UNIVERSIDADE ESTADUAL PAULISTA

"JULIO DE MESQUITA FILHO"

Campus de Araraquara - Instituto de Química

Programa de Biotecnologia

ISABELA BRUNELLI AMBROSIO

AVALIAÇÃO DE LINHAGENS PATRILÍNEAS DE 23 Y-STRs NAS POPULAÇÕES DOS ESTADOS DE SÃO PAULO E RIO DE JANEIRO

ARARAQUARA - SP

2019

ISABELA BRUNELLI AMBROSIO

Avaliação de linhagens patrilíneas de 23 Y-STRs nas populações dos Estados de São Paulo e Rio de Janeiro

Tese apresentada ao Instituto de Química, Universidade Estadual Paulista "Júlio de Mesquita Filho", como parte dos requisitos para obtenção do título de Doutora em Biotecnologia.

Área de concentração: Biologia Molecular

ORIENTADORA: Profa. Dra. Regina Maria Barretto Cicarelli

ARARAQUARA
2019

FICHA CATALOGRÁFICA

A496a

Ambrosio, Isabela Brunelli
Avaliação de linhagens patrilíneas de 23 Y-STRs nas populações dos estados de São Paulo e Rio de Janeiro / Isabela Brunelli Ambrosio. – Araraquara : [s.n.], 2019
175 f. : il.

Tese (doutorado) – Universidade Estadual Paulista, Instituto de Química
Orientador: Regina Maria Barretto Cicarelli

1. Cromossomo Y. 2. Polimorfismo (Genética). 3. DNA - Análise. 4. Genética da população humana. 5. Genética forense. I. Título.

Bibliotecária Responsável: Valéria Aparecida Moreira Novelli - CRB/8 4634

UNIVERSIDADE ESTADUAL PAULISTA

Câmpus de Araraquara

CERTIFICADO DE APROVAÇÃO

TÍTULO DA TESE: "Avaliação de linhagens patrilíneas de 23 Y-STR as populações dos estados de São Paulo e Rio de Janeiro"

AUTORA: ISABELA BRUNELLI AMBRÓSIO

ORIENTADORA: REGINA MARIA BARRETTO CICARELLI

Aprovada como parte das exigências para obtenção do Título de Doutora em BIOTECNOLOGIA, pela Comissão Examinadora:

Profª. Drª. REGINA MARIA BARRETTO CICARELLI
Departamento de Ciências Biológicas / Faculdade de Ciências Farmacêuticas - UNESP - Araraquara

Drª JOYCE APARECIDA MARTINS LOPES FERRAZ
Departamento de Toxicologia / Superintendência da Polícia Técnico-Científica - São José do Rio Preto

Profa. Dra. MARIA LEONOR RODRIGUES DE SOUZA BOTELHO DE GUSMÃO
Departamento de Medicina Legal / Universidade do Estado do Rio de Janeiro - UERJ - Rio de Janeiro

Prof. Dr. DANILO FAUSTINO BRAGANHOLI
Departamento de Ciências Biológicas / Faculdade de Ciências Farmacêuticas - UNESP - Araraquara

Prof. Dr. MARCELO MALAGHINI
Laboratório de Genética Molecular Forense / Instituto de Criminalística - LGMF/IC - Curitiba - PR

Araraquara, 08 de agosto de 2019

UNIVERSIDADE ESTADUAL PAULISTA

Câmpus de Araraquara

CERTIFICADO DE APROVAÇÃO

TÍTULO DA TESE: "Avaliação de linhagens patrilíneas de 23 Y-STR na população dos estados de São Paulo e Rio de Janeiro"

AUTORA: ISABELA BRUNELLI AMBRÓSIO
ORIENTADORA: REGINA MARIA BARRETTO CICARELLI

Aprovada como parte das exigências para obtenção do Título de Doutora em BIOTECNOLOGIA, pela Comissão Examinadora:

Profª. Drª. REGINA MARIA BARRETTO CICARELLI
Departamento de Ciências Biológicas / Faculdade de Ciências Farmacêuticas - UNESP - Araraquara

Drª. JOYCE APARECIDA MARTINS LOPES FERRAZ
Departamento de Toxicologia / Superintendência da Polícia Técnico-Científica - São José do Rio Preto

Profa. Dra. MARIA LEONOR RODRIGUES DE SOUZA BOTELHO DE GUSMÃO
Departamento de Medicina Legal / Universidade do Estado do Rio de Janeiro - UERJ - Rio de Janeiro

Prof. Dr. DANILO FAUSTINO BRAGANHOLI
Departamento de Ciências Biológicas / Faculdade de Ciências Farmacêuticas - UNESP - Araraquara

Prof. Dr. MARCELO MALAGHINI
Laboratório de Genética Molecular Forense / Instituto de Criminalística - LGMF/IC - Curitiba - PR

Araraquara, 08 de agosto de 2019

Dedicatória

"Aos meus pais, Antônio e Maria Inês, minhas maiores inspirações e os principais responsáveis por esta conquista."

AGRADECIMENTOS

Aos meus pais, Antônio e Maria Inês, que sempre apoiaram minhas escolhas, acreditando nos meus sonhos, fazendo o possível e o impossível para que eu pudesse realizá-los.

À minha irmã Fernanda e meu marido Davi que sempre estiveram ao meu lado, me dando as maiores forças e os melhores conselhos, os quais sempre me mantiveram no caminho certo.

Aos meus avós, família e amigos que de alguma forma sempre estiveram presentes dando força, acreditando e torcendo pelo meu sucesso tanto pessoal quanto profissional.

À Profa. Regina Maria Barretto Cicarelli, pelas orientações dadas, confiança em mim depositada e conhecimento, o que foi essencial durante esses anos e que contribuíram com minha formação acadêmica, profissional e pessoal. O meu muito obrigado por ter me recebido em seu laboratório e compartilhado seus conhecimentos.

À professora Dra. Leonor Gusmão, que mesmo com inúmeros compromissos profissionais sempre se mostrou solícita, preocupando-se apenas em transmitir seu tamanho conhecimento, nos inspirando a sermos melhores profissionais. Além de ter disponibilizado as amostras da população do Rio de e Janeiro. A senhora foi essencial para a finalização deste trabalho, muito obrigada por tanto.

À Fernanda Polverari, a amiga que o trabalho me deu. Obrigada pela parceria, conversas, risos, pelo apoio e todo o conhecimento compartilhado, sem você essa caminhada não seria a mesma.

Ao Danilo, por todo apoio no decorrer deste projeto, seu conhecimento e experiência profissional foram essenciais.

À Nathália, por toda ajuda e colaboração para desenvolver deste trabalho, além da amizade.

À Bianca, pela grande colaboração na finalização deste trabalho, por ser a grande responsável por nos proporcionar momentos de tantas risadas, pela amizade e claro seus deliciosos cafés pelas manhãs.

À Flávia e Isadora, pessoas e profissionais incríveis. Obrigada pela amizade, pelos conhecimentos transmitidos, trocas de experiência e momentos de descontração. É sempre bom trabalhar com vocês.

Ao Instituto de Química e Faculdade de Ciências Farmacêuticas da Universidade Estadual Paulista Júlio de Mesquita Filho, pelo suporte dado, tanto pelo núcleo profissional quanto pela infraestrutura, sendo possível o desenvolvimento desse projeto.

Ao CNPQ pela bolsa de doutorado concedida.

À FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo), pelo auxílio financeiro concedido durante o andamento deste projeto, sendo possível sua realização.

"É MUITO IMPORTANTE QUE O HOMEM TENHA IDEAIS. SEM ELES, NÃO SE VAI A PARTE ALGUMA. NO ENTANTO, É IRRELEVANTE ALCANÇÁ-LOS OU NÃO. É APENAS NECESSÁRIO MANTÊ-LOS VIVOS E PROCURAR ATINGI-LOS."

DALAI – LAMA

RESUMO

A análise dos polimorfismos do DNA é a melhor ferramenta encontrada para resolução de casos de identificação humana, sendo os marcadores STR (*Short Tandem Repeat*) localizados em regiões autossômicas os principais e mais utilizados para esta finalidade. Os marcadores genéticos presentes nos cromossomos sexuais (X e Y) e no DNA mitocondrial podem auxiliar as análises de forma eficiente. A análise de Y-STRs é uma ferramenta importante, sendo empregada em casos criminais, principalmente os que envolvem crimes sexuais, quando ocorre mistura das amostras do agressor com a da vítima. São também úteis na determinação da ancestralidade biogeográfica paterna, além de permitirem inferir relações de parentesco, principalmente em casos de reconstrução genética (com ausência do suposto pai) e, como uma alternativa confirmatória em casos de exclusão. Mutações espontâneas podem ocorrer em células germinativas paternas, podendo ser transmitidas para os filhos. Com isso, o estudo aprofundado das taxas de mutação para os Y-STRs é muito importante, quer do ponto de vista antropológico, pois permite efetuar a datação das linhagens do cromossomo Y, quer do ponto de vista forense para a interpretação correta dos resultados destes marcadores quando aplicados em processos de investigação de parentesco ou de criminalística biológica. Neste contexto, o uso de um banco de dados da população local pode ser essencial para o cálculo adequado do índice de parentesco e análises forenses. No presente trabalho, foram analisados 23 Y-STRs por meio do kit comercial PowerPlex®Y23 (Promega Corporation, USA) em 401 casos de investigação de paternidade, das populações dos estados de São Paulo e do Rio de Janeiro. Foram observadas 38 mutações em 9.897 transferências alélicas, obtendo-se uma taxa de mutação global de $3,84X10^{-3}$ (95% IC, $2,72X10^{-3}$ a $5,27X10^{-3}$). Em todos os casos foram observados ganho ou perda de apenas 1 unidade de repetição. Avaliou-se também as frequências alélicas e haplotípicas, bem como os parâmetros estatísticos forenses, para os 23 loci do cromossomo Y. A maior diversidade gênica foi observada no locus duplo DYS385 (0,9165) e no DYS481 (0,7940). A diversidade haplotípica se mostrou maior para a população de São Paulo (1,0000 +/- 0,0005) que para a do Rio de Janeiro (0,99995 +/- 0,0005), e este conjunto de marcadores se mostrou com alto

poder de discriminação para ambas as populações (São Paulo - 1,0000; Rio de Janeiro - 0,99502). No estudo comparativo entre as populações, ambas mostraram-se muito próximas geneticamente às populações europeias. A previsão dos haplogrupos confirmou os resultados obtidos entre as distâncias genéticas e a forte influência europeia na linhagem patrilínea de ambas as populações, quando comparada com as africanas e nativo-americana.

Palavras-chave: Y-STR. Taxas de mutação. São Paulo. Rio de Janeiro. Identificação Humana.

ABSTRACT

The DNA polymorphisms analysis is the best tool for solving cases of human identification, and STR (Short Tandem Repeat) markers in autosomal chromosomes are the most used for this purpose. Genetic markers present on the sex chromosomes (X and Y) and mitochondrial DNA can efficiently assist analyzes. Y-STR analysis is an important tool used in criminal cases, especially those involving sexual crimes, when a sample of offender occurs with the victim, paternal biogeographic ancestry, and kinship inference relations, especially cases with alleged father absence, and alternative cases of exclusion as a confirmatory. Spontaneous mutations can occur in the germ cells of parents and can be transmitted for descendants. Thus, in-depth Y-STRs mutation rates studies is important both from the anthropological point of view, since it allows dating the Y chromosome lineages, or in forensic point of view for results correct interpretation these markers when applied in kinship investigation processes or biological criminalistics. In addition, in order to determine the profile probability found to belong to the individual tested, it would be necessary to use a local population database to calculate paternity index and forensic analysis. In this study, we analyzed 401 cases related to paternity research with 23 Y-STRs panel, from São Paulo and Rio de Janeiro population. A total of 38 mutations were identified in 9,897 father-son allelic transfers, obtaining an overall mutation rate of 3.84x10-3 (95% CI, 2.72x10-3 to 5.27x10-3). In all cases there was only one locus mutated with gain/loss and one repeat was observed. We evaluated haplotype and allele frequencies, as well as statistical forensic parameters, for 23 Y-STR of PowerPlex®Y23 loci. The highest gene diversity was observed for double locus DYS385 (0.9165) and single locus DYS481 (0.7940). The haplotype diversity was higher for São Paulo population (1.0000 +/- 0.0005) than for Rio de Janeiro (0.99995 +/- 0.0005), and this set of markers showed a high power of discrimination in both populations (São Paulo - 1.0000; Rio de Janeiro - 0.99502). In the comparative study between populations, both were very genetically close to European populations. The haplogroup prediction confirmed the results obtained between genetic distances and the strong European influence on the patrilineal lineage of both populations, when compared with to Africans and Native Americans.

Keywords: Y-STR. Mutation Rates. São Paulo. Rio de Janeiro. Human Identification.

LISTA DE FIGURAS

Figura 1. Tipos de polimorfismos que são utilizados na HID..... 24

Figura 2. Representação esquemática dos cromossomos sexuais X e Y..... 28

Figura 3. Esquema representativo da estrutura do cromossomo Y e a localização dos Y-STRs utilizados na HID..... 31

Figura 4. Distribuição dos 23 Y-STRs presentes no kit comercial Powerplex®Y 23..... 32

Figura 5. Eletroferograma do perfil alélico do DNA comercial obtido com o Powerplex®Y 23, seguindo protocolo estabelecido no laboratório..... 57

Figura 6. Eletroferograma do perfil alélico do DNA comercial 2800 obtido com o Powerplex®Y 23 seguindo protocolo adaptado..... 58

Figura 7. Correlação entre o comprimento dos alelos presentes na população estudada para o Powerplex®Y 23 e as mutações observadas..... 65

Figura 8. Eletroferogramas parciais de perfis alélicos de três amostras obtidos com o Powerplex®Y 23 da triplicação no locus DYS19..... 68

Figura 9. Eletroferogramas parciais de perfis alélicos de duas amostras obtidos com o Powerplex®Y 23 do alelo nulo encontrado no locus DYS576..... 69

Figura 10. Valores das diversidades gênicas entre os 23 Y-STRs obtidas com o Powerplex®Y 23 nas populações de São Paulo e Rio de Janeiro..... 73

Figura 11. Média de coincidência entre os haplótipos da população de São Paulo quando confrontados com dados do YHRD para diferentes conjuntos de marcadores Y-STR (Powerplex®Y 23, YFiler®, PowerPlex® Y System e Haplótipo mínimo)..... 77

Figura 12. Média de coincidência entre os haplótipos da população do Rio de Janeiro quando confrontados com dados do YHRD para diferentes conjuntos de marcadores Y-STR (Powerplex®Y 23, YFiler®, PowerPlex® Y System e Haplótipo..... 78

Figura 13. Gráfico MDS bidimensional elaborado a partir das distâncias genéticas (*Fst*) entre as populações estudadas (São Paulo e Rio de Janeiro) com populações de outras regiões brasileiras e provenientes de outros países..... 83

Figura 14. Classificação das amostras de São Paulo segundo a ancestralidade dos haplogrupos..... 88

Figura 15. Classificação das amostras do Rio de Janeiro segundo a ancestralidade dos haplogrupos..... 89

LISTA DE TABELAS

Tabela 1. Descrição dos casos com relações de parentesco confirmadas utilizados nas análises de mutações da população do Estado de São Paulo..... 45

Tabela 2. Descrição dos casos com relações de parentesco confirmadas utilizados nas análises de mutações da população do estado do Rio de Janeiro..... 45

Tabela 3. Comparação da utilização dos reagentes na reação de PCR seguindo recomendações do fabricante e protocolo adaptado..... 47

Tabela 4. Marcadores Y-STR analisados com o kit Powerplex®Y 23 e sequência repetitiva do polimorfismo..... 48

Tabela 5. Informação relativa ao número de transferências alélicas analisadas por marcador referente a cada população estudada..... 50

Tabela 6. Amostras utilizadas para comparação entre as populações, número de haplótipos analisados (N), origem ancestral da população e respectivas referências..... 53

Tabela 7. Número de mutações observadas para as populações de São Paulo e Rio de Janeiro separadamente..... 59

Tabela 8. Dados relativos às mutações observadas nas populações de São Paulo e Rio de Janeiro, respectivos intervalos de confiança (95%), dados disponíveis no YHRD e valores obtidos neste trabalho combinados com os disponíveis no YHRD, referentes a cada marcador Y-STR analisado..... 60

Tabela 9 Dados de todas as mutações observadas nas populações de São Paulo e Rio de Janeiro. 62

Tabela 10. Tipos de mutação observadas nos alelos presentes na população estudada para o Powerplex®Y 23..... 63

Tabela 11. Frequências relativas das mutações, alelos, e tipos de mutações observadas nas populações de São Paulo e Rio de Janeiro..... 64

Tabela 12. Número (N) de variantes alélicas observadas nas populações de São Paulo e Rio de Janeiro referente aos 23 Y-STRs analisados..... 66

Tabela 13. Frequências alélicas, número de alelos e diversidade gênica (DG) dos 23 marcadores Y-STR obtidos com o Powerplex®Y 23 nas populações de São Paulo e Rio de Janeiro..... 71

Tabela 14. Diversidade haplotípica e capacidade de discriminação das populações estudadas e em conjunto com populações de outras regiões do Brasil..... 75

Tabela 15. Coincidência entre os perfis haplotípicos dentre as populações estudadas... 76

Tabela 16. Capacidade de discriminação de diferentes conjuntos de Y-STR da 79

população estudada em comparação com dados da literatura.....

