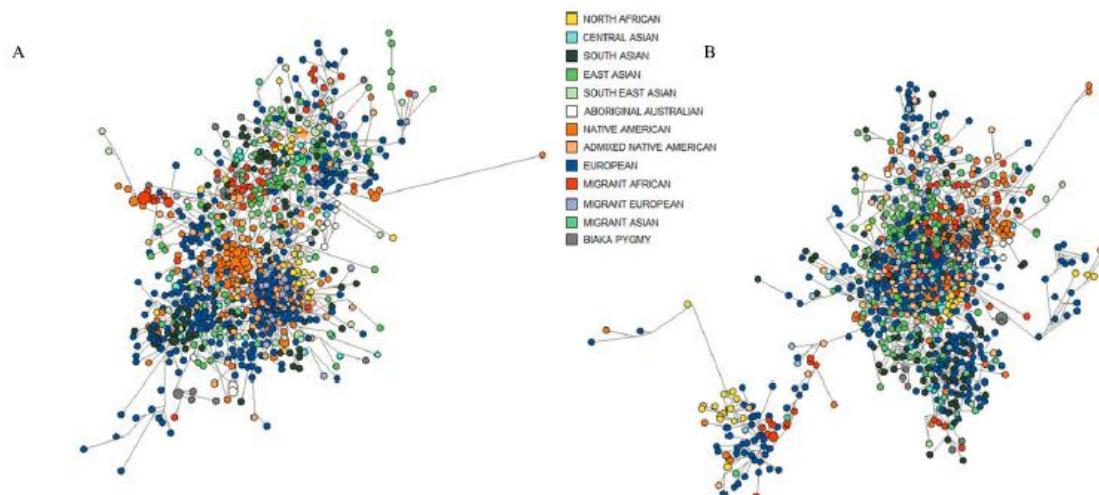


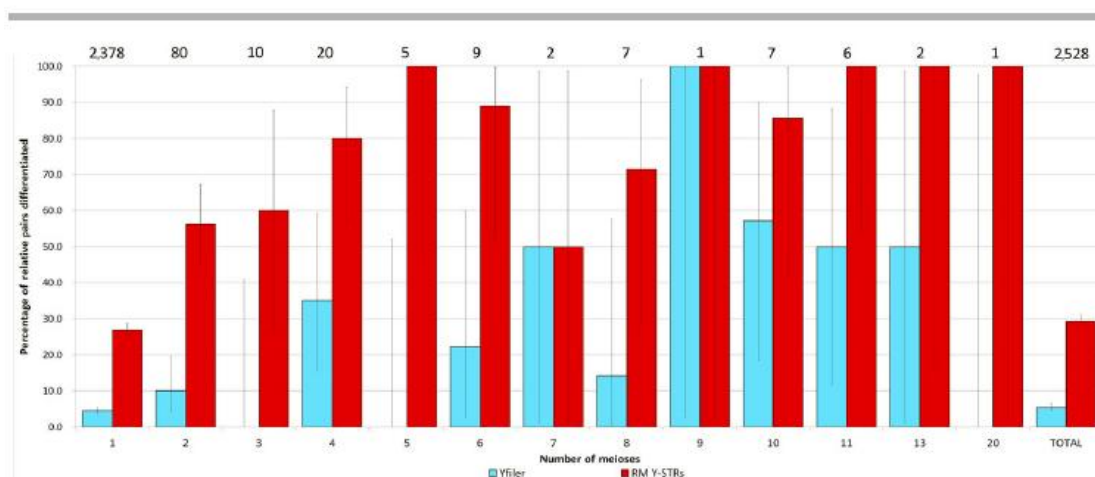
Figure 4. Weighted median-joining networks from single-copy Y-STRs for 1,000 individuals randomly selected from the total dataset, with regional geographic groups indicated by colors (see inset legend). **A:** Based on 15-loci Yfiler haplotypes (excluding DYS385a/b). **B:** Based on 10-loci RM Y-STR haplotypes (excluding DYF399S1, DYF403S1, and DYF404S1).

Table 3. RM Y-STR Haplotype Characteristics in a Global Set of 12,272 Individuals from 111 Populations Summarized for Groups Based on Urbanization Status

Population classification	Populations	Individuals	Haplotypes	Singletons	Haplotype diversity	Proportion of nonunique haplotypes
Endogamous	7	558	512	478	0.999614	0.143
Rural	20	1,369	1,354	1,340	0.999982	0.021
Urban	51	7,198	7,164	7,131	0.999999	0.009
Mixed	26	2,379	2,361	2,342	0.999994	0.015
Unknown	7	770	767	764	0.999990	0.008

Table 4. Comparison of RM Y-STRs and Yfiler Haplotype Characteristics in a Global Set of 7,784 Individuals from 65 Populations Summarized for Groups Based on Urbanization Status

Population classification	N	Yfiler ht	RM Y-STR ht	Yfiler singletons	RM Y-STR singletons	Yfiler haplotype diversity	RM Y-STR haplotype diversity	Yfiler proportion of nonunique haplotypes	RM Y-STR proportion of nonunique haplotypes	Yfiler percentage average allelic difference	RM Y-STR percentage average allelic difference
Endogamous	466	419	445	380	427	0.99947	0.99978	0.101	0.045	68%	83%
Rural	714	602	708	537	703	0.99916	0.99997	0.157	0.008	70%	85%
Urban	4,974	4,577	4,951	4,318	4,929	0.99994	0.99998	0.080	0.005	68%	83%
Mixed	1,422	1,332	1,407	1,267	1,392	0.99988	0.99999	0.063	0.011	67%	84%
Unknown	208	181	205	169	202	0.99684	0.99986	0.130	0.014	64%	83%

**Figure 5.** Empirical male relative differentiation using 2,528 paternal relative pairs separated by one to 20 meioses for Yfiler (light blue bars) and the RM Y-STR set (dark red bars). The data combine the new 2,372 relative pairs from the current study with the previously used 156 relative pairs [Ballantyne et al., 2012]. Values above the bars indicate the absolute number of relative pairs the estimated differentiation rate is based upon. Error bars represent binomial confidence intervals.

markers and four additional Y-STRs, this value is 0.092 (95% CI 0.077–0.107), nearly twofold higher than for Yfiler, but more than twofold lower than for the RM Y-STR set studied here.

To compare the theoretical expectation for the RM Y-STR and the Yfiler sets with empirical data, we genotyped 2,372 pairs of male relatives previously confirmed by autosomal DNA analysis (2,339 father–son pairs, 30 brother pairs, and three uncle–nephew pairs), confirmed the observed allelic differences by independent genotyping (and some by sequencing), and combined these new data with those from the 156 relative pairs separated by one to 20 paternal genetic transfers (or meioses) described previously [Ballantyne et al., 2012]. In this combined dataset, the RM Y-STR set allows differentiation by at least one allelic difference (i.e., mutation) in at least one locus in 742 (29%) of the total 2,528 pairs related by one to

20 generations, whereas Yfiler only allowed the separation of 118 (5.5%) of a subset of 2,161 of these relative pairs (Fig. 5). In particular, the RM Y-STR set differentiated fathers from their sons in 26.9% of the cases versus 4.5% with Yfiler, and brothers from each other in 56.3% of the cases versus 10.0% with Yfiler.