Tabela 17. Distâncias genéticas (*Fst*) entre as populações de São Paulo 1 e Rio de Janeiro 1 deste trabalho com a população de São Paulo 2 e Rio de Janeiro 2..... 80

Tabela 18. Distâncias genéticas (*Fst*) entre populações (abaixo da diagonal) e valor de *p* correspondente (acima da diagonal)..... 81

Tabela 19. Frequência dos haplogrupos na população de São Paulo (N=200) baseada nos resultados obtidos pelo Nevgen..... 86

Tabela 20. Frequência dos haplogrupos na população do Rio de Janeiro (N=201) baseada nos resultados obtidos pelo Nevgen..... 87

LISTA DE ABREVIATURAS E SIGLAS

AS-STR	*Short Tandem Repeat of autossomic chromosomes*
BNPG	Banco Nacional de Perfis Genéticos
CD	Capacidade de discriminação
DG	Diversidade gênica
DNA	Ácido desoxirribonucleico
mtDNA	DNA mitocondrial
nDNA	DNA nuclear
EUA	Estados Unidos da América
HID	*Human IDentification*
HLA	*Histocompatibility leucocyte antigen*
IBGE	Instituto Brasileiro de Geografia e Estatística
IC	Intervalo de confiança
InDel	Inserção/Deleção
IP	Índice de Paternidade
ISFG	*Internacional Society for Forensic Genetics*
Kb	kilo pares de bases (1.000pb)
LD	*Linkage Disequilibrium*
Locus	Região genômica na qual se localizam os alelos
Mb	Mega pares de base (1.000.000pb)
MDS	Multidimensional scaling
MSY	*Male-specific region of the Y chromosome*
NGS	*Next Generation Sequencing*
pb	Par de bases
PCR	*Polymerase Chain Reaction*
PAR	*Pseudoautosomal region*
PPY	PowerPlex® Y

PPY23	Powerplex®Y 23
SWGDAM	*Scientific Working Group on DNA Analysis Methods*
q.s.p.	Quantidade suficiente para
RFU	Relative fluorescence unit
RM	*Rapidly Mutating*
rpm	Rotações por minuto
SNP	*Single Nucleotide Polymorphism*
STR	Short Tandem Repeat
UERJ	Universidade Estadual do Rio de Janeiro
UNESP	Universidade Estadual Paulista Júlio Mesquita Filho
VNTR	*Variable Number of Tandem Repeats*
X-STR	*Short Tandem Repeat of X chromosome*
YHRD	*Y-Chromosome STR Haplotype Reference Database*
Yfiler	*AmpFLSTR® Yfiler® PCR Amplification Kit*
Y-SNPs	*Single Nucleotide Polymorphism of Y chromosome*
Y-STR	*Short Tandem Repeat of Y chromosome*

SUMÁRIO

1. Introdução..... 19

2. Revisão da Literatura..... 21

2.1 - Marcadores genéticos na Identificação Humana..... 22

2.1.1 - *Short Tandem Repeats* (STR)..... 24

2.1.2 - *Single Nucleotide Polymorphisms* (SNPs)..... 25

2.1.3 - Insertion/Deletion Polymorphisms (InDel)..... 26

2.2 - Cromossomo Y..... 27

2.2.1 – Aplicação dos marcadores genéticos do cromossomo Y..... 29

2.2.2 - Utilização de marcadores Y-STR..... 30

2.2.3 - Mutação no cromossomo Y..... 34

2.3 - Bases de dados..... 35

2.4 - Aspectos históricos do povoamento do Brasil..... 36

2.5 – Justificativa..... 38

3. Objetivo..... 40

3.1 - Objetivos específicos..... 41

4. Materiais e Métodos..... 43

4.1 - Aspectos éticos..... 44

4.2 – Casuística..... 44

4.3 - Coletas das amostras biológicas e extração do DNA..... 46

4.4 - Análises dos 23Y-STR..... 46

4.4.1 - Reação de PCR..... 46

4.4.2 - Análise dos produtos amplificados..... 49

4.5 - Análises estatísticas..... 49

4.5.1 - Avaliação de mutações..... 49

4.5.2 - Diversidade genética e de relevância forense..... 51

5. Resultados e Discussão..... 55

5.1 - Padronização da utilização do Kit Powerplex®Y 23..... 56

5.2 - Análises de mutações..... 59

5.3 - Variantes Alélicas..... 66

5.4 - Análises haplotípicas..... 74

5.5 - Comparações genéticas entre as populações..... 79

5.6 - Ancestralidades dos haplogrupos..... 85

6. Conclusões..... 90

7. Referências..... 93

Anexo I..... 103

1. INTRODUÇÃO

A identificação humana (*Human IDentification* – HID) é o processo científico pelo qual se determina a identidade de uma pessoa, por meio do estabelecimento de um conjunto de caracteres que a individualize, fazendo-a igual apenas a si mesma (FRANÇA, 2011). A Genética Forense é um ramo dedicado à identificação de indivíduos ou amostras biológicas, por meio da análise de DNA (ácido desoxirribonucleico), para fins jurídicos, tanto no âmbito do Direito civil quanto no criminal. Para esses fins, são realizados exames periciais em amostras com origem conhecida, que consistem na determinação de um perfil genético, contribuindo para o esclarecimento de relações biológicas de parentesco, identificação de desconhecidos, ou para identificação dos contribuintes de amostras biológicas, relacionadas a processos de natureza criminal, confrontando-as com o suspeito. Desde a sua primeira utilização em um caso de imigração no Reino Unido (JEFFREYS *et al.*, 1985b), as tecnologias de identificação por perfis genéticos têm sofrido grande avanço. Atualmente constitui uma ferramenta indispensável na investigação criminal, sendo aceita rotineiramente em processos judiciais em todo o mundo.

Na espécie humana, a composição genética de cada indivíduo resulta da contribuição de 50% do genoma materno e 50% do paterno. Na composição do genoma apenas 0,1% pode variar, sendo o restante exatamente igual entre todos os indivíduos (KARKI *et al*., 2005). Nessas regiões variáveis do DNA, denominadas polimórficas, estão presentes sequências nucleotídicas utilizadas pela Genética Forense como marcadores moleculares (BUTLER, 2005; HOLMES; RUSSELL; WALLEY, 2003).

2. REVISÃO DA LITERATURA

2.1 Marcadores genéticos na Identificação Humana

Até 1985, a identificação humana era baseada em análises de grupos sanguíneos do sistema ABO (JORDE; WATKINS; BAMSHAD, 2001), verificação de enzimas e pelo complexo do Antígeno Leucocitário Humano - HLA (*Histocompatibility Leucocyte Antigen*) (CALABREZ, 1999). No entanto, estes sistemas proteicos apresentavam baixo nível discriminativo entre os indivíduos apresentando certas limitações tanto na quantidade e tipo de amostra biológica utilizada, bem como na precisão dos resultados (WEEDN; SWARNEN, 1998).

O avanço em biologia molecular iniciou-se em 1985 com o geneticista inglês Alec Jeffreys e seus colaboradores. Este grupo identificou a existência de um padrão aleatório de determinadas regiões de DNA, que diferiam entre os indivíduos, localizadas na região não codificante.

Este padrão é composto por sequências de DNA que se repetem em *tandem*, ou seja, a mesma sequência repete-se um número variável de vezes na mesma região cromossómica e o número dessas repetições varia entre os indivíduos, o que ficou conhecido como DNA "*fingerprinting*" ou impressões digitais do DNA (JEFFREYS; WILSON; THEIN, 1985), porém para a aplicação de tal técnica, era necessária uma quantidade satisfatória de DNA disponível, o que nem sempre ocorre na área de genética forense.

Juntamente a esta descoberta surgiu uma nova metodologia laboratorial, designada por PCR (*Polymerase Chain Reaction*), técnica que permite a multiplicação exponencial do número de cópias de uma única molécula de DNA (SAIKI *et al*., 1988), contribuindo para a realização de análises em todo o tipo de amostras biológicas, mesmo com quantidades mínimas de material biológico disponível. O momento do desenvolvimento da PCR (MULLIS *et al*, 1987) marcou uma grande transição na área da genética forense, permitindo uma ampliação do uso do DNA. Estas características técnicas contribuíram para o desenvolvimento de metodologias inovadoras fundamentais na genética forense e, atualmente, permitem a utilização de reações em *multiplex*, que é a amplificação simultânea de múltiplas regiões do DNA, adicionando mais de um par de iniciadores na reação de PCR.

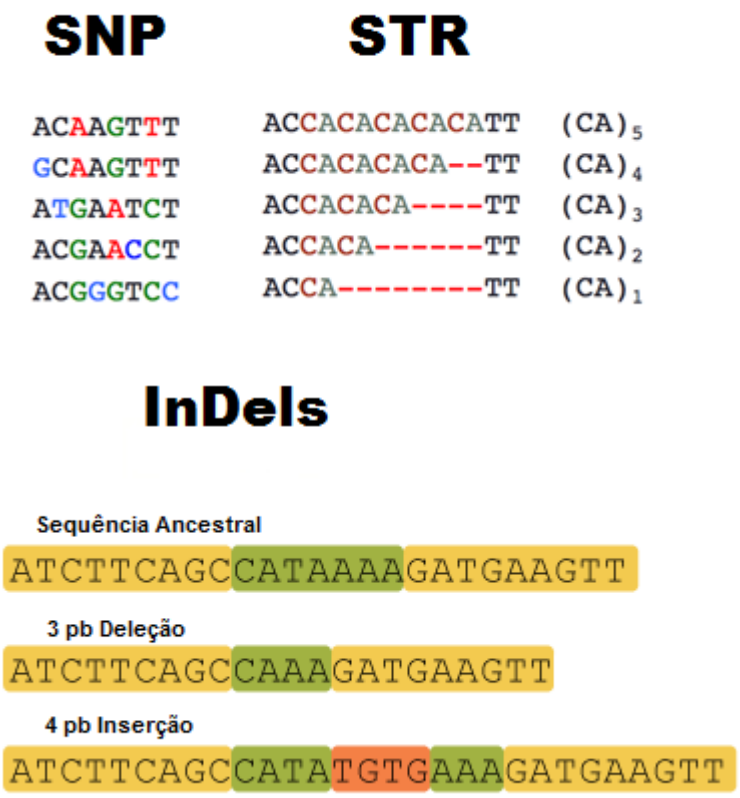
O genoma humano é um conjunto de toda a informação genética existente nas células do organismo. É constituído pelo DNA nuclear (nDNA) e DNA mitocondrial (mtDNA). O nDNA é composto por 46 cromossomos (22 pares de cromossomos homólogos e 1 par de cromossomos sexuais X/Y), sendo biparental, ou seja, é resultado da contribuição de 50% do genoma materno e 50% paterno, estando localizado no interior do nucleo de células somáticas, com apenas uma cópia por célula. O mtDNA está localizado no interior da organela celular mitocôndria em forma de dupla fita e apresenta herança exclusivamente materna (BUTLER, 2005).

Em todos os cromossomos existem sequências com função codificante e outras sequências com função não codificante. As regiões codificantes, correspondentes a 5% do genoma nuclear total e contêm a informação necessária para a síntese de proteínas (SAFERSTEIN, 2006). As regiões não codificantes correspondentes a 90-95% do genoma nuclear, e é a região que mais apresenta variabilidade entre os indivíduos, exceto no caso de gêmeos monozigóticos (KAYSHYAP *et al*., 2004) e, consequentemente, um elevado nível de polimorfismo, sendo por esse motivo muito úteis na identificação genética dos indivíduos e consequentemente em aplicações forenses.

A palavra “polimorfismo” tem origem grega e significa “muitas formas”, Polimorfismos genéticos são variações nas sequências de DNA que ocorrem na população de forma estável. Se a variação é encontrada em uma frequência superior a 1% da população denomina-se polimorfismo. São classificados de acordo com as alterações ocorridas na sequência de DNA, podendo ser polimorfismos de sequência ou de comprimento (BEIGUELMAN, 1996).

Polimorfismos de sequência consistem na substituição de um ou mais nucleotídeos na sequência de DNA. Dentre esses polimorfismos, os *Single Nucleotide Polymorphisms* (SNPs) são a forma mais frequente de variação genética. Os polimorfismos de comprimento são originados por inserções ou deleções de um ou mais nucleotídeos ao longo da sequência, resultando em variações no comprimento. Nesse grupo, além das inserções ou deleções (InDel) de alguns pares de bases (pb), também deve-se destacar as repetições consecutivas, como os *Short Tandem Repeats* (STR) (BUTLER, 2005). Todos os polimorfismos mencionados estão representados na figura 1.

Figura 1: Tipos de polimorfismos que são utilizados na HID.



Fonte: adaptado de www.wiki.uiowa.edu e https://hackbrightacademy.com

2.1.1. Short Tandem Repeats (STRs)

STRs ou Microssatélites são formados pela repetição de unidades com 2 a 7pb que se repetem geralmente de 5 a 30 vezes, definindo as formas variantes dos alelos presentes na população (Figura 1). Essas variações são decorrentes de mutações que favorecem pequenas mudanças, geralmente de uma unidade de repetição, em decorrência de erros de polimerização durante a replicação do DNA, provocados pela enzima DNA polimerase (BUTLER, 2007).

Os STRs estão localizados no nDNA, tanto nos autossomos como nos cromossomos sexuais (X e Y). As sequências de repetição dos STRs são nomeadas pelo comprimento da unidade de repetição, por exemplo, di-, tri- e tetranucleotídeos. Esses três exemplos são mais comuns que penta ou hexanucleotídeos no genoma humano. Os tetranucleotídeos são os mais

utilizados para identificação humana, pois apresentam em média 15% de *stutter* (produtos de amplificação que apresentam uma unidade de repetição a mais ou a menos do que um alelo verdadeiro, que ocorre devido a erro da DNA polimerase durante a PCR). Já os di- e trinucleotídeos apresentam 30% de *stutter* ou mais, o que acaba dificultando a interpretação de perfis genéticos, principalmente em casos de misturas. Além de que os heterozigotos que apresentam alelos distintos de comprimento muito próximos, são melhor discriminados na eletroforese capilar em comparação aos di- e trinucleotídeos (BUTLER, 2011).

Os STRs tornaram-se marcadores de DNA muito utilizados na identificação humana, pois apresentam facilidade de amplificação dessas pequenas regiões ao longo do genoma, podendo ser realizadas em *multiplex*, devido ao fragmento apresentar tamanho reduzido (~100pb a 400pb). Além de apresentar número variável de repetições entre os indivíduos (BUTLER, 2011), já que são marcadores multialélicos. A interpretação dos dados obtidos para estes marcadores é simples, e a detecção (sistemas de eletroforese) é automatizada, fazendo deste um método rápido (BUTLER, 2005).

O caráter multialélico e a herança biparental, a qual permite recombinação alélica entre o par de cromossomos homólogos, conferem maior poder de discriminação aos marcadores STR autossômicos (AS-STR) e, portanto, estes são os mais utilizados na rotina forense (BODNER *et al.*, 2016; ROEWER, 2013, LIN; LI; IP, 2017). No entanto, em casos específicos, marcadores STR dos cromossomos Y (Y-STR) e X (X-STR) são utilizados como ferramentas complementares às análises devido ao seu padrão de herança.

2.1.2. Single Nucleotide Polymorphisms (SNPs)

Os SNPs consistem na substituição de uma única base localizada em um determinado fragmento de DNA (Figura 1), sendo a forma mais simples e comum de variação genética (DAWSON *et al*., 2001). Podem estar localizados no nDNA, tanto nos autossomos como nos cromossomos sexuais ou no mtDNA.

Normalmente apresentam-se como marcadores bialélicos (ancestral e derivado), e podem ocorrer em regiões codificadoras ou com função regulatória, porém, na maior parte das vezes são encontrados em espaços intergênicos, sem função determinada. É o polimorfismo mais estudado do DNA podendo ter grande impacto sobre a forma como os seres humanos respondem às doenças, fatores ambientais, drogas e outras terapias. (SOBRINO; BRION; CARRACEDO, 2005).

Os SNPs estão distribuídos abundantemente no genoma ocorrendo em média entre 1-2Kb (CLIFFORD *et al*., 2000). Por serem marcadores bialélicos, são mais estáveis que os STRs, apresentando baixa taxa de mutação; geralmente são os marcadores de escolha em estudos de origem ancestral (JOBLING; GILL, 2004), evolução e história das populações humanas em escalas de tempo mais profundas (JAKOBSSON *et al*., 2008).

Na área forense, apesar de apresentarem uma diversidade genética menor que os STRs, os SNPs têm várias características desejáveis como marcadores: podem ser analisados em fragmentos muito curtos, o que é importante para melhorar o sucesso de amplificação em DNA altamente degradado, e por apresentarem baixa taxa de mutação, tornam-se interessantes em testes de parentesco complexos, com participação de parentes mais distantes, relacionados por via paterna.

A análise desses marcadores pode ser facilmente automatizada e realizada em *multiplex*. Seu protocolo de amplificação e análise pode ser mais complexo, envolvendo passos de purificação e reação de *single-base extension* (SBE) (PEREIRA; GUSMÃO, 2012), porém atualmente podem ser detectados por diferentes técnicas como *Next Generation Sequencing* (NGS), sequenciamento direto, PCR em tempo real, dentre outras.

2.1.3. Insertion/Deletion Polymorphisms (InDels)

Os polimorfismos de inserção/deleção (InDel) são polimorfismos de comprimento, caracterizados pela inserção ou deleção de um ou mais nucleotídeos em uma determinada região do genoma (Figura 1). A utilização desses marcadores é relativamente recente, sendo identificados e

caracterizados em 2002 por Weber e colaboradores (2002), e então destacados em estudos genéticos (PEREIRA, GUSMÃO, 2012).

Estão presentes tanto nos cromossomos autossômicos quanto nos cromossomos sexuais (X e Y), além de também serem encontrados no DNA mitocondrial. Amplamente distribuídos no genoma, representam 20% de todos os polimorfismos do genoma humano (WEBER, *et al*., 2002), podendo ser encontrados em média a cada 7,2kb (MILLS, 2006).

Esses marcadores apresentam baixa taxa de mutação em comparação com os STRs, e por isso, são potenciais marcadores para estudo de ancestralidade (YANG *et al*., 2005, BRAGANHOLI, 2016). O tamanho do fragmento a ser amplificado na PCR é pequeno (50 a 150pb) o que facilita a análise em amostras degradadas (ZIDKOVA *et al*., 2011). Também podem ser analisados em *multiplex*, sendo as reações tão simples quanto aquelas dos STRs (TORRES *et al.*, 2014).

Todas essas características vêm atraindo cada vez mais a atenção dos cientistas da área de identificação humana, pois torna a análise desses polimorfismos uma ferramenta complementar importante, acrescentando robustez em análises forenses, principalmente em amostras degradadas ou casos com resultado inconclusivo, além de se mostrarem eficientes em avaliações de ancestralidade (BRAGANHOLI, 2016).

2.2. Cromossomo Y

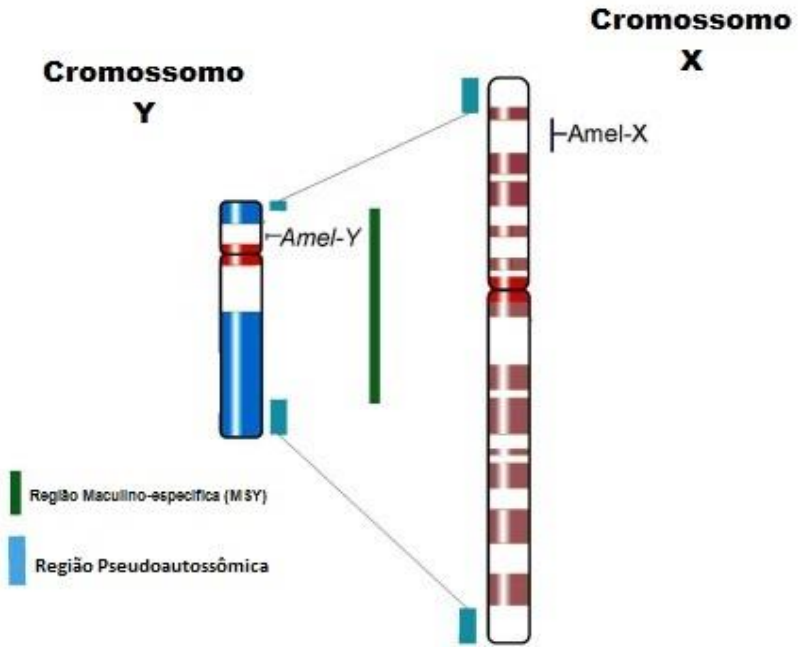
O cromossomo Y, exclusivo do sexo masculino, é um dos menores cromossomos do genoma humano, representando aproximadamente 2% do material nuclear (CORACH; PENACINO; SALA, 1999). Com cerca de 68 milhões de pares de bases, é consideravelmente menor que o cromossomo X (~155Mb). Sua estrutura está dividida em duas regiões pseudoautossômicas (*Pseudoautosomal region* – PAR1 e 2; ~3Mb) e uma região masculino-específica do cromossomo Y (*Male-Specific region of the Y chromosome* – MSY; ~65Mb) (Figura 2).

As regiões PAR1 e PAR2 estão localizadas nas extremidades dos braços curto e longo do cromossomo e são homólogas às regiões do cromossomo X e, portanto, participam na recombinação destes cromossomos

na meiose. Correspondem a aproximadamente 5% da sequência total do cromossomo Y (ROEWER *et al*., 2001; XUE *et al*., 2009).

A região MSY é assim denominada por ser a parte do cromossomo Y que exibe herança paterna clássica, e nela estão localizados os marcadores de interesse em estudos de patrilinhagem. A MSY representa a maior parte do cromossomo Y (95% da sequência total), e não recombina com nenhum outro cromossomo (SKALETSKY *et al*., 2003).

Figura 2: Representação esquemática dos cromossomos sexuais X e Y.



Fonte: adaptado de Butler (2012).

A ausência de recombinação genética deste cromossomo implica que todas as sequências localizadas fora das regiões PAR1 e PAR2 sejam transmitidas às gerações seguintes como um bloco, ou seja, na forma de haplótipos, sem alterações genéticas a não ser que ocorra alguma mutação (CORACH; PENACINO; SALA, 1999, GUSMÃO; BRIÓN; GONZÁLEZ-NEIRA, 2000). Desta forma, é possível afirmar que este cromossomo é o único nos indivíduos do sexo masculino, que é haplóide, por só existir um alelo em cada locus, salvo em situações de duplicação no mesmo locus com herança uniparental paterna, transmitido de pai para filho. Estas características são importantes na área de genética forense, pois permite identificar linhagens

paternas, fazendo deste cromossomo uma poderosa ferramenta de rastreio, traçando a história, movimentos e ancestralidade de populações humanas de forma semelhante ao mtDNA, que é utilizado para estudar linhagens maternas (BOSCH, 1999, KAYSER, 2017).

A MSY humana é constituída por sequências altamente repetitivas, nas quais se localizam diversos Y-STRs. Deste modo, a diversidade genética desta região torna a utilização do cromossomo Y de extrema utilidade na área de identificação humana, sendo utilizada quando a análise de AS-STR não for suficiente para obter conclusões adequadas.

2.2.1 Aplicação dos marcadores genéticos do cromossomo Y

O cromossomo Y, tal como os restantes cromossomos nucleares, apresenta sequências polimórficas diferenciadas, podendo agrupar-se em duas categorias principais: *loci* bialélico, o qual exibe dois possíveis alelos, e *loci* multialélico.

Marcadores bialélicos são constituídos por variantes de sequência, e são classificados em haplogrupos (KAYSER; de KNIJFF, 2011). Neles estão constituídos os Y-SNPs, que apresentam baixas taxas de mutação ($2x10^{-8}$ por base por geração) (CORACH; PENACINO; SALA, 1999), e permitem a construção de uma árvore filogenética única (GUSMÃO; BRIÓN; GONZÁLEZ-NEIRA, 2000).

Os marcadores multialélicos incluem 2 minissatélites (*Variable Number Tandem Repeat* - VNTR) e várias centenas de marcadores de regiões de repetições consecutivas curtas, as variantes de comprimento designadas por Y-STR (*Y-Chromosome Short Tandem Repeats*) (KAYSER *et al*., 2004; BUTLER, 2005). Esses loci multialélicos são caracterizados como haplótipos, e podem ser usados para diferenciar os cromossomos Y com alta resolução devido a sua alta taxa de mutação (KAYSER *et al*., 2000).

A utilização destes marcadores pode ser particularmente útil em casos de paternidade complexos. Os exames de parentesco mais simples envolvem o estudo do trio Suposto Pai/Mãe/Filho; no entanto, existem muitos processos de parentesco que recorrem aos familiares dos supostos progenitores, nomeadamente meios irmãos, supostos avós ou tios paternos, identificando

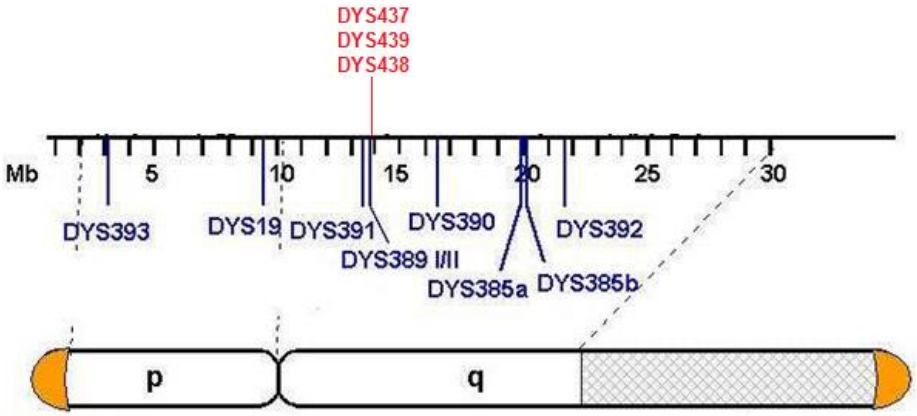
relações de homens com mesma origem patrilinear (FRANCEZ *et al*., 2012). Nesse contexto, os laboratórios forenses têm adotado Y-STR para auxiliar em investigações de parentesco, pois apesar de não permitir diferenciar os indivíduos masculinos relacionados por via paterna, é uma ferramenta muito interessante que permite excluir uma relação de paternidade quando o filho envolvido é do gênero masculino (THOMSEN *et al*, 2009).

A aplicação dos marcadores específicos do cromossomo Y na área forense tem crescido significativamente nas últimas décadas com a descoberta e validação de novos marcadores. O seu valor nesta área é apoiado ao fato de grande parte dos crimes violentos serem efetivamente provocados por um ou vários indivíduos do sexo masculino (JARRETA, 1999).

2.2.2 Utilização de marcadores Y-STR

Os estudos iniciais com estes Y-STR foram efetuados com um STR designado por DYS19 (ROEWER, *et al*., 1993), e desde então foram descobertos e validados para a área forense um grande número de sequências polimórficas importantes, não só para esta área, mas também para estudos populacionais (CORACH; PENACINO; SALA, 1999). O primeiro conjunto de marcadores recomendados para utilização forense ficou conhecido como “haplótipo mínimo”, sendo composto por sete *loci* de cópia única – DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392 e DYS393 – e um *locus* altamente polimórfico, de múltiplas cópias – o DYS385a/b (PASCALLI *et al*., 1999; ROEWER *et al*., 2001) Com o objetivo de incrementar o poder discriminativo destes marcadores genéticos, em janeiro de 2003, o grupo americano *Scientific Working Group on DNA Analysis Methods* (SWGDAM) recomendou a utilização dos marcadores DYS437, DYS438 e DYS439 (BUTLER, 2003, MULERO, *et al*, 2006, GUSMÃO *et al*., 2006); este conjunto ficou denominado de “haplótipo estendido”. Posteriormente a estes 12 marcadores, foram adicionados DYS448, DYS456, DYS458, DYS635 e GATA H4, para garantir um maior poder discriminativo entre as linhagens paternas (GUSMÃO; SILVA, 2010). Subsequentemente, foram reportados diversos loci adicionais que, combinados com os do haplótipo mínimo, aumentaram o poder de discriminação dos marcadores Y-STR.

Figura 3: Esquema representativo da estrutura do cromossomo Y e a localização dos Y-STR utilizados em HID.

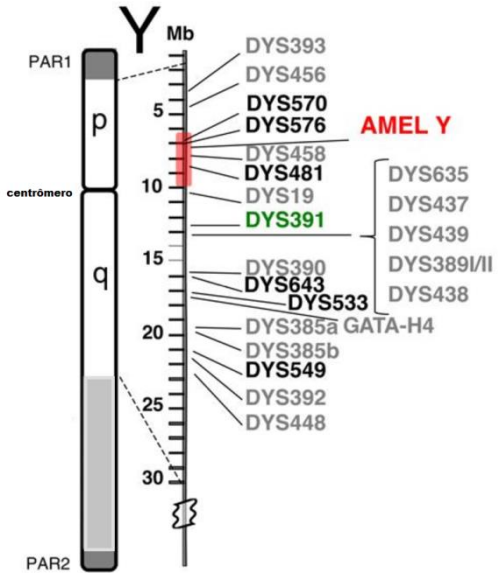


Nota: A letra *p* indica o braço curto do cromossomo e *q* o longo. Em amarelo destacam-se as regiões de recombinação genética (PAR1 e PAR2). Em azul destacam-se os marcadores que englobam o conjunto "haplótipo mínimo" e em vermelhos os marcadores adicionais que em conjunto formam o haplótipo estendido.
Fonte: adaptado de https://strbase.nist.gov//ystrpos1.htm

Com o grande número de polimorfismos Y-STRs que vem sendo descrito, bem como o desenvolvimento de novos kits *multiplex* incorporando um número crescente destes marcadores, como é o caso do kit comercial Powerplex®Y 23, o qual analisa 22 marcadores, totalizando 23 loci. A distribuição destes marcadores no cromossomo Y está indicada na figura 4.

Em todos os marcadores, com exceção do marcador DYS385, há um único alelo, excluindo casos em que ocorre duplicação. O locus DYS385 está localizado em uma sequência palindrômica no braço longo do cromossomo Y, sendo constituída por duas regiões de DNA (denominadas de “a” e “b”) separadas entre si por 40.000pb. Para a amplificação dos alelos de ambas as regiões são utilizados os mesmos pares de *primers* tendo como resultado um conjunto de dois alelos e sua análise é realizada em um único bloco (NIEDERSTÄTTER, 2004).

Figura 4: Distribuição dos 23 Y-STR presentes no kit comercial Powerplex®Y 23.



Fonte: Figura adaptada de www.promega.com.br.

Nota: A região destacada em vermelho indica uma região que pode ser excluída em alguns indivíduos, fazendo com que marcadores presentes nessas regiões estejam ausentes em um perfil Y-STR. Em verde, destaca o marcador que nesses casos é utilizado como indicador de sexo.

Espera-se que num futuro próximo os laboratórios forenses sejam capazes de usar conjuntos de Y-STRs altamente discriminativos, principalmente por serem descritos cada vez mais marcadores considerados *Rapidly Mutating* (RM) (BALLANTYNE, *et al*, 2012), que apresentam alta taxa mutacional, aumentando cada vez mais o potencial para distinguir entre parentes masculinos que pertencem à mesma linhagem paterna.

Os marcadores Y-STRs, por apresentarem herança paterna exclusiva, podem ser utilizados também em outros tipos de estudo, como na determinação do sexo, em estudos de genética de populações e de evolução humana, bem como nas análises de fluxo migratório diferencial (PENA *et al*., 2000; TARAZONA-SANTOS *et al*., 2001; MARTÍNEZ *et al*., 2007; WATKINS *et al*., 2008). No entanto, devido à existência de variações nas frequências alélicas entre diferentes grupos étnicos e populacionais, é importante determinar as frequências dos marcadores Y-STRs nas diferentes populações (PARRA *et al*., 2003; CORACH *et al*., 2010; BRENNER, 2013).

Em casos forenses, sua principal aplicação está relacionada a casos de violência sexual, quando existe mistura de material celular da vítima e do suspeito, em quantidades desproporcionais, na grande maioria dos casos com quantidade de material celular feminino consideravelmente superior à do suspeito. Nessas situações, a análise de marcadores AS-STR torna-se complexa, pois o componente majoritário é mais favorável à amplificação, podendo mascarar o perfil genético minoritário ou até suprimir a sua amplificação. Além disso, a presença de mais de um agressor também dificulta a individualização do perfil genético de cada um deles. Nesses casos, a análise de Y-STR é eficaz na identificação do perfil masculino, pois apenas o agressor possui o cromossomo Y, e a presença de mais de um alelo em cada marcador auxilia na identificação da quantidade de agressores envolvido naquele caso (SILVA *et al*, 2006).

Os Y-STRs constituem um haplótipo único cujos *loci* estão em forte desequilíbrio de ligação (*Linkage Disequilibrium*, LD), desse modo não é válida a utilização de cálculos estatísticos baseando-se em frequências alélicas, como para marcadores AS-STR, herdados independentemente. Assim, para o cromossomo Y, torna-se necessário obter a frequência de um perfil haplotípico na população referência da amostra em questão, que representa o número de vezes que um determinado perfil ocorre na população. Para isso, é necessário um amplo banco de dados, representativo da população em questão e com controle de qualidade dos haplótipos depositados. (ROEWER; WILLUWEIT, 2018, KRAWCZAK, 2001).

Um perfil genético do cromossomo Y é coincidente com o perfil da linhagem paterna se existir coincidência entre todos os alelos das amostras que estão sendo comparadas. Nos casos de investigação de parentesco, os marcadores masculinos são analisados em indivíduos relacionados por via paterna, como por exemplo, num caso em que o pai esteja desaparecido. Se existir uma coincidência, entre o perfil masculino dos intervenientes sugere-se existir então um laço de parentesco entre eles pela via paterna. Se não existir essa coincidência, será necessário avaliar quais os marcadores em que foi detectada alguma incompatibilidade, considerando a possibilidade de a mesma ser devida à mutação genética. Vale salientar nestes casos que a distância entre gerações de indivíduos é um fator que contribui para aumentar a

probabilidade de ocorrência de diferenças alélicas entre os indivíduos. Deste modo, essas incompatibilidades podem ser resultados de mutações, e por isso não devem ser avaliadas necessariamente como exclusões de laços de parentesco entre os indivíduos. Assim, as taxas de mutação são particularmente importantes e devem ser obtidas para utilização na rotina dos casos forenses.