It should be noted, however, that only for the 2,378 father–son pairs investigated here, was the sample size large enough to allow reliable conclusions, whereas for more distantly related males, future studies need to deliver more data. However, as the chance of mutations increases with each meiosis, it is expected that the more distantly related men are, the higher the chance that they will have different RM Y-STR haplotype; a trend that can be seen in our data (Fig. 5). The importance of sufficient sample size is illustrated in the following for father–son pairs. Our initial estimate of

father–son differentiation was 70% based on only 20 pairs [Ballantyne et al., 2010], which was subsequently revised to 49% based on 39 pairs, including the initial 20 pairs [Ballantyne et al., 2012]. In the present study, we managed to increase the number of father–son pairs drastically to 2,378 pairs, now achieving a differentiation rate of 26.9% (95% confidence interval 25.1%–28.7%). Notably, the empirical rate achieved, now based on a large sample size, corresponds more closely to the theoretical estimate of 19.5%.

Removing DYF403S1a+b from the RM Y-STR panel resulted in the nondifferentiation of 84 relative pairs that could be differentiated when DYF403S1a+b was included; the overall relative differentiation rate dropped to 26%. However, the difference only occurred in pairs with fewer than five meioses separating the individuals—in all other pairs, there were sufficient mutations at other RM Y-STR loci to allow differentiation (data not shown).

As pointed out previously [Ballantyne et al. 2012], RM Y-STRs are not particularly useful for paternity testing or familial searching because of their high mutation rates and the resultant power of differentiation of male relatives. If, however, they are used for such purposes, correction factors would be needed to compensate for the likely observed mutations in the probability calculations. It has been suggested before that the presence of at least three mutations at any Y-STR loci (most commonly with the 17 Yfiler Y-STRs) is sufficient to exclude paternity [Kayser and Sajantila 2001; Gjertson et al., 2007]. In the current dataset however, 3.8% of the DNA-confirmed father–son pairs displayed three or four RM Y-STR mutations; as such, the number of mutations constituting exclusion should be increased when using the RM Y-STRs. As outlined elsewhere [Ballantyne et al. 2012], instead of an ad hoc cut-off, a dynamic threshold shall be used, estimated from the number of Y-STR loci genotyped and their locus-specific mutation rates. Because three mutations have also been observed occasionally among 17 Yfiler loci in DNA-confirmed father–son pairs [Goedbloed et al. 2009], this notion also applies to conventional Y-STR sets such as Yfiler, which due to their moderate mutation rates in principle are more appropriate for paternity and kinship testing than RM Y-STRs.

Conclusions

The large worldwide dataset compiled, presented, and analyzed here demonstrates the exceptional value of the RM Y-STR panel for differentiating male lineages on a global and regional scale. Many of the current limitations of Y-chromosome applications are reduced with this RM Y-STR set, providing increased utility and effectiveness to the genetic analysis of male populations and lineages. In particular, this 13 RM Y-STR set provides near-complete paternal lineage differentiation in general populations as well as in populations with otherwise reduced Y-chromosome diversity, due to peculiarities in population history or cultural practices. The effects of this near-complete male individualization will be of great benefit to numerous fields using the Y-chromosome genetics to investigate male lineages, such as in genealogical studies (e.g., to detect extrapair paternity or adventitious haplotype matches within different surnames/lineages), population genetic and genetic history studies (e.g., to assist in differentiating between lineages shared between populations with common history), and in forensic applications (e.g., to reduce the inclusion of innocent individuals in investigations due to adventitious haplotype matches). Moreover, the immense value of this RM Y-STR set to differentiate between both close and distant male relatives will have beneficial effects on these same fields, especially in forensic genetics, providing increased confidence that haplotype matches between unknown samples such

as those from crime scenes and reference samples such as those from suspects are due to individual identity, rather than relatedness.

On the other hand, however, the extreme degree of RM Y-STR haplotype diversity highlights the limitations of the current approach for placing a statistical weight on Y-STR haplotype matches by using empirically derived haplotype frequency estimates obtained from Y-STR haplotype reference databases. Although already noted with the currently used Y-STR sets such as Yfiler, the problem that an observed haplotype is not present in the large and growing reference database, and how to deal with rare haplotypes in obtaining frequency estimates and estimating match probabilities for which no consensus has been reached yet [Krawczak, 2001; Brenner, 2010; Buckleton et al., 2011; Willuweit et al., 2011], will become drastically increased when using RM Y-STRs, as can be seen from the data presented here. Therefore, new statistical solutions shall be developed to estimate the weight of a RM Y-STR haplotype match with particular relevance for forensic applications.

To assist future studies utilizing RM Y-STRs, complete dataset of RM Y-STR and Yfiler haplotypes obtained from the 12,272 individuals analyzed here is made available via Supp. Table S6, with allele frequencies for individual populations, regional groups, and the complete dataset summarized in Supp. Table S7. Relative pair haplotypes for RM Y-STRs and Yfiler are available in Supp. Table S8.

Acknowledgments

The authors thank the following colleagues for their expert technical assistance: Linda S. H. Gan (Singapore) for the Singapore samples, Minttu Hedman (Helsinki) for the Finnish samples, Helena Nilsson (Linköping) for the Swedish samples, Kevin Cloete (Cape Town) for the South African samples, Nancy Vanderheyden (Leuven) for the Belgium samples, Baowen Cheng (Kunming) and Yanan Sun (Kunming) for the Kunming Chinese samples, Juliana Januzzi (Rio de Janeiro) for the Rio Brazil samples, Maria Lourdes D. Honrado (Quezon City) for the Filipino samples, and Harald Niederstätter (Innsbruck) for the Innsbruck Austrian samples as well as for his comments on the manuscript. Oscar Lao (Rotterdam) is acknowledged for statistical advice and help in some calculations. Certain commercial equipment, instruments, and materials are identified to specify experimental procedures as completely as possible. In no case does such identification imply a recommendation or endorsement by the National Institute of Standards and Technology (USA), nor does it imply that any of the materials, instruments, or equipment identified are necessarily the best available for the purpose.

Disclosure statement: KNB and MKay, together with two colleagues from Life Technologies, are coinventors on a registered patent about the RM Y-STRs used here and their forensic utility; however, no personal or institutional benefits were received. All other authors declare no conflict of interest.