2.2.3. Mutação no cromossomo Y

Uma mutação é considerada se existirem incompatibilidades em poucos *loci* envolvendo uma ou duas unidades de repetição na transmissão hereditária das características genéticas. Nestes casos a coincidência de perfis genéticos de dois indivíduos alegadamente pertencentes à mesma linhagem paterna deve ser reconsiderada, dada a informação existente relativa às taxas de mutação (KAYSER, 2000). As mutações em AS-STR ou Y-STR são o resultado de erros no DNA durante a fase da replicação, que normalmente implicam no ganho ou perda de uma unidade de repetição, sendo assim transmitidos à geração seguinte.

Em testes de paternidade, a observação de três incompatibilidades pode ser considerada como exclusão de parentesco (GJERTSON *et al*., 2007), porém com utilização de Kits comerciais contendo maior número de marcadores, a observação destas diferenças pode ser maior.

Para a aplicação de marcadores Y-STRs em paternidade e teste de identidade incluindo parentes do sexo masculino, é necessário levar-se em conta as taxas de mutação. Com base em uma taxa média de mutação de $2{,}8 \times 10^{-3}$, na comparação de haplótipos incluindo nove Y-STRs, por exemplo, que é o haplótipo mínimo, pode ocorrer pelo menos um alelo incompatível entre pai e filho, em cerca de 1 em 40 pares analisados. Este valor aumenta para 1 em cada 20 pares pai/filho quando 18-19 STRs são analisados. Uma grande quantidade de dados é necessária para estimar as taxas de mutação, o que pode ser crucial para a interpretação dos resultados genéticos em determinadas situações. Portanto, além dos esforços que estão sendo feitos em publicações de dados populacionais, a publicação de dados de mutação de

pares pai/filho com a paternidade confirmada deve ser incentivada (PINTO, GUSMÃO, AMORIM *et al*., 2014).

Em genética populacional, a estimativa aproximada das taxas de mutação dos marcadores de cromossomo Y é particularmente importante para a datação da origem das linhagens paternas, com importância em estudos antropológicos. Em termos forenses, o conhecimento das taxas de mutação é necessário para a interpretação mais correta dos dados relativos aos Y-STR, sobretudo nos testes de investigação de parentesco (KAYSER, 2000, VIEIRA-SILVA *et al*, 2009).

2.3 Bases de dados

Quando identificado um haplótipo, seja para uso em área criminal, relações de parentesco, ou estudos antropológicos, é necessário estimar a frequência com que este haplótipo ocorre na população de referência. (VELDMAN, 2007). Esta frequência é calculada usando bancos de dados populacionais forenses, sendo um pré-requisito para a quantificação do valor probatório da correspondência entre os perfis Y-STRs (ROEWER, 2018).

Em nível internacional existe um repositório relativo a marcadores do cromossomo Y (Y-STR ou Y-SNPs), aplicados na área forense, o YHRD (*Y-Chromosome STR Haplotype Reference Database*), criado por Lutz Roewer juntamente com Sascha Willuweit, no ano de 2000 (https://.yhrd.org); atualmente, dispõem de 284.406 haplótipos referentes ao conjunto de haplótipo mínimo, sendo que 11.398 são referentes à população brasileira. Este repositório reúne informação de todos os laboratórios associados que têm contribuído com os seus dados, inclusive com informação relativa às taxas de mutação. No entanto, tratam-se de taxas de mutações conjuntas, e a aplicação de uma mesma frequência haplotípica ou de taxa de mutação de um determinado marcador, para diferentes regiões, só é válida se não houver sub-estruturação da população, ou seja, diferenças estatisticamente significativas entre as distribuições de haplótipos Y-STR nas populações. Essa informação pode ser obtida de laboratório recorrendo a estudos de pares pai/filho com paternidade confirmada (KAYSER *et al*., 2000).

No caso das populações brasileiras, as análises moleculares vêm demonstrando que a maior parte da variabilidade genética ocorre dentro das populações, tendo menor influência as diferenças inter-populacionais (BARCELOS *et al*., 2006; GODINHO *et al*., 2008, PALHA *et al*., 2012, RESQUE *et al*., 2016). Esta diversidade genética existente em indivíduos de um mesmo grupo e entre diferentes grupos populacionais, pode ser avaliada a partir da distribuição das frequências alélicas de marcadores moleculares. As variações observadas nestas frequências permitem estimar o grau de divergência entre populações, cuja variabilidade pode ser decorrente de diversos fatores, destacando-se a migração, como um dos mais impactantes (DALTON, 2010). Outro importante fator neste cenário consiste nas variações possivelmente decorrentes das taxas de mutações que favorecem pequenas mudanças nas frequências alélicas de uma população (GÓES *et al*., 2005; SOARES VIEIRA *et al*., 2009; DE ANDRADE *et al*., 2009).

Com a publicação da Lei 12.654/12 (BRASIL, 2012), o Brasil foi um dos países que mais recentemente aderiu ao projeto de criação de um banco de perfis genéticos, porém, utilizado apenas para fins criminais. De acordo com o Ministério da Justiça, atualmente, o Banco Nacional de Perfis Genéticos brasileiro (BNPG) contém aproximadamente 17.361 perfis genéticos de condenados, 545 de investigados e 9.111 de vestígios de local de crime. No Brasil, até o momento, 559 investigações foram auxiliadas por essa ferramenta (https://www.justica.gov.br/sua-seguranca/seguranca-publica/ribpg/relatorio, acessado em agosto de 2019).

Assim, para melhor aproveitamento, estudos populacionais referentes à população brasileira, que permita um incremento em bancos de dados genéticos, se tornam ainda mais importantes, pois estes dados podem ser aplicados para obtenção das frequências dos perfis depositados no BNPG em nossa população, cooperando ainda mais com a Justiça brasileira.

2.4 Aspectos históricos do povoamento do Brasil

Do ponto de vista genético, a população brasileira é considerada uma das mais heterogêneas do mundo (PALHA, *et al*., 2012), com importante

contribuição de três grandes grupos populacionais: Europeus, Africanos e Nativo Americanos.

Os portugueses foram os primeiros estrangeiros a chegar ao país, em meados de 1500, na região Nordeste. Nesta época, houve uma grande diminuição da população indígena da região, devido à resistência aos portugueses, além da ocorrência de epidemias que dizimaram aldeias, as quais os nativos não tinham anticorpos naturais. Na região Norte, a localização das aldeias em áreas mais afastadas dos núcleos coloniais, dificultou o apresamento de indígenas e o alastramento de doenças, contribuindo para que o “despovoamento” ali fosse menos trágico do que fora no litoral (IBGE).

No final do século XVI, iniciaram-se as viagens de reconhecimento do interior do país, conhecidas como “entradas” ou “bandeiras” dos sertanistas, em busca de metais e indígenas escravos (BUENO, 2009). Ao longo do século XVI e XVII, as atividades econômicas dos colonos da região eram realizadas basicamente por escravos indígenas, que foram aprisionados ao longo das expedições pelo sertão. No final do século XVIII, a economia açucareira passa a ter grande importância, obrigando os colonos a intensificar a importação de escravos africanos (MONTEIRO, 1994). Os portugueses foram os responsáveis por trazerem em torno de 3,6 milhões de africanos de forma compulsória, sendo o país do continente americano que mais importou escravos africanos. (RIBEIRO, 1995).

A partir de 1882, as plantações de cana de açúcar foram substituídas pelas plantações de café, que se mostraram mais promissoras; com isso iniciou-se a imigração de europeus com a finalidade de prover a mão-de-obra nas lavouras de café. Assim, iniciaram-se construções de estradas de ferro, expandindo a ligação do interior com o litoral, o que acelerou significativamente a imigração em massa de italianos, espanhóis e mais tarde japoneses, que se dirigiram principalmente para as regiões Sudeste e Sul do país. O número de indivíduos que chegaram foi de aproximadamente 6 milhões.

O início do período republicano (1889) e o surto de industrialização trouxeram a modernização aos estados de São Paulo e Rio de Janeiro, sendo marcada pela elevação da concentração populacional. A imigração de estrangeiros foi um grande contribuinte para esse aumento, especialmente de italianos, sírios, judeus e libaneses, com dados indicando que, entre 1908 e

1920, chegaram em Santos 190 mil imigrantes subvencionados e 340 mil espontâneos (MOTA, 2003).

Ao mesmo tempo em que pessoas de diversos países e continentes estavam chegando a diferentes regiões do Brasil, movimentos migratórios importantes estavam ocorrendo dentro do território, principalmente devido a interesses econômicos.

Com isso, a história de povoamento do país se mostrou heterogênea nas diferentes regiões geopolíticas do país (Norte, Nordeste, Centro-Oeste, Sudeste e Sul), e a forma de distribuição desses grupos populacionais, variou muito conforme a região, o que é demonstrado em diversos estudos (FIGUEIREDO, *et al*., 2015, OLIVEIRA *et al*., 2014, RESQUE *et al*., 2016), fato este que se dá pelo grande tamanho territorial (mais de 8,5 milhões de km^2). Como resultado, a região Norte abriga indivíduos com maior ascendência dos nativo-americanos; a região Nordeste apresenta maior contribuição africana; o Sudeste e o Sul são as regiões onde a contribuição europeia é mais predominante e a região Centro-Oeste recebeu indivíduos provenientes de todas as outras regiões brasileiras, principalmente do Nordeste e Sudeste, sendo a última região colonizada do Brasil (PALHA *et al*., 2012). Todos esses fatos devem ser refletidos na composição de bases de dados brasileiros, ressalvando a importância de estudos populacionais direcionados para as populações de cada região do país.

2.5 Justificativa

O estudo genético das populações humanas tem interesse não só do ponto de vista forense como do ponto de vista antropológico. Na genética forense, os cálculos para a valorização da prova só podem ser realizados após o estudo das características e frequências genéticas da população à qual pertencem os intervenientes, designada como população de referência.

Até agora, mais de 200 polimorfismos Y-STR têm sido descritos, mas estudos sobre suas taxas de mutação ainda são insuficientes, uma vez que são considerados apenas um número restrito de marcadores (RESQUE *et al*., 2016).

As taxas podem ser estimadas pela aplicação de modelos estatísticos, porém estimativas obtidas a partir da observação direta em casos de investigação de parentesco produzem resultados mais fidedignos (KAYSER *et al*., 2000).

A quantidade de estudos realizados com marcadores autossômicos, apresentam grande contraste com os poucos dados disponíveis de alguns marcadores presentes no painel de marcadores do cromossomo Y (Powerplex®Y 23 - Promega Corporation, USA), como por exemplo o DYS576 e o DYS570, que são considerados RM, por apresentarem taxa de mutação na ordem de grandeza de 10^{-2} (BALLANTYNE *et al*, 2012); tais marcadores, por apresentarem esta característica, estão presentes também no kit comercial de AS-STR PowerPlex® Fusion 6C System (Powerplex®Y 23 - Promega Corporation, USA), utilizados por inúmeros laboratórios de genética forense. Por essa razão, estudos referentes a esses marcadores para diferentes populações se fazem necessários.

Por meio da análise de 23 Y-STR, este trabalho teve como objetivo principal verificar dados mutacionais a partir da transmissão de pais para filhos nas populações brasileiras dos estados de São Paulo e Rio de Janeiro, bem como caracterizar geneticamente estas populações, avaliando a relação entre elas comparando com diferentes regiões do Brasil e do mundo.

Os dados obtidos também poderão ser aplicados na rotina do laboratório de Investigação de Paternidade da Faculdade de Ciências Farmacêuticas – UNESP – Araraquara – SP, principalmente para solucionar casos de investigações de paternidade envolvendo indivíduos do sexo masculino na ausência do suposto pai, sendo necessária uma reconstrução genética. Além de que estudos de taxas de mutação Y-STR para alguns marcadores ainda são escassos na literatura.

3. OBJETIVOS

Avaliar a ocorrência de mutações em 23 marcadores STRs do cromossomo Y, em casos de inclusão de paternidade, bem como caracterizar as populações dos estados de São Paulo e Rio de Janeiro, pela análise desses marcadores para utilização em identificação humana.

3.1 Objetivos específicos:

- Avaliar 23 Y-STR em 401 casos de investigação de paternidade, com paternidade confirmada, de indivíduos residentes nos estados de São Paulo (n=200) e Rio de Janeiro (n=201) e avaliar os marcadores que apresentaram mutações;

- Comparar os dados mutacionais obtidos entre as populações estudadas e aqueles disponíveis no YHRD;

- Estimar as frequências haplotípicas e alélicas, além da diversidade gênica do conjunto de 23 microssatélites do cromossomo Y (DYS19, DYS385a, DYS385b, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS576, DYS570, DYS448, DYS533, DYS481, DYS549, DYS635, DYS643, DYS458, DYS456 e Y-GATA-H4) das populações dos estados de São Paulo e Rio de Janeiro;

- Avaliar a eficiência forense dos 23 marcadores do cromossomo Y para ambas as populações, e comparar com os dados disponíveis na literatura para outras populações brasileiras;

- Verificar a relação genética das populações estudadas com outras do Brasil e do mundo, por meio da diferenciação genética (*Fst*);

- Identificar os diferentes haplótipos encontrados na população dos estados de São Paulo e Rio de Janeiro e submetê-los ao YHRD auxiliando na caracterização destas populações, ampliando assim a divulgação dos dados obtidos e fornecendo uma ferramenta útil para a investigação forense;

- Classificar as populações de São Paulo e Rio de Janeiro conforme a ancestralidade, por meio da previsão dos haplogrupos obtidos com a ferramenta online Nevgen.

4. MATERIAIS E MÉTODOS

4.1 Aspectos éticos

Este trabalho foi aprovado pelo Comitê de Ética em Pesquisa da Faculdade de Ciências Farmacêuticas da UNESP de Araraquara com o código: CAAE 88408218.0.0000.5426.

4.2 Casuística

Com o objetivo de avaliar ocorrências de mutações no cromossomo Y, foram utilizadas amostras sanguíneas de indivíduos oriundos dos estados de São Paulo e Rio de Janeiro.

As amostras da população do estado de São Paulo foram coletadas no Laboratório de Investigação de Paternidade da Faculdade de Ciências Farmacêuticas – UNESP – Araraquara – SP, de indivíduos que recorreram ao laboratório para realização de exame de parentesco. As amostras coletadas envolveram duos de pais e filhos, e casos de reconstruções genéticas envolvendo indivíduos do sexo masculino. Os casos selecionados tiveram as relações de paternidade confirmadas previamente por painéis de STR autossômicos, em um princípio de razão de verossimilhança que geraram valor de índice de paternidade (IP) ≥ 10.000.

Para a população do estado de São Paulo foram analisados 200 casos, com um total de 409 amostras. Os casos estão descritos na Tabela 1.

Tabela 1: Descrição dos casos com relações de parentesco confirmadas utilizados nas análises de mutações da população do Estado de São Paulo.

Amostras da população de São Paulo				
	Casos analisados	**Reconstrução Genética**	**Descrição dos casos**	**Total de Amostras**
	186	não	1P+1F	372
	5	não	1P + 2F	15
	3	sim	1IP +1F	6
	3	sim	2IP+ 1F	9
	2	sim	1FB+1F	4
	1	sim	2FB+1F	3
Total	200			409

P, pai; F, filho; ISP, irmão do pai; FB, filho biológico.

Fonte: Este trabalho.

As amostras da população do estado do Rio de Janeiro foram cedidas gentilmente pelo Prof. Elizeu F. Carvalho, sendo provenientes de indivíduos que recorreram ao Laboratório de Diagnóstico por DNA da Universidade do Estado do Rio de Janeiro (UERJ) para realização de exames de parentesco. Foram analisados 201 casos, com um total de 407 amostras. Os casos estão descritos na Tabela 2.

Tabela 2: Descrição dos casos com relações de parentesco confirmadas utilizados nas análises de mutações da população do estado do Rio de Janeiro.

Amostras da população do Rio de Janeiro				
	Casos analisados	**Reconstrução genética**	**Descrição dos casos**	**Total de amostras**
	198	não	1P+1F	396
	2	não	1P + 2F	6
	1	não	1P + 4F	5
Total	201			407

P, pai; F, filho.

Fonte: Este trabalho.

Para todos os casos que envolveram pais com mais de um filho foram realizadas análises independentes para cada filho, pois em todos os casos os filhos apresentaram idades diferentes e, portanto concebidos em diferentes gestações.

4.3 Coletas das amostras biológicas e extração do DNA

As amostras de sangue foram coletadas em papel FTA® (Whatman® GE) utilizando caneta lancetadora (G.Tech), e devidamente armazenados até momento da extração do DNA, que foi realizada conforme adaptação das metodologias propostas por Singer-Sam, Tanguay e Riggs (1989), com auxílio da resina Chelex 100 (Bio-Rad Laboratories).

Para a extração do DNA, 1 disco de 3mm de papel de filtro FTA® (Whatman® GE), contendo a amostra sanguínea do indivíduo, foi cortado com o auxílio da caneta Harris Micro Punch® (Whatman® GE) e depositado em tubo de 0,2ml. Os discos foram agitados por 5 segundos em 50µL de água ultrapura (Milli-Q ® - Milipore Corporation). O sobrenadante foi descartado e os discos incubados com 50µL de água ultrapura (Milli-Q ® - Milipore Corporation) por 30 minutos em temperatura ambiente. O sobrenadante foi descartado e os discos foram incubados em 100µL de Chelex 100 (Bio-Rad Laboratories) na concentração de 5% por pelo menos 1 hora à 56ºC. Após esse período, a solução foi agitada e incubada por 8 minutos à 100ºC. A solução foi centrifugada a 13.000 rpm por 3 minutos e transferiu-se o sobrenadante para novo microtubo, o qual foi armazenado à 4ºC.

As amostras da população do Rio de Janeiro nos foram cedidas como DNA já extraído, sendo a extração realizada previamente no Laboratório de Diagnóstico de DNA- EURJ, com auxílio da resina Chelex 100 ou por extração com Fenol-Clorofórmio.

4.4 Análises dos 23Y-STRs

4.4.1 Reação de PCR

Neste trabalho foi utilizado o Kit comercial Powerplex®Y 23 (Promega Corporation, USA; PPY23 – 23 marcadores) para amplificação dos marcadores (Tabela 4).

A reação de amplificação por PCR para todas as amostras de DNA foi realizada em um sistema único em *multiplex*. Neste trabalho, as reações foram preparadas com adaptações que reduzem o volume final da reação para 5µL por amostra, porém mantendo a proporção estabelecida pelo fabricante (Tabela 3).

Tabela 3: Comparação da utilização dos reagentes na reação de PCR seguindo recomendações do fabricante e protocolo adaptado.

	Recomendações do fabricante	**Protocolo adaptado**
Water, Amplification Grade (µl)	q.s.p. 25	2,7
PowerPlex® Y23 5X Master Mix (µl)	5	1
PowerPlex® Y23 10X Primer Pair Mix (µl)	2,5	0,5
DNA extraído (ng)	~0,5ng/µl (DNA comercial)	~0,2-4ng/µl
Volume Final (µl)	25	5

Fonte: Este trabalho.

Os fragmentos foram amplificados em termociclador Veriti® Thermal Cycler (Applied Biosystems® – Thermo Fischer Scientific), nas seguintes condições:

96°C - 2 min
94°C - 10 seg
61°C - 1 min — 28 ciclos
72°C - 30 seg
60°C - 20 min

Tabela 4: Marcadores Y-STRs analisados com o kit Powerplex®Y 23 e sequência repetitiva do polimorfismo.

STR	**Sequência Repetitiva 5´-> 3´**
DYS576	AAAG
DYS389I/II	(TCTG) (TCTA)
DYS448	AGAGAT
DYS19	TAGA
DYS391	TCTA
DYS481	CTT
DYS549	GATA
DYS533	ATCT
DYS438	TTTTC
DYS437	TCTA
DYS570	TTTC
DYS635	TSTA composto
DYS390	(TCTA) (TCTG)
DYS439	AGAT
DYS392	TAT
DYS643	CTTTT
DYS393	AGAT
DYS458	GAAA
DYS385a/b	GAAA
DYS456	AGAT
Y-GATA-H4	TAGA

Fonte: Powerplex®Y 23 (Promega Corporation, USA) manual.

4.4.2 Análise dos produtos amplificados

A separação e detecção dos produtos amplificados foram realizadas por eletroforese capilar, no analisador genético ABI 3500 (Applied Biosystems® – Thermo Fischer Scientific), utilizando o polímero POP-4 (Applied Biosystems® – Thermo Fischer Scientific) e capilar de 36 cm (Applied Biosystems® – Thermo Fischer Scientific). Cada eletroforese foi acompanhada com uma amostra de um *ladder*. As amostras foram preparadas em placa de 96 poços conforme recomendações do fabricante, sendo da seguinte forma:

- 1 μL do produto de PCR
- 9,5 μL de formamida HI-DI (Applied Biosystems® – Thermo Fischer Scientific)
- 0,5 μL de WEN ILS500 (Promega Corporation, USA)

Os resultados da eletroforese capilar, foram analisados em eletroferogramas no software Gene Mapper v ID-X-1.2 (Applied Biosystems® – Thermo Fischer Scientific). A tipagem ou classificação dos alelos foi efetuada por comparação com o Allelic Ladder fornecido com o kit e a sua nomenclatura é feita de acordo com as recomendações da DNA Comission of the International Society of Forensic Genetics (ISFG) (GUSMÃO *et al*, 2006)

4.5. Análises estatísticas

4.5.1 Avaliação de mutações

As mutações foram identificadas por observação direta dos Y-STRs obtidos dos casos analisados, por meio da detecção de diferenças alélicas entre os indivíduos referentes a cada caso, conforme recomendado por outros autores (KAYSER, 2000).

Na generalidade dos casos, a mutação genética é caracterizada pela diferença do número de unidades de repetição entre o perfil do pai e do filho, num único marcador, devido à deleção ou inserção de uma ou mais unidades de repetição. O reconhecimento de uma mutação é admitido sempre que existe

uma incompatibilidade entre pai e o filho. Nas amostras em que isso ocorreu, a análise foi repetida desde início do processo de extração de DNA para confirmação dos resultados.

As taxas de mutações por marcador foram calculadas dividindo o número de mutações encontradas para um determinado locus pelo número de transferências alélicas, e a taxa geral de mutação foi calculada dividindo o número total de mutações observadas pelo número total de transferências alélicas. Os números de transferências alélicas referente a cada marcador analisado estão disponíveis na tabela 5. Intervalos de confiança (IC) para taxas de mutação (global e por marcador) foram estimados, assumindo um intervalo de confiança de 95%.

Tabela 5: Informação relativa ao número de transferências alélicas analisadas por marcador referente a cada população estudada.

	Rio de Janeiro	**São Paulo**	**Total**
DYS576	206	223	429
DYS389 I	206	225	431
DYS448	207	224	431
DYS389 II	206	224	430
DYS19	206	228	434
DYS391	206	224	430
DYS481	206	224	430
DYS549	206	224	430
DYS533	206	224	430
DYS438	206	224	430
DYS437	206	224	430
DYS570	206	224	430
DYS635	206	224	430
DYS390	206	224	430
DYS439	209	224	433
DYS392	205	224	429
DYS643	206	224	430
DYS393	206	224	430

DYS458	206	224	430
DYS385	412	448	860
DYS456	206	224	430
YGATAH4	206	224	430
Total	**4741**	**5156**	**9897**

Nota: O número de meioses por marcador variou devido a deleções, duplicações e triplicações de alelos.

Fonte: Este trabalho.

4.5.2 Diversidade genética e de relevância forense

Para os cálculos estatísticos dos parâmetros forenses, foi selecionado apenas um haplótipo correspondente a cada caso analisado, já que para tais cálculos não deve haver parentesco entre os indivíduos analisados. O número total de haplótipos utilizados foi de 401.

A diversidade haplotípica, definida como a probabilidade de encontrar dois perfis genéticos masculinos diferentes na população (EXCOFFIER, 2005), bem como desvio padrão, foi estimada de acordo com Nei, utilizando a fórmula: $HD = (n/(n-1)) (1-\Sigma pi^2)$, sendo n o número de amostras e pi a frequência haplotípica.

Com a utilização do software Arlequin v 3.5.1.2, foram realizadas as seguintes abordagens: a) frequências alélicas e diversidade gênica de cada marcador analisado b) nº de coincidência entre os perfis genéticos; c) distâncias genéticas (*Fst*) entre pares de populações (população estudada e outras descritas na literatura – tabela 6). Os valores de P foram obtidos após 50.000 permutações. A correção para múltiplos testes proposta por Bonferroni foi utilizada (BONFERRONI, 1936) para reduzir as chances de se obter um resultado falso-positivo quando muitas comparações são realizadas em um único conjunto de dados. Esse método consiste na divisão do nível de significância estipulado (em nosso caso, 0,05) pelo número de testes que foram executados.

Utilizou-se o software STATISTICA ver.8.0 (StatSoft) para a obtenção da escala multidimensional (MDS) plot baseada nas distâncias genéticas obtidas pelo cálculo de *Fst*, que foi plotado em um gráfico de 2 dimensões.

A probabilidade de coincidência entre os haplótipos foi calculada como a soma do quadrado das frequências haplotípicas (PURPS *et al.,* 2014). A Capacidade de discriminação (CD) do conjunto de marcadores utilizado para cada população foi feita utilizando a fórmula: CD = Nº de haplótipos diferentes/Nº total de indivíduos.

Para todos os parâmetros estatísticos o número de repetições em DYS389II foi subtraído de DYS389I. Alelos nulos, intermediários e duplicado não foram desconsiderados. DYS385a/b foi considerado como um bloco, a ordem dos alelos foi referente ao tamanho dos fragmentos (menor para o maior) conforme indicados nos eletroferogramas.

Para avaliar a relação genética das populações estudadas com outras do Brasil e do mundo, foram utilizadas populações referências que se mostraram com grande influência na história de povoamento do Brasil, além da proximidade geográfica (Tabela 6).

Tabela 6: Amostras utilizadas para comparação entre as populações, número de haplótipos analisados (N), origem ancestral da população e respectivas referências.

População	N	Origem	Referência
Brasil			
São Paulo 1	200	Mistura	Este trabalho
Rio de Janeiro 1	201	Mistura	Este trabalho
São Paulo 2	120	Mistura	PURPS *et al.* (2014)
Rio de Janeiro 2	123	Mistura	PURPS *et al*. (2014)
Manaus	204	Mistura	Dados colaborativos
Brasília	410	Mistura	FERREIRA *et al*. (2015)
São Gabriel da Cachoeira	61	Nativo Americana	PURPS *et al*. (2014)
América do Sul			
Bolívia	56	Nativo Americana	PURPS *et al*. (2014)
Bolívia	44	Mistura	PURPS *et al* (2014)
Argentina	257	Mistura	PURPS *et al*. (2014)
Peru	83	Mistura	PURPS *et al*. (2014)
Equador	270	Mistura	TOSCANINI *et al*. (2018)
Paraguai	537	Mistura	RIBEIRO *et al*. (2018)
América Central			
Panamá	100	Mistura	PURPS *et al*. (2014)
Costa Rica	166	Mistura	PURPS *et al*. (2014)
Europa			
Espanha	706	Europeia	PURPS *et al*. (2014)
Portugal	248	Europeia	PURPS *et al*. (2014)
Itália	550	Europeia	PURPS *et al*. (2014)

Alemanha	634	Europeia	PURPS *et al*. (2014)
África			
Zimbábue	55	Africana	PURPS *et al*. (2014)
Populações Norte Americanas			
Africanos	509	Africana	PURPS *et al*. (2014)
Europeus	534	Europeia	PURPS *et al*. (2014)
Asiáticos	96	Asiática	PURPS *et al*.(2014)
Hispânicos	403	Europeia	PURPS *et al.* (2014)

Fonte: Este trabalho.

Apesar da forma mais precisa em classificar uma amostra em sua respectiva origem ancestral seja através de haplogrupos, pela tipagem de SNPs, foi realizada uma previsão dos respectivos haplogrupos de cada haplótipos obtido neste trabalho, utilizando a ferramenta online Y-DNA Haplogroup Predictor - Nevgen (http://nevgen.org), sendo classificadas conforme sua origem ancestral (ATHEY, 2005, ATHEY, 2006).

Todos haplótipos obtidos neste trabalho foram submetidos ao banco de dados YHRD (www.yhrd.org) sob número de acesso: YA004619 (Rio de Janeiro) e YA004620 (São Paulo).

5. RESULTADOS E DISCUSSÃO

5.1 Padronização da utilização do Kit Powerplex®Y 23

Com a atual economia que estamos vivenciando e a dificuldade de levantar recursos para desenvolver pesquisas, inicialmente foi realizada uma padronização do uso do kit comercial utilizado neste projeto para a análise dos 23 Y-STR. Conforme recomendação do fabricante é necessária à utilização de um volume final de 25µl do *mix* utilizado na reação de PCR. Tendo em vista a importância da otimização de custos dos métodos realizados na rotina dos laboratórios, sem afetar sua sensibilidade, especificidade, fácil execução e interpretação dos dados, testou-se a redução do volume final da reação para 5µL, mantendo a proporção da concentração recomendada pelo fabricante dos componentes do Kit (*Master Mix e Pair Primers*), ou seja uma redução de 1/5 no volume final do Kit comercial.

No laboratório, para os exames de investigação de paternidade já se utilizava uma adaptação do protocolo original do Kit comercial, com uma redução para 1/2 do volume final do *mix*, porém não havia padronização do mesmo kit com redução do volume para 1/5.

Além da economia feita na utilização do Kit, o volume de DNA utilizado também foi menor, uma vez que a reação padronizada no laboratório utilizava entre 1 e 1,5µl (+/- 0,5-5ng/µl) de DNA extraído com o mesmo protocolo de extração de DNA; a partir desta redução, passou-se a utilizar um volume de 0,8µl (0,2-4ng/µl) (Tabela 3).

Para a comparação dos resultados, foram realizadas concomitantemente reações seguindo o protocolo previamente estabelecido pelo laboratório e o adaptado, utilizando, em ambas, amostras do DNA comercial 2800 (Promega Corporation, USA) como controle positivo.

Observando os eletroferogramas do controle positivo (Figura 5) seguindo o protocolo estabelecido no laboratório e o com as adaptações realizadas neste trabalho (Figura 6), foi possível notar que, apesar da redução de RFU (unidade de fluorescência relativa), houve adequada amplificação de todos os marcadores nas amostras com volume reduzido.

Figura 5: Eletroferograma do perfil alélico do DNA comercial obtido com o Powerplex®Y 23, seguindo protocolo estabelecido no laboratório. RFU, unidade de fluorescência relativa. **Fonte:** Este trabalho.

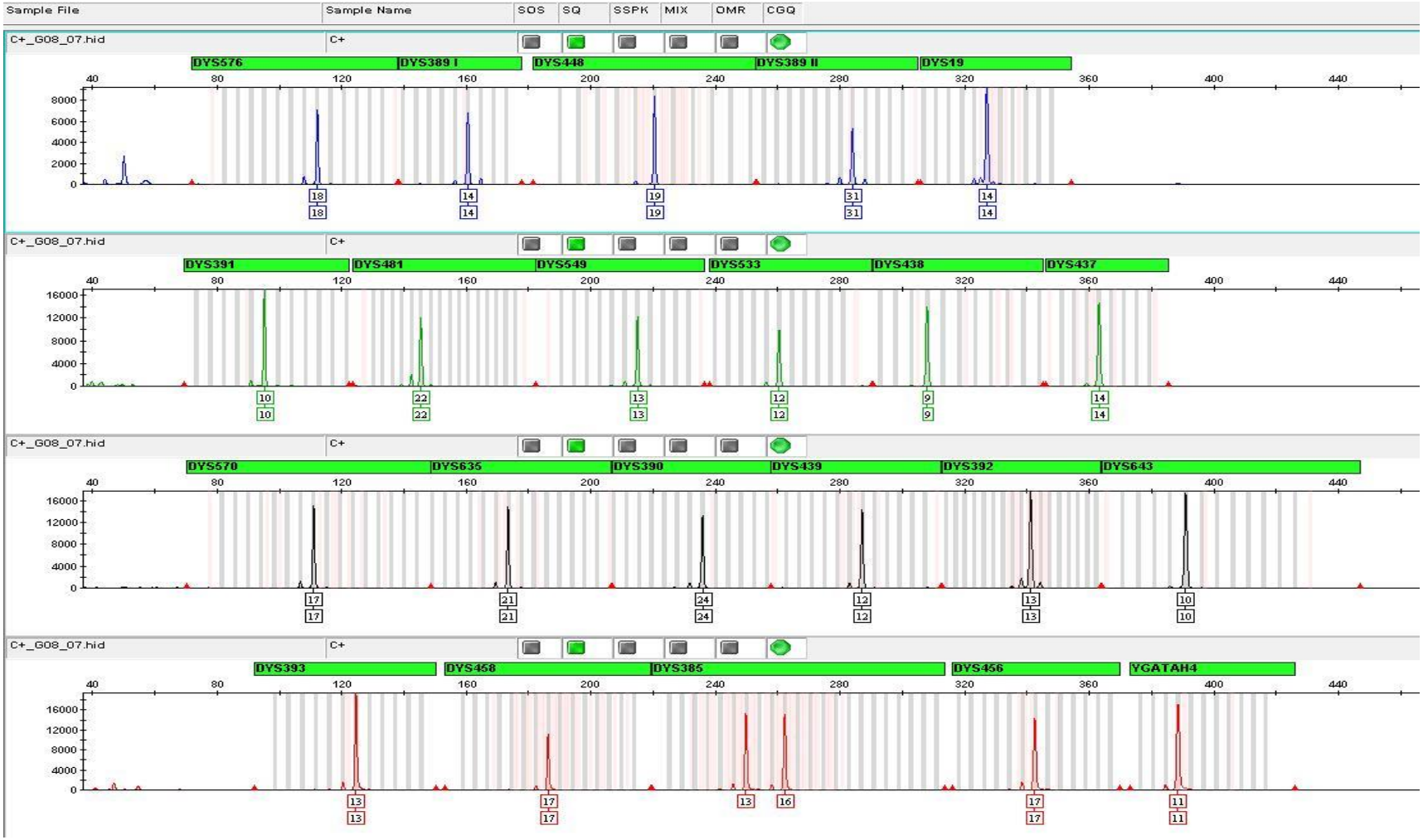
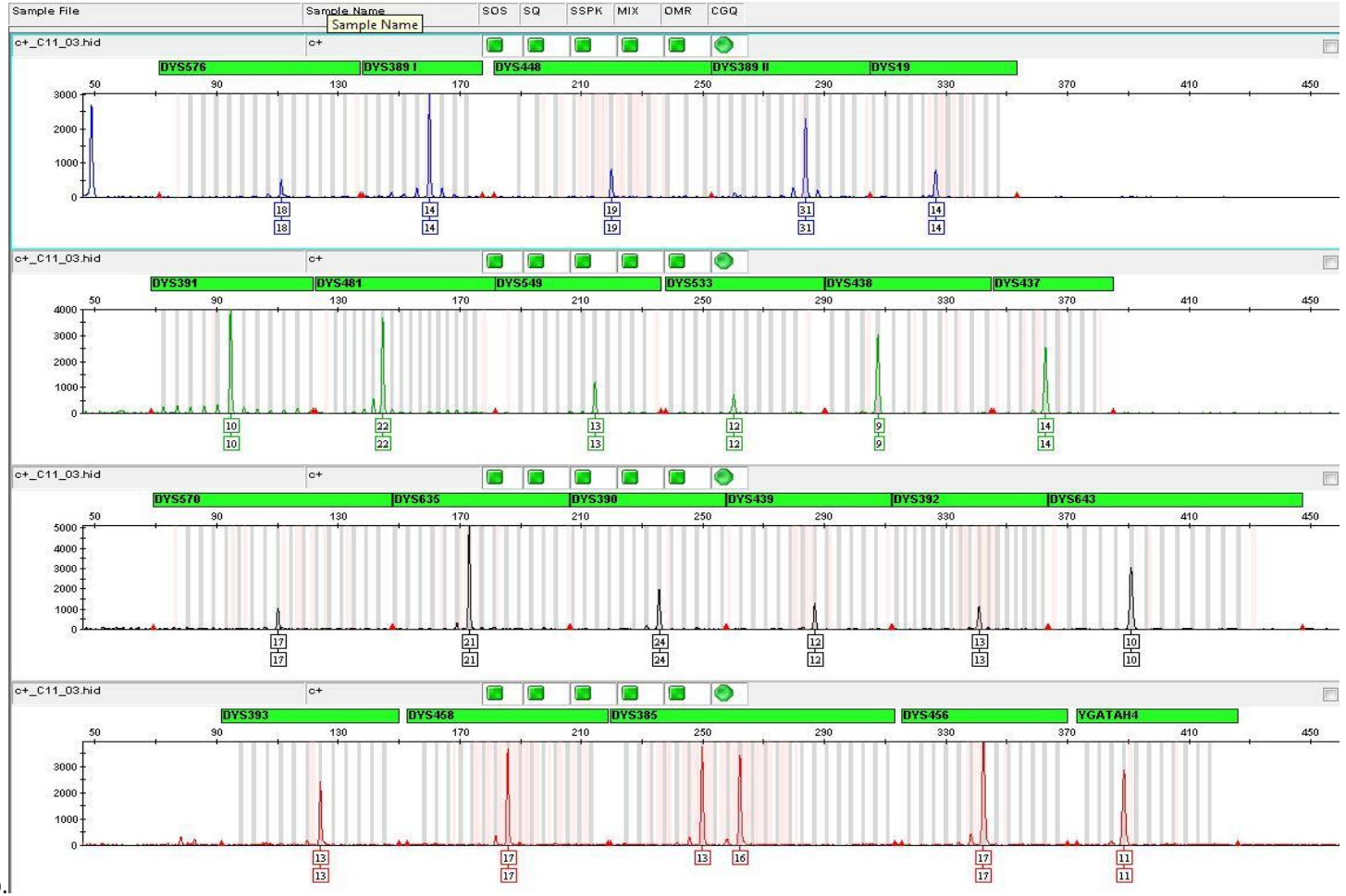