References

- Ballantyne KN, Goedbloed M, Fang R, Schaap O, Lao O, Wollstein A, Choi Y, van Duijn K, Vermeulen M, Brauer S, Decorte R, Poetsch M, et al. 2010. Mutability of Y-chromosomal microsatellites: rates, characteristics, molecular bases, and forensic implications. *Am J Hum Genet* 87:341–353.
- Ballantyne KN, Keerl V, Wollstein A, Choi Y, Zuniga SB, Ralf A, Vermeulen M, de Knijff P, Kayser M. 2012. A new future of forensic Y-chromosome analysis: rapidly mutating Y-STRs for differentiating male relatives and paternal lineages. *Forensic Sci Int Genet* 6:208–218.
- Brenner CH. 2010. Fundamental problem of forensic mathematics—the evidential value of a rare haplotype. *Forensic Sci Int Genet* 4:281–291.
- Buckleton JS, Krawczak M, Weir BS. 2011. The interpretation of lineage markers in forensic DNA testing. *Forensic Sci Int Genet* 5:78–83.
- Coble MD, Loreille OM, Wadhams MJ, Edson SM, Maynard K, Meyer CE, Niederstätter H, Berger C, Berger B, Falsetti AB, Gill P, Parson W, et al. 2009. Mystery solved: the identification of the two missing Romanov children using DNA analysis. *PLoS One* 4:e4838.

- Excoffier L, Lischer HEL. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* 10:564–567.
- Gjertson DW, Brenner CH, Baur MP, Carracedo A, Guidet F, Luque JA, Lessig R, Mayr WR, Pascali VR, Prinz M, Schneider PM, Morling N. 2007. ISFG: recommendations on biostatistics in paternity testing. *Forensic Sci Int Genet* 1:223–231.
- Goedbloed M, Vermeulen M, Fang RN, Lembring M, Wollstein A, Ballantyne K, Lao O, Brauer S, Kruger C, Roewer L, Lessig R, Ploski R, et al. 2009. Comprehensive mutation analysis of 17 Y-chromosomal short tandem repeat polymorphisms included in the AmpFISTR Yfiler PCR amplification kit. *Int J Legal Med* 123:471–482.
- Gusmão L, Butler JM, Carracedo A, Gill P, Kayser M, Mayr WR, Morling N, Prinz M, Roewer L, Tyler-Smith C, Schneider PM. 2006. DNA Commission of the International Society of Forensic Genetics (ISFG): an update of the recommendations on the use of Y-STRs in forensic analysis. *Int J Legal Med* 120:191–200.
- Hammer MF, Karafet TM, Redd AJ, Jarjanazi H, Santachiara-Benerecetti S, Soodyall H, Zegura SL. 2001. Hierarchical patterns of global human Y-chromosome diversity. *Mol Biol Evol* 18:1189–1203.
- Jobling MA, Tyler-Smith C. 2003. The human Y chromosome: an evolutionary marker comes of age. *Nat Rev Genet* 4:598–612.
- Kayser M, Caglia A, Corach D, Fretwell N, Gehrig C, Graziosi G, Heidorn F, Herrmann S, Herzog B, Hidding M, Honda K, Jobling M, et al. 1997. Evaluation of Y-chromosomal STRs: a multicenter study. *Int J Legal Med* 110:125–133, 141–129.
- Kayser M, Sajantila A. 2001. Mutations at Y-STR loci: implications for paternity testing and forensic analysis. *Forensic Sci Int* 118:116–121.
- Kayser M, Brauer S, Schädlich H, Prinz M, Batzer MA, Zimmerman PA, Boatín BA, Stoneking M. 2003. Y chromosome STR haplotypes and the genetic structure of U.S. populations of African, European, and Hispanic ancestry. *Genome Res* 13:624–634.
- Kayser M, Lao O, Anslinger K, Augustin C, Bargel G, Edelmann J, Elias S, Heinrich M, Henke J, Henke L, Hohoff C, Illing A, et al. 2005. Significant genetic differentiation between Poland and Germany follows present-day political borders, as revealed by Y chromosome analysis. *Human Genet* 117:428–443.
- King TE, Jobling M. 2009. What's in a name? Y chromosomes, surnames and the genetic genealogy revolution. *Trends Genet* 25:351–360.
- Krawczak M. 2001. Forensic evaluation of Y-STR haplotype matches: a comment. *Forensic Sci Int* 118:114–115.
- Meirmans PG, Hedrick PW. 2011. Assessing population structure: FST and related measures. *Mol Ecol Resour* 11:5–18.
- Oota H, Settheetham-Ishida W, Tiwawech D, Ishida T, Stoneking M. 2001. Human mtDNA and Y-chromosome variation is correlated with matrilineal versus patrilineal residence. *Nat Genet* 29:20–21.
- Parkin EJ, Kraayenbrink T, Opgenort JRML, van Driem GL, Tuladhar NM, de Knijff P, Jobling MA. 2007. Diversity of 26-locus Y-STR haplotypes in a Nepalese population sample: isolation and drift in the Himalayas. *Forensic Sci Int* 166:176–181.
- Poznik GD, Henn BM, Yee MC, Sliwerska E, Euskirchen GM, Lin AA, Snyder M, Quintana-Murci L, Kidd JM, Underhill PA, Bustamante CD. 2013. Sequencing Y chromosomes resolves discrepancy in time to common ancestor of males versus females. *Science* 341:562–565.
- Qamar R, Ayub Q, Mohyuddin A, Helgason A, Mazhar K, Mansoor A, Zerjal T, Tyler-Smith C, Mehdi SQ. 2002. Y-chromosomal DNA variation in Pakistan. *Am J Hum Genet* 70:1107–1124.
- Rebala K, Mikulich AI, Tsybovsky IS, Siváková D, Džupinková Z, Szczerkowska-Dobosz A, Szczerkowska Z. 2007. Y-STR variation among Slavs: evidence for the Slavic homeland in the middle Dnieper basin. *J Human Genet* 52:406–414.
- Roewer L, Croucher PJP, Willuweit S, Lu TT, Kayser M, Lessig R, De Knijff P, Jobling M, Tyler-Smith C, Krawczak M. 2005. Signature of recent historical events in the European Y-chromosomal STR haplotype distribution. *Human Genet* 116:279–291.
- Roewer L. 2009. Y chromosome STR typing in crime casework. *Forensic Sci Med Pathol* 5:77–84.
- Shi W, Ayub Q, Vermeulen M, Shao RG, Zaniga SB, Van der Gaag K, De Knijff P, Kayser M, Xue Y, Tyler-Smith C. 2010. A worldwide survey of human male demographic history based on Y-SNP and Y-STR data from the HGDP-CEPH populations. *Mol Biol Evol* 27:385–393.
- Taylor D, Nagle N, Ballantyne KN, van Oorschot RA, Wilcox S, Henry J, Turakulov R, Mitchell RJ. 2012. An investigation of admixture in an Australian Aboriginal Y-chromosome STR database. *Forensic Sci Int Genet* 6:532–538.
- Underhill PA, Shen P, Lin AA, Jin L, Passarino G, Yang WH, Kauffman E, Bonne-Tamir B, Bertranpetit J, Francalacci P, Ibrahim M, Jenkins T, et al. 2000. Y chromosome sequence variation and the history of human populations. *Nat Genet* 26:358–361.
- van Oven M, Hammerle JM, van Schoor M, Kushnick G, Pennekamp P, Zega I, Lao O, Brown L, Kennerknecht I, Kayser M. 2011. Unexpected island effects at an extreme: reduced Y chromosome and mitochondrial DNA diversity in Nias. *Mol Biol Evol* 28:1349–1361.
- van Oven M, van Geystelen A, Kayser M, Decorte R, Larmuseau MHD. 2014. Seeing the wood for the trees: a minimal reference phylogeny for the human Y chromosome. *Hum Mutat* 35:187–191.
- Weir BS, Hill WG. 2002. Estimating F-statistics. *Annu Rev Genet* 36:721–750.
- Willuweit S, Caliebe A, Andersen MM, Roewer L. 2011. Y-STR frequency surveying method: a critical reappraisal. *Forensic Sci Int Genet* 5:84–90.
- Willuweit S, Roewer L. 2007. Y chromosome haplotype reference database (YHRD): update. *Forensic Sci Int Genet* 1:83–87.
- Xue Y, Wang Q, Long Q, Ng BL, Swerdlow H, Burton J, Skuce C, Taylor R, Abdellah Z, Zhao Y, Asan, MacArthur DG, et al. 2009. Human Y chromosome base-substitution mutation rate measured by direct sequencing in a deep-rooting pedigree. *Curr Biol* 19:1453–1457.