Figura 6: Eletroferograma do perfil alélico do DNA comercial 2800 obtido com o Powerplex®Y 23 seguindo protocolo adaptado. RFU, unidade de fluorescência relativa. **Fonte:** Este trabalho.



5.2 Análises de mutações

Em 401 casos com paternidade confirmada, totalizando 9.897 transferências alélicas, foram identificadas 38 mutações (Tabela 7). As mutações foram observadas em 15 dos 23 loci avaliados, sendo que em 5 dos 15 marcadores mutados foi observada apenas uma mutação.

A maioria das mutações foi observada na população de São Paulo, em um total de 24 (Tabela 7), resultando em uma taxa de mutação geral de $4{,}65x10^{-3}$ (95% IC, $2{,}98x10^{-03}$ a $6{,}92x10^{-03}$). Enquanto que na população do Rio de Janeiro foram observadas 14, das 38 mutações encontradas, resultando em uma taxa geral de $2{,}95x10^{-03}$ (95% IC, $1{,}62x10^{-03}$ a $4{,}95x10^{-03}$).

Tabela 7: Número de mutações observadas para as populações de São Paulo e Rio de Janeiro separadamente.

	Nº meioses	-1	+1	Total	Transferências alélicas	Taxa de mutação	Limite inferior (95%)	Limite superior (95%)
Rio de Janeiro	206	5	9	14	4741	$2{,}95X10^{-3}$	$1{,}62X10^{-3}$	$4{,}95X10^{-3}$
São Paulo	224	17	7	24	5156	$4{,}65X10^{-3}$	$2{,}98X10^{-3}$	$6{,}92X10^{-3}$
Total	430	22	16	38	9897	$3{,}84X10^{-3}$	$2{,}72X10^{-3}$	$5{,}27X10^{-3}$

Fonte: Este trabalho.

Para os marcadores em que foram observadas mutações, as taxas de mutação obtidas neste trabalho (Tabela 8), não se mostraram diferentes das já relatadas no banco de dados YHRD, já que a taxa de mutação de um marcador pode variar até um fator 10 (^10), diferentemente de taxas de mutação de populações subestruturadas as quais podem apresentar grandes variações (ANTÃO-SOUZA *et al*., 2017). A taxa de mutações de todos os marcadores variou de 0,00232 a 0,01150.

Tabela 8: Dados relativos às mutações observadas nas populações de São Paulo e Rio de Janeiro, respectivos intervalos de confiança (95%), dados disponíveis no YHRD e valores obtidos neste trabalho combinados com os disponíveis no YHRD, referentes a cada marcador Y-STR analisado.

		Este trabalho		IC95%		YHRD		YHRD + Este trabalho	
	Total	**Taxa de mutação**	**Limite Inferior**	**Limite Superior**	**Taxa de mutação**	**Mutações**	**Taxa de mutação**	**Limite Inferior (IC95%)**	**Limite Superior (IC95%)**
DYS576	4	9,32X10-3	2,50X10-3	2,40X10-2	1,26X10-2	37 em 2944	1,22X10-2	8,74X10-3	1,65X10-2
DYS389 I	1	2,32X10-3	5,90X10-5	1,30X10-2	2,72X10-3	41 em 15050	2,71X10-3	1,96X10-3	3,67X10-3
DYS448	1	2,32X10-3	5,90X10-5	1,30X10-2	1,39X10-3	11 em 7940	1,43X10-3	7,41X10-4	2,50X10-3
DYS389 II	3	6,98X10-3	1,40X10-3	2,00X10-2	4,33X10-3	65 em 15021	4,40X10-3	3,42X10-3	5,58X10-3
DYS19	0	0	0	0	2,20X10-3	37 em 16801	2,15X10-3	1,51X10-3	2,96X10-3
DYS391	1	2,33X10-3	5,90X10-5	1,30X10-2	2,53X10-3	41 em 16197	2,53X10-3	1,82X10-3	3,41X10-3
DYS481	3	6,98X10-3	1,40X10-3	2,00X10-2	4,88X10-3	12 em 2460	5,19X10-3	2,91X10-3	8,55X10-3
DYS549	3	6,98X10-3	1,40X10-3	2,00X10-2	3,33X10-3	8 em 2400	3,89X10-3	1,94X10-3	6,94X10-3
DYS533	0	0	0	0	3,68X10-3	9 em 2446	3,13X10-3	1,43X10-3	5,93X10-3
DYS438	0	0	0	0	3,51X10-4	4 em 11384	3,39X10-4	9,23X10-5	8,67X10-4
DYS437	1	2,33X10-3	5,90X10-5	1,30X10-2	1,32X10-3	15 em 11363	1,36X10-3	7,76X10-4	2,20X10-3
DYS570	3	6,98X10-3	1,40X10-3	2,00X10-2	1,14X10-2	30 em 2643	1,07X10-2	7,40X10-3	1,50X10-2
DYS635	1	2,33X10-3	5,90X10-5	1,30X10-2	4,21X10-3	37 em 8787	4,12X10-3	2,92X10-3	5,65X10-3
DYS390	4	9,30X10-3	2,50X10-3	2,40X10-2	2,08X10-3	34 em 16323	2,27X10-3	1,61X10-3	3,11X10-3
DYS439	5	1,15X10-2	3,80X10-3	2,70X10-2	5,46X10-3	62 em 11358	5,68X10-3	4,41X10-3	7,21X10-3
DYS392	0	0	0	0	4,96X10-4	8 em 16129	4,83X10-4	2,09X10-4	9,52X10-4
DYS643	0	0	0	0	1,21X10-3	3 em 2489	1,03X10-3	2,12X10-4	3,00X10-3
DYS393	0	0	0	0	1,07X10-3	16 em 14975	1,04X10-3	5,94X10-4	1,69X10-3
DYS458	3	6,98X10-3	1,40X10-3	2,00X10-2	6,17X10-3	49 em 7939	6,21X10-3	4,64X10-3	8,14X10-3
DYS385	2	2,33X10-3	2,80X10-4	8,40X10-3	2,51X10-3	70 em 27911	2,50X10-3	1,96X10-3	3,15X10-3
DYS456	3	6,98X10-3	1,40X10-3	2,00X10-2	4,41X10-3	35 em 7940	4,54X10-3	3,21X10-3	6,23X10-3
YGATAH4	0	0	0	0	3,01X10-3	27 em 8971	2,87X10-3	1,89X10-3	4,18X10-3
Total	**38**	**3,84X10-3**	**2,72X10-3**	**5,27X10-3**					

Fonte: Este trabalho.

O marcador que apresentou maior número de mutações foi o DYS439, com taxa de 1,15X10^{-2} (95% IC, 3,80x10^{-3} a 2,70x10-^{2}), que apesar de não ser considerado um marcador RM, resultado similar foi encontrado em um trabalho realizado com a população da Coréia (OH e*t a.,* 2015), seguido pelo marcador RM DYS576 com taxa de 9,32x10^{-3} (95% IC 2,50x10^{-3} a 2,40x10^{-2}). Ambos os marcadores apresentaram taxa de mutação acima do IC estimado para a taxa global de mutação de Y-STRs obtida neste trabalho. A variação do IC do marcador DYS390 apresentou-se acima do valor da taxa de mutação disponível no YHRD para este marcador.

A taxa de mutação geral para os 22 marcadores analisados foi de 3,84x10^{-3} (95% IC, 2,72x10^{-3} a 5,27x10^{-3}), mostrando semelhança com trabalhos realizados na população Sul do Brasil (Da FRÉ *et al*., 2015) e da Itália (TURRINA, *et al.,* 2015). Porém, este trabalho mostrou uma grande heterogeneidade dos valores de taxas de mutação para os marcadores analisados, portanto, não sendo correta a aplicação de uma taxa geral para diferentes marcadores Y-STR, mas levar em consideração quais os marcadores estão sendo avaliados e quantas transferências alélicas ocorreram nas análises de parentesco em questão (BALLANTYNE *et al*, 2012).

Com relação aos dados disponíveis no YHRD referentes à mutação por marcador, nota-se que os marcadores utilizados mais recentemente na rotina forense mundial apresentam quantidades de dados bem inferiores quando comparados a marcadores mais antigos (Tabela 8). Como é o caso do marcador DYS19, que foi o primeiro Y-STR identificado para utilização na rotina forense, e que apresenta um valor de taxa de mutação bem estabelecida, apresentando dados de 16801 análises meióticas. Já os dois marcadores RM presentes no PPY23 (DYS576 e DYS570) apresentam quantidade de dados disponíveis bem inferiores, com 2944 e 2643 análises meióticas respectivamente. Entretanto, estes marcadores estão entre os que apresentam maiores taxas mutacionais dentre os disponíveis no YHRD. Além disso, por estarem presentem no Kit comercial pelo potencial poder de discriminação entre linhagens patrilíneas, deve haver disponíveis dados de taxa de mutação mais fidedignos, e para isso, dados contendo maiores números de análises devem ser disponibilizados para estes marcadores.

Os valores das taxas de mutações disponíveis atualmente no YHRD para os marcadores avaliados foram atualizados, juntando-os aos deste trabalho e estão disponíveis na tabela 8.

Todos os dados referentes às mutações observadas nas populações estudadas estão disponíveis na tabela 9.

Tabela 9: Dados de todas as mutações observadas nas populações de São Paulo e Rio de Janeiro.

Marcador	Alelo do pai	Alelo do filho	Tipo de mutação
DYS385	13, 18	14, 18	+1
DYS385	16	15	-1
DYS389 I	13	12	-1
DYS389 II	32	33	+1
DYS389 II	30	29	-1
DYS389 II	31	30	-1
DYS390	23	24	+1
DYS390	24	23	-1
DYS390	24	23	-1
DYS390	21	22	+1
DYS391	16	15	-1
DYS437	15	14	-1
DYS439	11	12	+1
DYS439	13	14	+1
DYS439	12	13	+1
DYS439	11	10	-1
DYS439	14	13	-1
DYS448	21	20	-1
DYS456	15	16	+1
DYS456	16	15	-1
DYS456	17	16	-1
DYS458	20	19	-1
DYS458	16	17	+1
DYS458	21	20	-1
DYS481	24	25	+1
DYS481	28	27	-1
DYS481	23	24	+1
DYS549	13	12	-1
DYS549	12	11	-1
DYS549	12	13	+1
DYS570	17	18	+1
DYS570	18	17	-1
DYS570	22	23	+1
DYS576	18	17	-1

DYS576	17	18	+1
DYS576	19	18	-1
DYS576	17	18	+1
DYS635	23	22	-1

+1, ganho de 1 unidade de repetição; -1, perda de 1 unidade de repetição.

Fonte: Este trabalho.

Todas as mutações observadas neste estudo ocorreram em alelos com sequências idênticas repetitivas ≥11 (Tabela 10), estando em acordo com estudos (BRINKMANN *et al*., 1998; GUSMÃO *et al*., 2005; FAN & CHU 2007; MARTINEZ *et al*., 2017) realizados em STR autossômicos, os quais indicam maior possibilidade de mutação em sequência repetitiva de comprimento maior que 10 repetições homogêneas.

Tabela 10: Tipos de mutação observadas nos alelos presentes na população estudada para o Powerplex®Y 23.

Alelo	**Mutação**	**+1**	**-1**
7-10	0	0	0
11-15	12	6	6
16-20	12	4	8
21-25	10	5	5
26-30	2	0	2
31-33	2	1	1
Total	38	16	22

+1, ganho de 1 unidade de repetição; -1, perda de 1 unidade de repetição.

Fonte: Este trabalho.

Das 38 mutações observadas, 12 ocorreram em alelos de comprimento de 11 a 15 unidades de repetição, e 12 mutações entre os alelos com comprimento de 16 a 20 unidades de repetição. Não foram observadas mutações em alelos menores que 10 unidades de repetição. O estudo de Ballantyne e colaboradores (2010) mostrou que quanto maior o tamanho do fragmento, maior a chance de ocorrer mutação, além de que quanto maior número de alelos, ou seja, quanto mais repetições possíveis dentro do locus, maior a chance de mutações. O número de repetições presentes no alelo mais longo do locus pode ser usado para prever a mutabilidade, ou seja, se este locus apresentar alelos muito longos pode ter maior chance de sofrer mutação que outros com alelos mais curtos. Porém, se outro locus apresentar alelos

mais curtos, mas que a soma de todas as repetições presentes dentro deste locus seja maior, este passa a apresentar maior chance de sofrer mutação.

Em relação ao tipo de mutação, 22 das mutações observadas apresentaram perda de uma unidade de repetição e 16 apresentaram ganhos de uma unidade de repetição (Tabela 10). Não foram encontradas mutações com ganhos ou perdas de 2 ou mais unidade de repetição. Estudos não mostram haver diferenças significativas entre ganhos ou perdas de unidades de repetição (BALLANTYNE *et al.,* 2010), porém mostram que quanto maior o comprimento do alelo mutado, maior a chance da mutação resultar na perda de repetição (encurtamento do alelo:-1), enquanto que os alelos mais curtos tendem a apresentar ganhos de repetições (expansão do alelo:+1) com uma frequência significativamente maior. No gráfico (Figura 7), podemos notar que a linha de frequência relativa de mutações com perdas de unidades de repetição, se mantém próxima da linha da frequência relativa de mutações conforme o comprimento dos alelos vai aumentando, enquanto que a linha referente a mutações de ganhos de unidades de repetição sofre uma queda abrupta conforme aumento do comprimento dos alelos, confirmando dados de outros estudos (GUSMÃO et al., 2005; BALLANTYNE *et al*., 2010).

Foi realizada a correlação baseada nos comprimentos dos alelos encontrados em todos os marcadores (com exceção do DYS385a/b), com a frequências relativas (Tabela 11) dos alelos, das mutações, e tipos de mutações observadas (ganhos:+1 e perdas:-1)

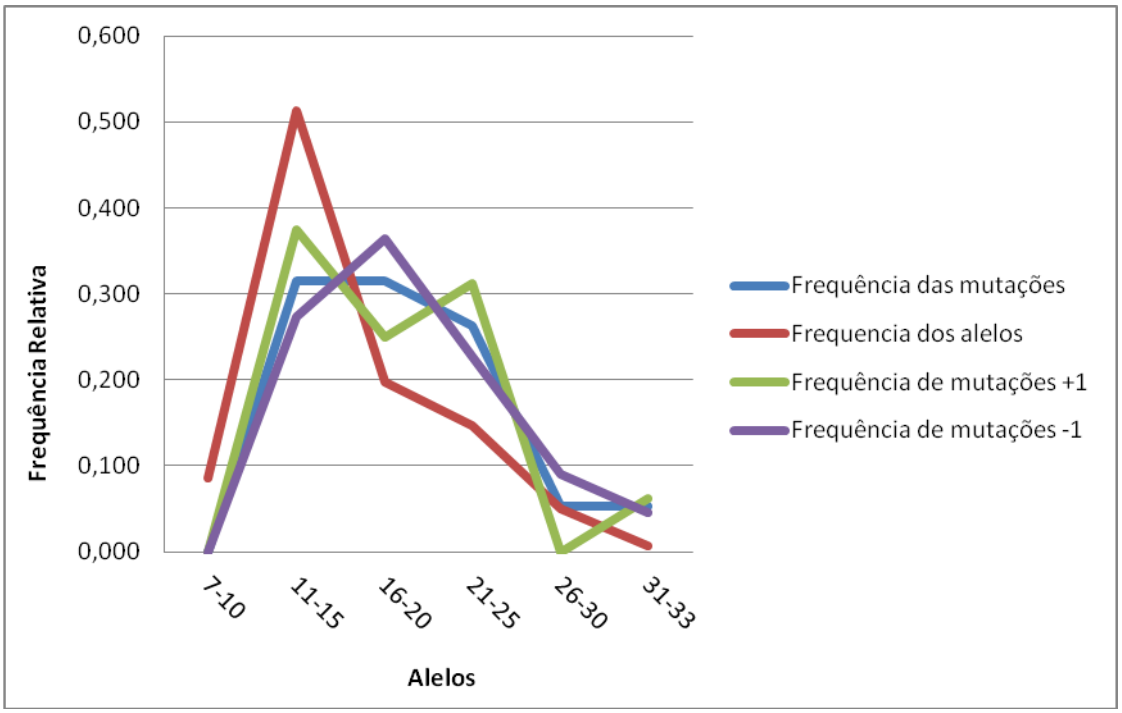
Tabela 11: Frequências relativas das mutações, alelos, e tipos de mutações observadas nas populações de São Paulo e Rio de Janeiro.

Alelo	Mutação	Frequência Alélica	+1	-1
7-10	0	0,0858	0	0
11-15	0,3158	0,5136	0,3750	0,2727
16-20	0,3158	0,1969	0,2500	0,3636
21-25	0,2632	0,1474	0,3125	0,2273
26-30	0,0526	0,0494	0	0,0909
31-33	0,0526	0,0069	0,0625	0,0455

Fonte: Este trabalho.

Os dados da tabela 11 foram representados em forma de gráfico (Figura 7) para permitir a correlação dos dados.

Figura 7: Correlação entre o comprimento dos alelos presentes na população estudada para o Powerplex®Y 23 e as mutações observadas.



Fonte: Este trabalho.

No gráfico (Figura 7), ao compararmos a linha da frequência relativa do comprimento dos alelos, com a de mutações, é possível notar que a linha de mutações se mantém abaixo da frequência relativa de alelos de comprimento de até 15 unidades de repetição, porém, acima de alelos com comprimento de 16 unidades de repetição, esta linha se mantém sempre acima. Isso nos mostra que mesmo com menor frequência relativa da ocorrência de alelos mais longos que 15 unidades de repetição nas populações de São Paulo e Rio de Janeiro, as mutações foram relativamente maiores para alelos de maior comprimento, confirmando os dados do outros estudos (GUSMÃO *et al*., 2005; BALLANTYNE *et al*., 2010) que indicam que alelos de maior comprimento tendem a sofrer mais mutações que alelos mais curtos.

O estudo de Ballantyne e colaboradores (2010) também mostrou que marcadores que apresentam sequência repetitiva complexa, ou seja, que são compostos por mais de um trecho de repetição, tendem a ocorrer mais

mutações que os com sequência repetitiva simples (apenas 1 única sequência de repetição homogênea ininterrupta). Curiosamente em todos os marcadores do painel utilizado que apresentam sequência complexa (DYS389I/II, DYS635 e DYS390) foi encontrada pelo menos 1 mutação.

5.3 Variantes Alélicas

Durante a análise dos 401 haplótipos, foram encontradas algumas variantes alélicas incomuns para este painel de marcadores. Foi realizado um levantamento no YHRD para identificar a ancestralidade das populações predominantes em que essas variantes são observadas. Estes dados estão representadas na tabela 12. Em todos os casos em que foram observadas variantes alélicas nos pais, as mesmas foram transmitidas para os filhos.

Tabela 12: Número (N) de variantes alélicas observadas nas populações de São Paulo e Rio de Janeiro e respectiva predominância ancestral referente aos 23 Y-STRs analisados.

	São Paulo		Rio de Janeiro			YHRD
Marcador	**N**	**Variante**	**N**	**Variante**	**Total**	**Ancestralidade**
DYS570	1	21.2	1	21.2	2	Europeu
DYS19	1	Triplicação (14, 15,17)	0		1	Africano
DYS389 I	1	Duplicação (29,31)	0		1	Asiático
DYS458	6	15.2/17.2/18.2/19.2/20.2	4	17.2/18.2/20.2/21.2	10	Africano/Asiático
DYS576	1	Alelo nulo	0		1	Europeu/Asiático
DYS385	1	12.3	1	12.2	2	Europeu/Asiático
DYS392	0		1	Alelo nulo	1	Africano
DYS448	0		1	Duplicação (20,21)	1	Africano
DYS439	0		3	Duplicação (10,11; 10,12; 12,13)	3	Europeu/Asiático
	11		11		22	

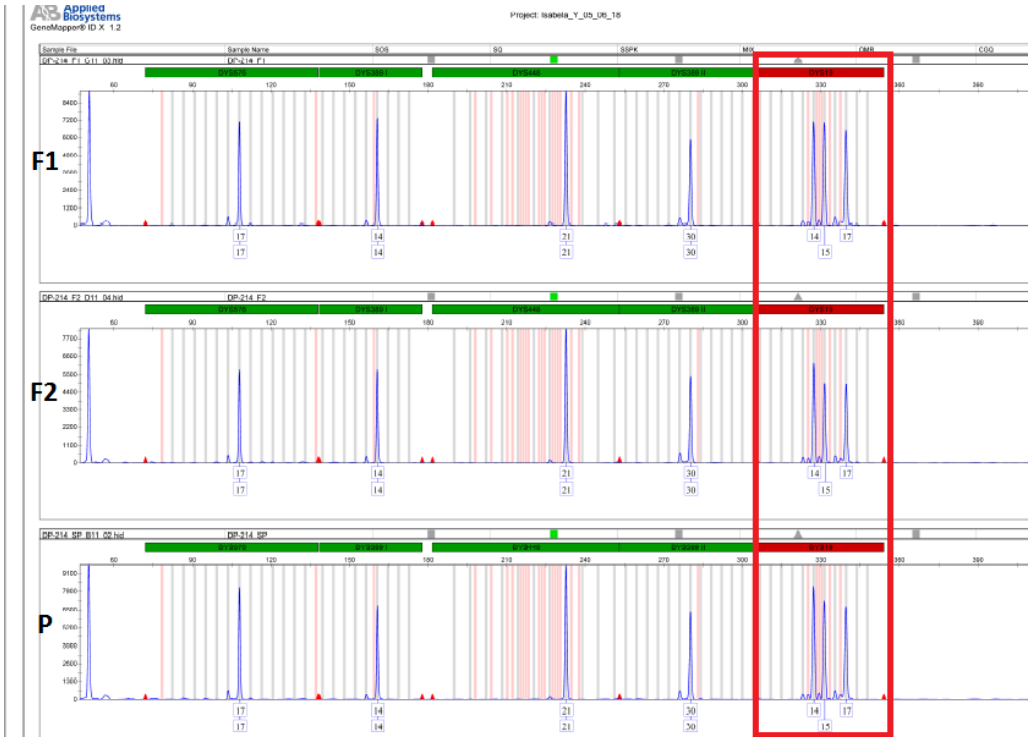
Fonte: Este trabalho.

Foram encontradas 22 variantes alélicas em 9 marcadores dos 23 Y-STR analisados. A maioria das variantes observadas foram alelos intermediários, sendo observados 14, que ocorreram em 3 marcadores diferentes (DYS570, DYS458 e DYS385). O marcador DYS458 foi o que concentrou maior ocorrência de alelos intermediários (10), o mesmo também

foi observado por PURPS e colaboradores (2014). Ao avaliar a ocorrência dessas variantes no YHRD, observamos que todos os alelos intermediários observados para o locus DYS458 são típicas de populações Afro asiáticas. Os alelos intermediários observados no locus DYS570 são predominantemente observados em populações europeias, e os observados no DYS385 são predominantemente encontrados em populações Euro asiáticas. A classificação das amostras em que essas variantes foram observadas confirmaram estes dados.

Um total de 6 variantes em números de cópias foram encontradas em 4 diferentes marcadores. Dessas variantes, 5 foram observadas duplicações de alelos, que ocorreram nos marcadores: DYS389 I, DYS448 e DYS439. Mais da metade das duplicações observadas foi vista no marcador DYS439, sendo 3 variantes no total, e todas essas variantes foram predominantemente observadas em populações de origem Euro asiáticas. A duplicação observada no locus DYS389 I, apresenta maior frequência em populações asiáticas e a duplicação no DYS448 em populações africanas. Uma variante de número de cópia apresentou amplificação de 3 alelos, e ocorreu em um caso em que havia um pai e dois filhos. Esta triplicação de alelos ocorreu no marcador DYS19 obtendo a amplificação dos alelos 14, 15 e 17. No trabalho de Purps e colaboradores (2019) também foi encontrado um caso de triplicação de alelos para este mesmo marcador, sendo encontrado em uma amostra afro-americana. Esta variante foi transmitida para os dois filhos, como mostrado na figura 8.

Figura 8: Eletroferogramas parciais de perfis alélicos de três amostras obtidos com o Powerplex®Y 23 da triplicação no locus DYS19.



F1, filho 1; F2, filho 2; P, pai.

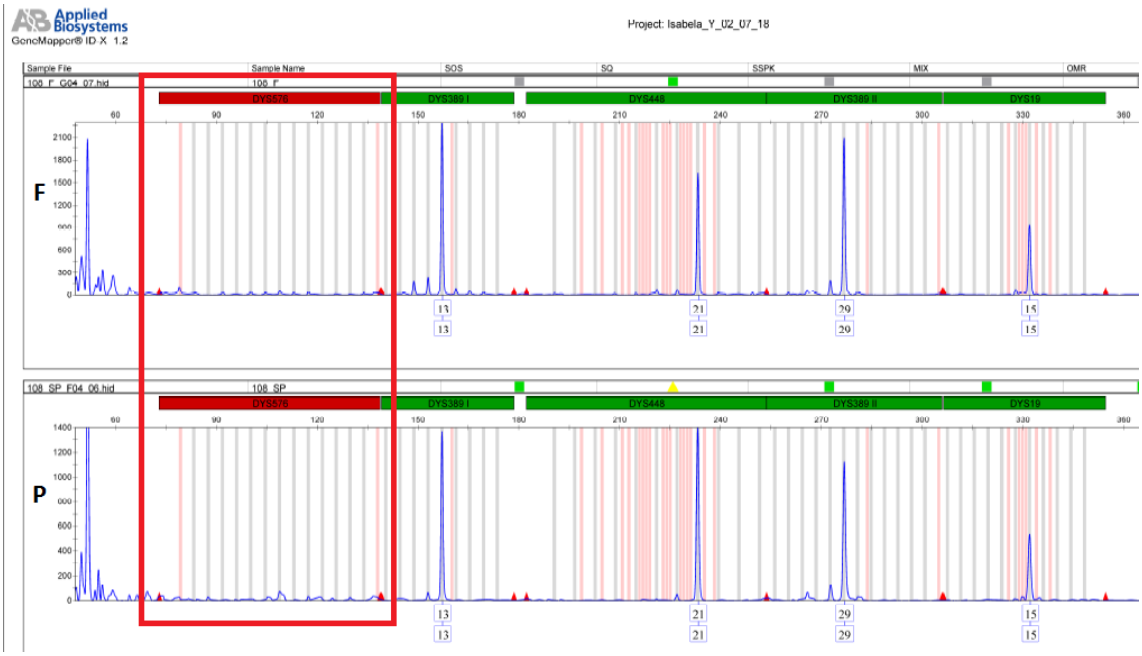
Nota: O marcador destacado em vermelho é o DYS19, o qual apresenta a amplificação de três alelos em todas as amostras (F1, F2 e P). **Fonte:** Este trabalho.

Segundo dados obtidos no YHRD, esta triplicação contendo os alelos 14, 15 e 17 apresenta frequência de 0,0003498%, ou seja, um tipo de variante extremamente rara. Já que foi relatado ao YHRD apenas uma única ocorrência. Existem muitos estudos que relatam a ocorrência de amplificações duplicadas no marcador DYS19 (PURPS *et al*., 2014, FIGUEIREDO *et al*., 2015), sendo geralmente o marcador que apresenta maiores variantes de números de cópias, porém mais de duas duplicações são raras, mas também observadas para outros marcadores (DECKER *et al*., 2008).

O estudo da frequência de amplificação de mais de um alelo por marcador no cromossomo Y é de extrema importância (DECKER *et al*., 2008), pois a análise do cromossomo Y é muito utilizada em investigações de crimes sexuais, os quais utilizam-se amostras de misturas, e esta duplicação e ou triplicação pode interferir na análise dos resultados, podendo ser mal interpretada durante a análise.

Encontramos também 2 ocorrências de alelos nulo (Figura 9), ou seja, não houve amplificação de nenhum alelo, o que pode ter ocorrido devido a mutações presentes na região de anelamento do *primer* que amplifica estas regiões (GUSMÃO *et al.*, 2006). Essas variantes ocorreram nos marcadores DYS576 e DYS392, respectivamente. Foram relatados ao YHRD 26 casos de alelo nulo para o marcador DYS576, e foi mais frequentemente observado em populações de origem Euro asiáticas e 34 para o DYS392, visto com maior frequência em populações africanas.

Figura 9: Eletroferogramas parciais de perfis alélicos de duas amostras obtidos com o Powerplex®Y 23 do alelo nulo encontrado no locus DYS576.



F, filho; P, pai.

Nota: O marcador destacado em vermelho é o DYS576, onde em ambas as amostra não houve amplificação de nenhum alelo para este marcador, sendo considerado como alelo nulo.

Fonte: Este trabalho.

No geral, as duas populações estudadas apresentaram mesmo número de variantes alélicas, sendo encontradas 11 variantes em cada população, e nas duas populações foram observados os diferentes tipos de variantes (alelos intermédios, variantes de número de cópias e alelos nulo).

Foram obtidas as frequências alélicas e diversidade gênica (DG) dos 23 marcadores Y-STR das populações de São Paulo e Rio de Janeiro e os dados

encontram-se na tabela 13. Foi verificada a distância genética entre as duas populações após o cálculo de *Fst*, que não mostrou haver diferenças significativas entre essas populações (*Fst* = 0,0011), por isso, estabeleceu-se então uma única tabela de frequências alélicas para as duas populações estudadas.

Tabela 13: Frequências alélicas, número de alelos e diversidade gênica (DG) dos 23 marcadores Y-STR obtidos com o Powerplex®Y 23 nas populações de São Paulo e Rio de Janeiro.