ANEXO B – Male lineage strata of Brazilian population disclosed by the simultaneous analysis of STRs and SNPs

Forensic Science International: Genetics 13 (2014) 264–268



Contents lists available at ScienceDirect

Forensic Science International: Genetics

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



Male lineage strata of Brazilian population disclosed by the simultaneous analysis of STRs and SNPs



Andréa M. Oliveira^a, Patricia M. Domingues^a, Verónica Gomes^b, António Amorim^{b,c}, Juliana Jannuzzi^a, Elizeu F. de Carvalho^a, Leonor Gusmão^{a,b,*}

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

^b IPATIMUP – Institute of Molecular Pathology and Immunology of the University of Porto, Porto, Portugal

^c FCUP – Faculty of Sciences of the University of Porto, Porto, Portugal

ARTICLE INFO

Article history:

Received 22 July 2014

Received in revised form 28 August 2014

Accepted 29 August 2014

Keywords:

Y-STRs

Y-SNPs

Brazil

Population stratification

ABSTRACT

Brazil has a large territory divided in five geographical regions harboring highly diverse populations that resulted from different degrees and modes of admixture between Native Americans, Europeans and Africans. In this study, a sample of 605 unrelated males was genotyped for 17 Y-STRs and 46 Y-SNPs aiming a deep characterization of the male gene pool of Rio de Janeiro and its comparison with other Brazilian populations. High values of Y-STR haplotype diversity (0.9999 ± 0.0001) and Y-SNP haplogroup diversity (0.7589 ± 0.0171) were observed. Population comparisons at both haplotype and haplogroup levels showed significant differences between Brazilian South Eastern and Northern populations that can be explained by differences in the proportion of African and Native American Y chromosomes. Statistical significant differences between admixed urban samples from the five regions of Brazil were not previously detected at haplotype level based on smaller size samples from South East, which emphasizes the importance of sample size to detected population stratification for an accurate interpretation of profile matches in kinship and forensic casework. Although not having an intra-population discrimination power as high as the Y-STRs, the Y-SNPs are more powerful to disclose differences in admixed populations. In this study, the combined analysis of these two types of markers proved to be a good strategy to predict population sub-structure, which should be taken into account when delineating forensic database strategies for Y chromosome haplotypes.

© 2014 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Brazil is a genetically highly diverse country as a result of extensive admixture between Native American, European and African populations in a vast and very heterogeneous territory [1,2].

Different patterns of genetic ancestry were revealed among the Brazilian populations from different regions, when assessed through mtDNA and autosomal markers (e.g. [3–9]). Nevertheless, no important population stratification have been disclosed until now for the Y chromosome gene pool, except for few African and

Native American descendent groups that are spread all over the country (e.g. [10–12]). A recent study from Palha et al. [13] showed no significant differentiation for 17 populations, distributed among the 5 geopolitical regions of Brazil, which was attributed to the high frequency of European male lineages in all populations.

Owing to the non-recombining nature of the Y chromosome, in forensic genetics it is important to study a large number of markers at the same time to increase levels of discrimination within populations. Concerning Brazil, data exist for a total of 17 states including 23 Y-STRs [13,14]. For the 17 Y-STR loci in the YFiler kit, there are data available for a total of 18 out of the 26 Brazilian states [13–18]; and haplotypic data for the remaining states have been published for smaller Y-STR sets [19].

Although less powerful to discriminate male lineages inside populations due to their low mutation rates and recurrence, Y-SNPs are useful to define haplogroups with high geographical specificity and are therefore more apt to detect population stratification in admixed populations from diverse continental origins.

* Corresponding author at: DNA Diagnostic Laboratory (LDD), Pavilhão Haroldo Lisboa da Cunha, UERJ – Universidade do Estado do Rio de Janeiro, São Francisco Xavier, 524, Maracanã, CEP: 20550-900 Rio de Janeiro, RJ, Brazil.
Tel.: +55 21 23342183; fax: +55 21 23340594.

E-mail address: lgusmao@ipatimup.pt (L. Gusmão).

<http://dx.doi.org/10.1016/j.fsigen.2014.08.017>

1872-4973/© 2014 Elsevier Ireland Ltd. All rights reserved.

Studies using Y-SNPs to characterize admixed populations in Brazil are few, and covering a small number of populations (e.g. Rio de Janeiro [20], Alagoas [21], Bahia [22], Manaus and Ribeirão Preto [23]).

In this work, a large sample of 605 non-related males from Rio de Janeiro population was genotyped for an extended group of both Y-STRs and Y-SNPs aiming a better characterization of Brazilian population stratification in what concerns male lineages contributions, allowing a more accurate interpretation of profile matches in kinship and forensic casework, as well as a deeper insight on the microevolution and demographic history of South America male gene pool.

2. Methods and materials

2.1. Population samples

A total of 605 samples from unrelated males living in Rio de Janeiro district, Brazil, were selected for this study (Fig. S1). Samples were obtained from paternity casework from the DNA Diagnostic Laboratory (LDD) with written informed consent of the donors. This study, together with the informed consent, was approved by the Ethics Committee of the State University of Rio de Janeiro (CAAE: 0067.0.228.000-09).

A total of 149 samples from Rio de Janeiro from previous reports [20,24,25] were also included in the present study, after being typed for 5 new Y-STRs and the full set of Y-SNPs.

2.2. Y chromosome STR typing

The AmpFISTR Yfiler kit was used to analyze 17 Y-STR loci (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS385, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS635 and GATA H4), following manufacturer's recommendations. Typing was carried out using ABI 3130 and ABI 3500 Genetic Analysers (Applied Biosystems) and analysis was performed using GeneMapper® ID Software v4.0 (Applied Biosystems). The Y-STR alleles were designated according to the ISFG recommendations [26].

Y haplotype data are included in the Y-STR Haplotype Reference Database (<http://www.yhrd.org>) under the following accession number: Rio de Janeiro, Brazil [Admixed] YA003720.

2.3. Y-chromosome SNP typing

In this study, 46 Y-SNPs were analyzed (Fig. S2). Haplogroup nomenclature follows Dulik et al. [27] for haplogroup Q, Trombetta et al. [28] for haplogroup E, Myres et al. [29] for haplogroup R, and Karafet et al. [30] for the remaining haplogroups. The Y-SNPs were chosen to discriminate the most frequent Native American, European and African haplogroups.

The 46 Y-SNPs were genotyped by PCR and SNaPshot reactions using the different multiplexes indicated in Supplementary Fig. S2. Multiplexes were carried out as previously described by Gomes et al. [31], except for multiplex Q that was performed as described by Vullo et al. [32]. The first multiplex to be typed in each sample was chosen after inferring the most likely haplogroup employing the freely available haplogroup predictor software (<http://www.hprg.com/hapest5/index.html>). In cases where no derived alleles were found using the first multiplex, the remaining multiplexes were also investigated and haplogroup classification was based on the results obtained. Besides not discriminating some sub-lineages inside haplogroups, the prediction program misclassified approximately 1% of the samples belonging to European haplogroups, and all samples belonging to the African haplogroups B2a*-M150 and E2b1-M85.

2.4. Data analysis

Haplogroup frequencies were determined by direct counting. Population comparisons, haplotype frequencies and genetic diversities according to Nei [33] were estimated with the software Arlequin 3.5.1.2 [34]. In population comparisons, the number of repetitions in DYS389I was subtracted from DYS389II and DYS385 was excluded. To match Arlequin STR data format: (a) duplicated alleles were represented as missing data; (b) null and non-consensus alleles were also represented as missing data and an additional locus was inserted with non-null/consensus alleles coded as 1 and null/intermediate alleles as 10 repeats.

Population pairwise genetic distances were conducted based on F_{ST} and R_{ST} and the significance was tested with 10,000 permutations. Pairwise genetic distances were visualized in two dimensional space using the multi-dimensional scaling (MDS) analysis included in the StatSoft, Inc. (2007) program, STATISTICA (data analysis software system), ver.8.0 (www.statsoft.com).

3. Results and discussion

The genotyping results of a population sample of 605 unrelated individuals from Rio de Janeiro for 17 Y-STR and 46 Y-SNP markers are listed in Table S1.

3.1. Y-chromosomal diversity

A high Y-STR based haplotype diversity was observed (0.9999 ± 0.0001); with just 12 haplotypes being shared by two individuals and just one observed three times. The haplotype diversity value observed in Rio de Janeiro was similar to those previously obtained for other samples from Brazil that have been studied for the same set of Y chromosome STR markers [13–18].

The observed haplogroup diversity was 0.7589 (± 0.0171) that can be considered high when compared to the values estimated (for the same group of Y-SNPs) in Portuguese (0.652) [35], in Sub-Saharan African Bantu groups (0.681 and 0.753 in Angola [36] and Equatorial Guinea [37], respectively) and in Native Americans (0.265 and 0.709 for Toba and Colla populations, respectively) [38], which is explained by the presence of haplogroups from very diverse continental origins.

The analysis of the Y-SNPs did not allow discriminating lineages with identical Y-STR haplotypes, i.e. all chromosomes sharing the same haplotype also belong to the same haplogroup.

3.2. Comparison of the Y chromosome haplotypes in Brazil

Haplotype frequencies in Rio de Janeiro were compared with those from samples from other 17 Brazilian states, which have been studied for the same markers (see states location and Y-STR haplotype data references in Fig. S1). Population comparisons were performed through F_{ST} analysis based on number of different alleles (FST-like) and sum of squared differences (RST-like), after excluding the DYS385 from the AmpFISTR Yfiler markers' set (see results in Table S2 and Fig. S3). The results revealed no significant differences between Rio de Janeiro and two other populations from South East Brazil (São Paulo and Minas Gerais), for a significance level of 0.00031 obtained after applying the Bonferroni correction for 161 pairwise comparisons. When comparing Rio de Janeiro with samples from other regions, significant differences were only found with two populations from the North region (Pará and Roraima), when accessed by both F_{ST} and R_{ST} values.

Samples were pooled to represent the 5 geopolitical Brazilian regions (see the delimitations of the geopolitical regions in Fig. S1) and compared with each other, as well as with reference samples from Portuguese (pooled data from [13,14,39]), Africans [13,37,40]

Table 1

Matrix showing the pairwise F_{ST} values (below diagonal) between Brazilian regions and reference samples from Africa (AFR), Europe (EUR) and Native America (NAM), obtained for the haplotypes defined by the 16-markers set. The corresponding differentiation p values are indicated in the above diagonal.

	N	NE	CW	SE	S	AFR	EUR	NAM
FST-like								
N	*	0.00020	0.00525	<SE–06	<SE–06	<SE–06	<SE–06	<SE–06
NE	0.00568	*	0.07841	0.05267	0.04366	<SE–06	0.00436	<SE–06
CW	0.00246	0.00152	*	0.03366	0.30334	<SE–06	0.39452	<SE–06
SE	0.00383	0.00131	0.00133	*	0.01525	<SE–06	0.00178	<SE–06
S	0.00418	0.00148	0.00018	0.00091	*	<SE–06	0.06772	<SE–06
AFR	0.14984	0.13059	0.15733	0.12772	0.14404	*	<SE–06	<SE–06
EUR	0.00551	0.00505	0.00005	0.00346	0.00112	0.17371	*	<SE–06
NAM	0.14155	0.14924	0.15930	0.13724	0.14544	0.20796	0.17148	*
RST-like								
N	*	<SE–06	0.00218	<SE–06	<SE–06	<SE–06	0.05039	<SE–06
NE	0.01917	*	0.00208	0.03485	0.00337	<SE–06	<SE–06	<SE–06
CW	0.00482	0.00804	*	0.01713	0.06950	<SE–06	0.06732	<SE–06
SE	0.00851	0.00249	0.00271	*	0.01416	<SE–06	0.00119	<SE–06
S	0.00763	0.00498	0.00160	0.00146	*	<SE–06	0.00693	<SE–06
AFR	0.27659	0.20360	0.26386	0.21337	0.24235	*	<SE–06	<SE–06
EUR	0.00205	0.01310	0.00238	0.00582	0.00399	0.29182	*	<SE–06
NAM	0.17778	0.22505	0.18906	0.16791	0.18235	0.48162	0.21607	*

Note: Significant value after applying Bonferroni's correction ($p < 0.00179$) are in bold.

and Brazilian Native Americans [13,41]. South East, the region to each Rio de Janeiro belongs, presented significant differences with the North, which is indeed the most differentiated region of Brazil (Table 1). When compared with the three reference samples, South East Brazil is more close to the Europeans than to Africans or Native Americans, although significant differences could be detected with European samples from Portugal, based on F_{ST} genetic distance (Table 1). Nonetheless, it is worth noting that South East region presents the smallest distance observed with Africans (Table 1).

3.3. Y chromosome haplogroups in Rio de Janeiro

A total of 23 different haplogroups were detected in the sample from Rio de Janeiro population, which ancestry can be traced to Europe, Africa or Native America (Table 2). As expected from the STR results, as well as from previous publications concerning the paternal ancestry of Brazilian admixed populations (e.g. [1,20,21,23,42,43]), the Y chromosomes in Rio de Janeiro belong

Table 2

Absolute and relative frequencies of 23 Y chromosome haplogroups found in 605 unrelated male samples from Rio de Janeiro population; and inferred recent African, European and American continental origin of each haplogroup.

Haplogroup	Inferred origin	N	Freq. (%)
B2a*-M150	Africa	4	0.66
DE*-YAP	Africa	3	0.50
E1b1a1*-M2	Africa	47	7.77
E1b1a1F-M191	Africa	26	4.30
E1b1a1g1c-M154	Africa	2	0.33
E1b1b1*-M35	Africa	3	0.50
E1b1b1a1-M78	Africa/Europe	29	4.79
E1b1b1b1-M81	Europe	32	5.29
E1b1b1c-M123	Europe	6	0.99
E2b1-M85	Africa	5	0.83
F*-M213	Europe	1	0.17
G-M201	Europe	22	3.64
I*-M170	Europe	41	6.78
I2a2-M26	Europe	10	1.65
J*-12f2a	Europe	10	1.65
J2-M172	Europe	39	6.45
K*-M9	Europe	4	0.66
L-M22	Europe	4	0.66
Q1a3a1a*-M3	America	14	2.31
R1a-SRY10831.2	Europe	8	1.32
R1b1b-M269	Europe	282	46.61
T-M70	Europe	12	1.98
Y*(xA,B,DE*,F*)-M168	Unknown	1	0.17

mainly to typically European haplogroups, followed by Africans, and just a small proportion of about 2% belong to Native American haplogroups. However, the frequency of African haplogroups increased from 8 to 15% when compared with the results of a previous study on a smaller sample from the same population [20].

Concerning Y chromosome SNP studies in Brazil, most have been focused on small isolated Afro-descendent or Native groups, or self-declared ethnic groups inside admixed populations (e.g. [3,6,44]). Moreover, few of them report the three main continental ancestral compositions of Brazilians in large admixed populations, which are the most relevant information in the forensic perspective. In Table 3, we compare the results from different studies that inferred ancestry proportions of Y chromosomes in Brazilian population groups. The results show a higher proportion of African haplogroups in Rio de Janeiro than in other populations; a value more than three times higher than the previous estimate for a general sample of 50 individuals from the same geopolitical region. The Native American component is consistently low in all studies, except for Manaus, in the North region, where it reaches a frequency of almost 10%.

4. Conclusions

In forensic genetics, it is important to accurately estimate the frequency of relevant profiles in reference populations, taking into account population boundaries and possible effects of sub-structure.

The size and demographic complexity of Brazil requires large studies including samples from different representative regions to disclose the full genetic diversity and its stratification across the country.

Concerning Y chromosome haplotype frequencies in Rio de Janeiro, previous publications only included a restricted number of

Table 3

Comparison of the Y chromosome ancestry proportions inferred based on Y-SNP haplogroups found in different population groups from Brazil.

Population	n	NAM (%)	EUR (%)	AFR (%)
Rio de Janeiro [this study]	605	2.32	81.95	15.73
Alagoas [21]	247	0.81	94.74	4.45
Ribeirão Preto [23]	65	1.54	95.38	3.08
Manaus [23]	42	9.52	88.10	2.38
North region [43]	49	0	98.00	2.00
North East [43]	49	0	95.90	4.10
South [43]	52	0	100.00	0.00
South East [43]	50	0	96.00	4.00

markers and/or samples. In the present work, the study of a large sample of 605 unrelated males living in Rio de Janeiro, for a high number of Y-STRs and Y-SNPs, showed no significant differences between Rio de Janeiro, São Paulo and Minas Gerais, all from South East region. Nevertheless, significant differences have been detected when comparing South East and North region populations, which were not previously noticed when using smaller size samples. Because of the very high levels of intra-population diversity that are associated with haplotypes defined by a high number of STRs, large samples are necessary to statistically support small differences between populations, or to detect population stratification.

The study of Y-SNPs in Rio de Janeiro revealed a high percentage of European haplogroups (81.95%), together with a noticeable proportion of African haplogroups (15.73%), which is usually lower in most urban admixed populations.

The main European contribution to the Y chromosome gene pool of most Brazilian populations explains the difficulty to detect significant population differences for Y-haplotypes in most studies, based on small size samples for extremely diverse haplotypes associated to large numbers of Y-STR markers. However, differences in Native American, European and African derived pools can be easily detected using SNPs, foreseeing differences in haplotype frequencies in large population databases.

Acknowledgements

This work was supported by FAPERJ (Fundação de Amparo à Pesquisa do Rio de Janeiro), Programa APQ1 2013/01, processo E-26/111.395/2013. IPATIMUP is an associate laboratory of the Portuguese Ministry of Science, Technology and Higher Education, and is partially supported by FCT, the Portuguese Foundation for Science and Technology. V.G. is supported by FCT through grant SFRH/BPD/76207/2011.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.fsigen.2014.08.017.

References

- [1] K. Abe-Sandes, W.A. Silva Jr., M.A. Zago, Heterogeneity of the Y chromosome in Afro-Brazilian populations, *Hum. Biol.* 76 (2004) 77–86.
- [2] M.D. Priore, R. Venâncio, *Uma Breve História do Brasil*, first ed., Planeta do Brasil, São Paulo, 2010.
- [3] M.C. Bortolini, W.A. Silva Jr., D.C. De Guerra, G. Remonato, R. Miranda, African-derived South American populations: a history of symmetrical and asymmetrical matings according to sex revealed by bi- and uni-parental genetic markers, *Am. J. Hum. Biol.* 11 (1999) 551–563.
- [4] J. Alves-Silva, M. da Silva Santos, P.E. Guimarães, A.C. Ferreira, H.J. Bandelt, The ancestry of Brazilian mtDNA lineages, *Am. J. Hum. Genet.* 67 (2000) 444–461.
- [5] A.K. Ribeiro-dos-Santos, J.M. Pereira, M.R. Lobato, B.M. Carvalho, J.F. Guerreiro, S.E. Batista Dos Santos, Dissimilarities in the process of formation of Curiaú, a semi-isolated Afro-Brazilian population of the Amazon region, *Am. J. Hum. Biol.* 14 (2002) 440–447.
- [6] T. Hunemeier, C. Carvalho, A.R. Marrero, F.M. Salzano, S.D. Pena, Niger-Congo speaking populations and the formation of the Brazilian gene pool: mtDNA and Y-chromosome data, *Am. J. Phys. Anthropol.* 133 (2007) 854–867.
- [7] N.P. Santos, E.M. Ribeiro-Rodrigues, A.K. Ribeiro-Dos-Santos, R. Pereira, L. Gusmão, Assessing individual interethnic admixture and population substructure using a 48-insertion-deletion (INSEL) ancestry-informative marker (AIM) panel, *Hum. Mutat.* 31 (2010) 184–190.
- [8] T.C. Lins, R.G. Vieira, B.S. Abreu, D. Grattapaglia, R.W. Pereira, Genetic composition of Brazilian population samples based on a set of twenty-eight ancestry informative SNPs, *Am. J. Hum. Biol.* 22 (2010) 187–192.
- [9] F.S. Manta, R. Pereira, A. Caiá, D.A. Silva, L. Gusmão, E.F. Carvalho, Analysis of genetic ancestry in the admixed Brazilian population from Rio de Janeiro using 46 autosomal ancestry-informative indel markers, *Ann. Hum. Biol.* 40 (2013) 94–98.
- [10] F.P. Leite, S.M. Callegari-Jacques, B.A. Carvalho, T. Kommers, C.H. Matte, Y-STR analysis in Brazilian and South Amerindian populations, *Am. J. Hum. Biol.* 20 (2008) 359–363.
- [11] T.J. Palha, E.M. Ribeiro-Rodrigues, A. Ribeiro-dos-Santos, J.F. Guerreiro, L.S. Moura, Male ancestry structure and interethnic admixture in African-descent communities from the Amazon as revealed by Y-chromosome STRs, *Am. J. Phys. Anthropol.* 144 (2011) 471–478.
- [12] C.E. Wiesel, M.R. Luizon, S.M. Sousa, L.M. Santos, Y.C. Muniz, Y-linked microsatellites in Amazonian Amerindians applied to ancestry estimates in Brazilian Afro-derived populations, *Am. J. Hum. Biol.* 25 (2013) 313–317.
- [13] T.J. Palha, L. Gusmão, E. Ribeiro-Rodrigues, J.F. Guerreiro, A. Ribeiro dos Santos, Disclosing the genetic structure of Brazil through analysis of male lineages with highly discriminating haplotypes, *PLoS ONE* 7 (2012).
- [14] J. Purps, S. Siegert, S. Willuweit, M. Nagy, C. Alves, R. Salazar, S.M. Angustia, L.H. Santos, K. Anslinger, B. Bayer, Q. Ayub, W. Wei, Y. Xue, C. Tyler-Smith, M.B. Bafluy, B. Martínez-Jarreta, B. Egged, B. Balitzki, S. Tschumi, D. Ballard, D.S. Court, X. Barrantes, G. Bähler, T. Wiest, B. Berger, H. Niederstätter, W. Parson, C. Davis, B. Budowle, H. Buri, 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. Davison, R. Decorte, M.H. Larmuseau, C. Ottoni, O. Rickards, D. Lu, C. Jiang, T. Dobosz, A. Jonkisz, W.E. Frank, I. Fura, C. Gehrig, V. Castella, B. Grskovic, C. Haas, J. Wobst, G. Hadzic, K. Drobnik, 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. Kraaijenbrink, A. Kondili, P. Miniati, M. Vouropoulou, L. Kovacevic, D. Marjanovic, I. Lindner, I. Mansour, M. Al-Azem, A.E. Andari, M. Marino, S. Furfuro, L. Locarno, P. Martin, G.M. Luque, A. Alonso, L.S. Miranda, H. Moreira, N. Mizuno, Y. Iwashima, R.S. Neto, T.L. Nogueira, R. Silva, M. Nastainczyk-Wulf, J. Edelmann, M. Kohl, S. Nie, X. Wang, B. Cheng, C. Núñez, M.M. Pancorbo, J.K. Olofsson, N. Morling, V. Onofri, A. Tagliabracci, H. Pamjav, A. Volgyi, G. Barany, R. Pawlowski, A. Maciejewska, S. Pelotti, W. Pepinski, M. Abreu-Glowacka, C. Phillips, J. Cárdenas, D. Rey-Gonzalez, A. Salas, F. Brisighelli, C. Capelli, U. Toscanini, A. Piccinini, M. Piglionica, S.L. Baldassarra, R. Ploski, M. Konarzewska, E. Jastrzebska, C. Robino, A. Sajantila, J.U. Palo, E. Guevara, J. Salvador, M.C. Ungria, J.J. Rodriguez, U. Schmidt, N. Schlauderer, P. Saukkou, P.M. Schneider, M. Sirkker, K.J. Shin, Y.N. Oh, I. Skitsa, A. Ampati, T.G. Smith, L.S. Calvit, V. Stenzl, T. Capal, 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. 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.
- [15] P. Sánchez-Diz, C. Alves, E.F. Carvalho, M. Carvalho, R. Espinheira, O. García, L. Pontes, M.J. Porto, D. Santapa, Population and segregation data on 17 Y-STRs: results of a GEP-ISFG collaborative study, *Int. J. Legal Med.* 122 (2008) 529–533.
- [16] S.P. Schwengber, T. Kommers, C.H. Matte, P.E. Raimann, B.A. Carvalho, F.P. Leite, M.A. Medeiros, L.F. Souza, C.S. Castro, F.G. Chassot, S.L. Bonatto, Population data of 17 Y-STR loci from Rio Grande do Sul state (South Brazil), *Forensic Sci. Int. Genet.* 4 (2009) e31–e33.
- [17] L.M. Gainé, M.M. de Pancorbo, F. Pinheiro, Y-chromosomal STR haplotype diversity in males from Santa Catarina, *J. Forensic Leg. Med.* 17 (2010) 92–95.
- [18] G.B. Barra, C.F. Chianca, T.H. Santa Rita, L.F. Velasco, C.F. de Souza, Haplotype diversity of 23 Y-chromosomal STRs in a population sample from the Federal District (Brazil) – a territory that arose from nothing, *Int. J. Legal Med.* (2014), <http://dx.doi.org/10.1007/s00414-013-0960-2> (Epub ahead of print).
- [19] T.C. Vieira, M.A. Gigorczak, D.M. Silva, R.G. Rodvalho, G.S. Santos, A.D. da Cruz, Y-STR haplotype diversity and population data for Central Brazil: implications for environmental forensics and paternity testing, *Genet. Mol. Res.* 13 (2014) 3404–3410.
- [20] D.A. Silva, E.F. Carvalho, G. Costa, L. Tavares, A. Amorim, L. Gusmão, Y-chromosome genetic variation in Rio de Janeiro population, *Am. J. Hum. Biol.* 18 (2006) 829–837.
- [21] D. Azevedo, L.A.F. Silva, L. Gusmão, E.F. Carvalho, Analysis of Y chromosome SNPs in Alagoas, Northeastern Brazil, *Forensic Sci. Int.: Genet. Suppl. Ser.* 2 (2009) 421–422.
- [22] E. Nascimento, E. Cerqueira, E. Azevedo, W. Freitas, D. Azevedo, The Africa male lineages of Bahia's people – North East Brazil: a preliminary SNPs study, *Forensic Sci. Int.: Genet. Suppl. Ser.* 2 (2009) 349–350.
- [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.
- [24] A.C.S. Góes, E.F. Carvalho, I. Gomes, D.S. Silva, E.H.F. Gil, 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.
- [25] P.M. Domingues, L. Gusmão, D.A. da Silva, A. Amorim, R.W. Pereira, E.F. Carvalho, Sub-Saharan Africa descendents in Rio de Janeiro (Brazil): population and mutational data for 12 Y-STR loci, *Int. J. Legal Med.* 121 (2007) 238–241.
- [26] L. Gusmão, J.M. Butler, A. Carracedo, P. Gill, M. Kayser, W.R. Mayr, N. Morling, M. Prinz, L. Roewer, C. Tyler-Smith, P.M. Schneider, DNA commission of the International Society for Forensic Genetics (ISFG): an update of the recommendations on the use of Y-STRs in forensic analysis, *Int. J. Legal Med.* 120 (2006) 191–200.
- [27] M.C. Dulik, A.C. Owings, J.B. Gaiakis, M.G. Vilar, A. Andre, C. Lennie, M.A. Mackenzie, I. Kritch, S. Snowshoe, R. Wright, J. Martin, N. Gibson, T.D. Andrews, T.G. Schurr, Y-chromosome analysis reveals genetic divergence and new founding native lineages in Athapaskan- and Eskimoan-speaking populations, *Proc. Natl. Acad. Sci. U S A* 109 (2012) 8471–8476.
- [28] B. Trombetta, F. Cruciani, D. Sellitto, R. Scozzari, A new topology of the human Y chromosome haplogroup E1b1 (E-P2) revealed through the use of newly characterized binary polymorphisms, *PLoS ONE* 6 (2011) e16073.
- [29] N.M. Myres, S. Rootsi, A.A. Lin, M. Järve, R.J. King, I. Kutuev, V.M. Cabrera, E.K. Khusnudinova, A. Pshenichnov, B. Yunusbayev, O. Balanovsky, E. Balanovsky, P. Rudan, M. Baldovino, R.J. Herrera, J. Chiaroni, J. Di Cristofaro, R. Villem, T. Kivisild, P.A. Underhill, A major Y-chromosome haplogroup R1b Holocene era founder effect in Central and Western Europe, *Eur. J. Hum. Genet.* 19 (2011) 95–101.

- [30] T.M. Karafet, F.L. Mendez, M.B. Meilerman, P.A. Underhill, S.L. Zegura, M.F. Hammer, New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree, *Genome Res.* 18 (2008) 830–838.
- [31] 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.
- [32] C. Vullo, V. Gomes, C. Romanini, A.M. Oliveira, O. Rocabado, J. Aquino, A. Amorim, L. Gusmão, Association between Y haplogroups and autosomal AIMS reveals intra-population substructure in Bolivian populations, *Int. J. Legal Med.* (2014), <http://dx.doi.org/10.1007/s00414-014-1025-x> (Epub ahead of print).
- [33] M. Nei, *Molecular Evolutionary Genetics*, Columbia University Press, New York, 1987.
- [34] L. Excoffier, H.E. Lischer, Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses Linux and Windows, *Mol. Ecol. Resour.* 10 (2010) 564–567.
- [35] S. Beleza, L. Gusmão, A. Lopes, C. Alves, I. Gomes, M. Giouzeli, F. Calafell, A. Carracedo, A. Amorim, Micro-phylogeographic and demographic history of Portuguese male lineages, *Ann. Hum. Genet.* 70 (2006) 181–194.
- [36] S. Beleza, L. Gusmão, A. Amorim, A. Carracedo, A. Salas, The genetic legacy of western Bantu migrations, *Hum. Genet.* 117 (2005) 366–375.
- [37] M. González, V. Gomes, A.M. López-Parra, A. Amorim, A. Carracedo, P. Sánchez-Diz, E. Arroyo-Pardo, L. Gusmão, The genetic landscape of Equatorial Guinea and the origin and migration routes of the Y Chromosome Haplogroup R-V88, *Eur. J. Hum. Genet.* 21 (2013) 324–331.
- [38] U. Toscanini, L. Gusmão, G. Berardi, V. Gomes, A. Amorim, A. Salas, E. Raimondi, Male lineages in South American native groups: evidence of M19 traveling south, *Am. J. Phys. Anthropol.* 146 (2011) 188–196.
- [39] M.L. Pontes, L. Cainé, D. Abrantes, G. Lima, M.F. Pinheiro, Allele frequencies and population data for 17 Y-STR loci (AmpF/STR® Y-filer™) in a Northern Portuguese population sample, *Forensic Sci. Int.* 170 (2007) 62–67.
- [40] M. Carvalho, P. Brito, A.M. Bento, V. Gomes, H. Antunes, H.A. Costa, V. Lopes, A. Serra, F. Balsa, L. Andrade, M.J. Anjos, F. Corte-Real, L. Gusmão, Paternal and maternal lineages in Guinea-Bissau population, *Forensic Sci. Int. Genet.* 5 (2011) 114–116.
- [41] L. Roewer, M. Nothnagel, L. Gusmão, V. Gomes, M. Gonzalez, D. Corach, A. Sala, E. Alechine, T. Palha, N. Santos, A. Ribeiro-dos-Santos, M. Geppert, S. Willuweit, M. Nagy, S. Zweyner, M. Baeta, C. Núñez, B. Martínez-Jarreta, F. Gonzalez-Andrade, E.F. Carvalho, D.A. Silva, J.J. Buales, D. Turbon, A. Lopez-Parra, E. Arroyo-Pardo, U. Toscanini, L. Borjas, C. Barletta, E. Ewart, S. Santos, M. Krawczak, Continent-wide decoupling of Y-chromosomal genetic variation from language and geography in native South Americans, *PLoS Genet.* 9 (2013) e1003460.
- [42] S.E. Batista-dos-Santos, J.D. Rodrigues, A.K. Ribeiro-dos-Santos, M.A. Zago, Differential contribution of indigenous men and women to the formation of an urban population in the Amazon region as revealed by mtDNA and Y-DNA, *Am. J. Phys. Anthropol.* 109 (1999) 175–180.
- [43] 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.
- [44] A.R. Marrero, F.P.N. Leite, B.A. Carvalho, L.M. Peres, T.C. Kommers, L.M. Cruz, F.M. Salzano, A. Ruiz-Linares, W.A. Silva Jr., M.C. Bortolini, Heterogeneity of the genome ancestry of individuals classified as White in the State of Rio Grande do Sul, *Am. J. Hum. Biol.* 17 (2005) 496–506.

ANEXO C- 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
 Prof^a. 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. S^a., que ao término da pesquisa encaminhe a esta comissão um sumário dos resultados do projeto.


 Prof^a. Patrícia Maria C. O. Duque
 Membro do Comitê de Ética em Pesquisa
 COMITÊ DE ÉTICA EM PESQUISA
 HUPE/UERJ