	DYS576	DYS389 I	DYS448	DYS389 II	DYS19	DYS391	DYS481	DYS549	DYS533	DYS438	DYS437	DYS570	DYS635	DYS390	DYS439	DYS392	DYS643	DYS393	DYS458	DYS456	YGATAH4	DYS385	Frequência Alélica	DYS385	Frequência Alélica
Alelo Nulo	0,00249															0,00249						10,13	0,00499	16,19	0,01746
7										0,00249							0,00249					10,14	0,00249	17,17	0,01247
8						0,00249			0,00499	0,00998							0,01496					10,15	0,00249	17,18	0,01746
9						0,05985			0,02244	0,10973					0,00998		0,09476				0,00499	11,11	0,02244	17,19	0,01247
10						0,53616		0,01995	0,06234	0,28678					0,08728	0,00499	0,42643				0,03491	11,12	0,00998	18,18	0,00998
10,11															0,00249							11,13	0,02494	18,19	0,00499
10,12															0,00249							11,14	0,25436	18,2	0,00748
11		0,01247				0,39401		0,20449	0,30673	0,18703					0,30673	0,50374	0,11222	0,00499			0,3616	11,15	0,04489	9,17	0,00249
12		0,202			0,00499	0,00748		0,36409	0,52618	0,38155	0,00249	0,00249			0,43641	0,04239	0,18454	0,1596	0,00249		0,53117	12,12	0,01995		
12,13															0,00249							12,13	0,00249		
13	0,00998	0,601			0,14214			0,32668	0,05736	0,02244	0,00249	0,00499			0,13466	0,38903	0,10973	0,68828		0,02244	0,06733	12,14	0,06234		
14	0,00998	0,17955			0,46384			0,06983	0,01995		0,4389	0,00499			0,01496	0,04489	0,05237	0,10973	0,02244	0,10973		12,15	0,01746		
14,15,17					0,00249																	12,16	0,00998		
15	0,05736	0,00499			0,26683			0,01496			0,4414	0,02743			0,00249	0,01247	0,00249	0,02993	0,10723	0,42893		12,17	0,00499		
15,2																			0,00249			12,19	0,00748		
16	0,16209				0,08229						0,10973	0,11721						0,00748	0,31172	0,30673		12,2,14	0,00249		
17	0,31671		0,00748		0,03491						0,00249	0,37905	0,00998						0,32918	0,11471		13,13	0,01746		
17,2																			0,00499			13,14	0,06484		
18	0,28429		0,05237		0,00249						0,00249	0,19202	0,00499	0,00249					0,14464	0,01746		13,15	0,03242		
18,2																			0,00499			13,16	0,02244		
19	0,12219		0,41397				0,00499					0,09975	0,01247						0,04239			13,17	0,02494		
19,2																			0,00499			13,18	0,01746		
20	0,02743		0,27431				0,00748					0,05985	0,05237						0,01247			13,19	0,00499		
20,2																			0,00499			14,14	0,02743		
20,21			0,00249																			14,15	0,01746		
21	0,00748		0,20698				0,07731					0,03491	0,23441	0,12718					0,00249			14,16	0,00998		
21,2												0,00748							0,00249			14,17	0,01995		
22			0,0399				0,38155					0,05736	0,13466	0,1197								14,18	0,00249		
23							0,17207					0,00998	0,40648	0,21197								14,19	0,00249		

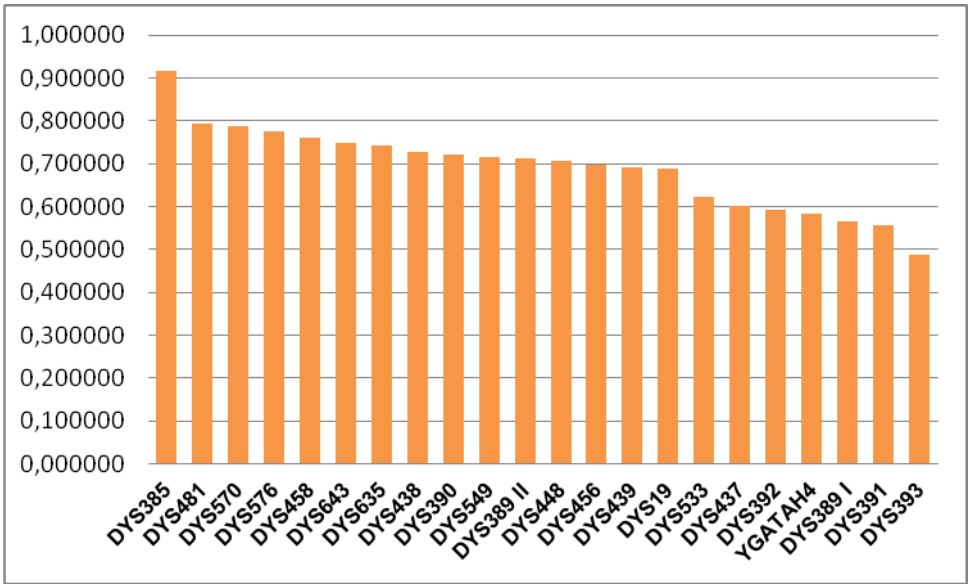
24			0,00249			0,09476						0,00249	0,13466	0,44389								14,20	0,00499
25						0,09227							0,00998	0,08229								14,21	0,00249
26				0,00499		0,04489								0,00998								15,15	0,02743
27						0,07731								0,00249								15,16	0,01746
28				0,14963		0,02993																15,17	0,01746
29				0,41397		0,01746																15,18	0,00748
29,31				0,00249																		15,19	0,00249
30				0,28429																		15,20	0,00249
31				0,11222																		16,16	0,02494
32				0,02494																		16,17	0,02494
33				0,00748																		16,18	0,02494
Nº Alelos	10	5	8	8	8	5	11	6	7	7	7	14	9	8	10	7	9	6	15	6	5		48
DG	0,77530	0,56700	0,70791	0,71390	0,68713	0,55502	0,79405	0,71519	0,62249	0,72636	0,60197	0,78867	0,74228	0,72244	0,69108	0,59238	0,74938	0,48900	0,76137	0,69767	0,58278		0,91651

Fonte: Este Trabalho.

A diversidade gênica é a probabilidade de que dois indivíduos escolhidos ao acaso na população tenham alelos distintos. Os valores de DG obtidos para as populações estudadas variaram de 0,4890 a 0,9165. Dos 23 Y-STR analisados neste trabalho o que apresentou maior diversidade gênica foi o DYS385 (0,9165). De certa forma este resultado já era esperado, uma vez que o DYS385 representa a amplificação simultânea de duas regiões idênticas do cromossomo Y, que são consideradas em forma de um bloco, ou seja, dois alelos para um único marcador. Este resultado se repete em vários estudos, sendo este marcador o que apresenta DG em diferentes populações provenientes de todos os continentes (PURPS *et al*., 2014).

O marcador DYS481 foi o que apresentou segundo maior valor (0,7940) de DG neste estudo, seguidos pelos dois marcadores RM DYS570 (0,78867) e DYS576 (0,77530). O marcador que apresentou menor valor de DG foi o DYS393 (0,4890) (Figura 10).

Figura 10: Valores das diversidades gênicas entre os 23 Y-STR obtidas com o Powerplex®Y 23 nas populações de São Paulo e Rio de Janeiro.



Fonte: Este Trabalho.

5.4 Análises haplotípicas

Na presença de perfis haplóides é importante estimar a diversidade haplotípica, definida como a probabilidade de dois indivíduos selecionados ao acaso em uma população apresentarem haplótipos diferentes para um conjunto de marcadores (EXCOFFIER, 2005).

A diversidade haplotípica foi calculada para cada população estudada, e então juntou-se com outras populações brasileiras retiradas da literatura (Tabela 6), considerando uma única população (Brasil), e calculou-se a diversidade haplotípica da população brasileira.

Ao analisar 200 haplótipos de indivíduos nascidos no estado de São Paulo, a diversidade haplotípica foi igual a 1,0000 +/- 0,0005, ou seja, dentre os haplótipos encontrados não houve nenhum igual ao outro, o que significa que para esse conjunto de marcadores há uma probabilidade baixa de dois indivíduos apresentarem o mesmo haplótipo nesta população. Cada haplótipo apresentou frequência haplotípica de 0,005. A capacidade de discriminação deste conjunto de marcadores para a população de São Paulo foi de 1,000 (Tabela 14).

Para a população do Rio de Janeiro foram considerados 201 haplótipos. A diversidade haplotípica foi de 0,99995 +/- 0,0005. Dentre os 201 haplótipos encontrados, 2 foram idênticos, sendo a frequência haplotípica desses haplótipos de 0,00995, enquanto os haplótipos únicos apresentaram frequência haplotípica de 0,00498. A capacidade de discriminação deste conjunto de marcadores para a população do Rio de Janeiro foi de 0,99502 (Tabela 13).

Estes valores mostram uma grande diversidade genética para as duas populações estudadas quando utilizado este conjunto de marcadores.

Tabela 14: Diversidade haplotípica e capacidade de discriminação das populações estudadas e em conjunto com populações de outras regiões do Brasil.

População	N	Nº de haplótipos diferentes	Diversidade haplotípica	Referência	Capacidade de discriminação
São Paulo 1	200	200	1,0000 +/- 0,0005	Este trabalho	1,00000
Rio de Janeiro 1	201	200	0,99995 +/- 0,0005	Este trabalho	0,99502
Manaus	204	203	0,99995 +/- 0,0005	Dados colaborativos	0,99510
Brasília	410	327	0,9987 +/- 0,0003	Ferreira *et al* (2015)	0,79756
Rio de Janeiro 2	123	123	1,0000 +/- 0,0010	PURPS *et al* (2014)	1,00000
São Gabriel da Cachoeira	61	58	0,9984 +/- 0,0033	PURPS *et al* (2014)	0,95082
São Paulo 2	120	120	1,0000 +/- 0,0011	PURPS *et al* (2014)	1,00000
São Paulo total	320	319	0,99998 +/- 0,0005	Este trabalho + PURPS *et al* (2014)	0,99688
Rio de Janeiro total	324	323	0,99998 +/- 0,0005	Este trabalho + PURPS *et al* (2014)	0,99691
Brasil	**1319**	**1214**	**0,99984 +/- 0,0005**		**0,92039**

Fonte: Este trabalho.

Ao juntarmos as populações de São Paulo e Rio de Janeiro deste trabalho com as de PURPS e colaboradores (2014), obteve-se diversidade haplotípica de 0,99998 +/- 0,0005 para ambas populações, com capacidade de discriminação de 0,99688 e 0,99691 para São Paulo e Rio de Janeiro respectivamente. Quando englobamos todas as populações, em uma única, denominada de Brasil, temos no total 1.319 haplótipos, sendo que 1.214 foram únicos. Esta população obteve diversidade haplotípica de 0,99984 +/- 0,0005, com capacidade de discriminação de 0,92039 para o PPY23. O haplótipo mais frequente foi compartilhado por 5 indivíduos.

Foi calculado o valor correspondente de probabilidade de *match* entre os haplótipos da população brasileira para este conjunto de marcadores, e obteve-se um valor de $9{,}18 \times 10^{-4}$, sendo a probabilidade de *match* de 1 em cada 1089 haplótipos avaliados. Este valor se assemelha ao estudo de PURPS *et al*. (2014), em que se obteve uma probabilidade de match de 1 a cada 1086 haplótipos para populações da América Latina, mostrando que os dados aqui

obtidos se encontram em conformidade com os da literatura para este mesmo painel de marcadores.

Em relação à coincidência dos haplótipos avaliados da população brasileira (Tabela 15), a que apresentou maior coincidência de perfis foi a população de Brasília com ela mesma (62), seguida por Brasília com Manaus (4), Brasília com Rio de Janeiro e com São Paulo (3). Sendo assim a população de Brasília além de apresentar maior coincidência de perfis, haplotipicos com as outras populações avaliadas, ou seja, inter-populacional, foi a que apresentou também maior coincidência de perfis intra-populacional.

Tabela 15: Coincidência entre os perfis haplotípicos dentre as populações estudadas.

	São Paulo 1[1]	**Rio de Janeiro 1[1]**	**Manaus[2]**	**Brasília[3]**	**Rio de Janeiro 2[4]**	**São Gabriel da Cachoeira[4]**	**São Paulo 2[4]**
São Paulo 1[1]	-						
Rio de Janeiro 1[1]	-	1					
Manaus[2]	-	-	1				
Brasília[3]	3	-	4	62			
Rio de Janeiro 2[4]	-	-	3	2	-		
São Gabriel da Cachoeira4	-	-	-	1	-	3	
São Paulo 2[4]	1	-	-	2	-	-	-

[1], Este Trabalho; [2], Dados colaborativos; [3], Ferreira *et al* (2015); [4], PURPS *et al* (2014).

Fonte: Este trabalho.

O grande número de coincidências de haplótipos da população de Brasília com outras populações do Brasil se mostra em conformidade com a história da cidade, que foi criada para ser a capital do Brasil, recebendo indivíduos de todas as partes do país (IBGE). A população do Rio de Janeiro deste estudo apenas apresentou uma única coincidência de perfil, sendo essa com outro perfil da própria população.

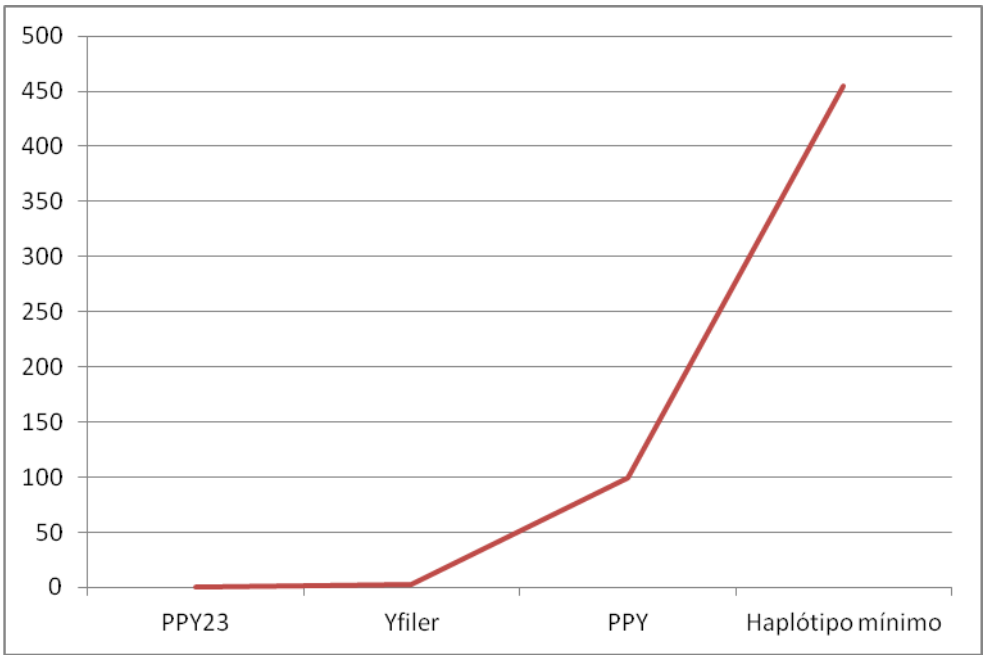
Todos os haplótipos da população de São Paulo e Rio de Janeiro deste estudo foram confrontados com os haplótipos incluídos no YHRD para busca de possíveis coincidências haplotípicas. Esta busca primeiramente foi realizada com o painel de marcadores PPY23. Para a população de São Paulo, e a

média de coincidência com os haplótipos depositados no YHRD foi de 0,08. Dos 200 haplótipos de São Paulo, 16 apresentaram *match*, sendo que destes, 6 foram com haplótipos provenientes de populações brasileiras.

Para a população do Rio de Janeiro, a média de coincidência foi de 0,085. Dos 201 haplótipos, 17 apresentaram *match* com outros depositados no YHRD, sendo que desses, 6 foram coincidentes com haplótipos de populações brasileiras.

Após esta análise, para averiguar a coincidência entre os haplótipos deste estudo com os depositados no YHRD para outros painéis de Y-STR retirou-se alguns marcadores que não estão contidos nesses outros painéis de marcadores: AmpFLSTR® Yfiler® PCR Amplification Kit (Applied Biosystems® – Thermo Fischer Scientific - YFiler) que possui 17 Y-STR; PowerPlex® Y System (Promega Corporation, USA - PPY) que possui 12 Y-STR; e o conjunto de haplótipo mínimo com 9 marcadores Y-STR. A figura 11 mostra a média de coincidências de haplótipos da população de São Paulo para os conjuntos de Y-STR citados acima.

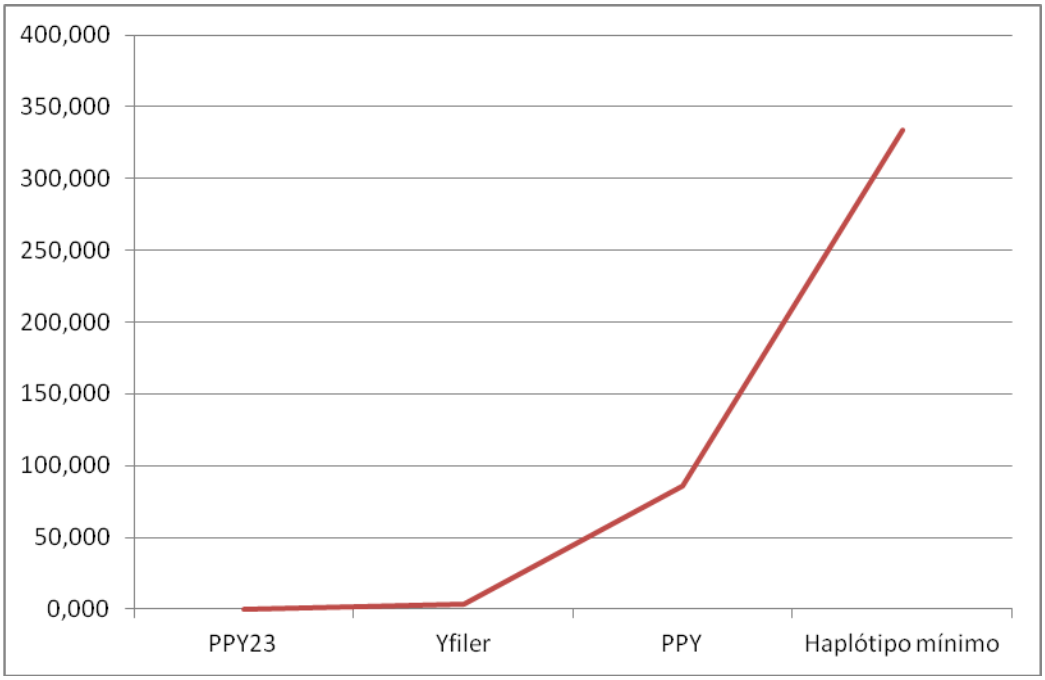
Figura 11: Média de coincidência entre os haplótipos da população de São Paulo quando confrontados com dados do YHRD para diferentes conjuntos de marcadores Y-STR (Powerplex®Y 23, YFiler®, PowerPlex® Y System e Haplótipo mínimo).



Fonte: Este trabalho.

Os resultados mostram um elevado aumento da média de coincidência entre haplótipos quando se diminui a quantidade de marcadores analisados, Resultados semelhantes ocorreram para a população do Rio de Janeiro (Figura 12).

Figura 12: Média de coincidência entre os haplótipos da população do Rio de Janeiro quando confrontados com dados do YHRD para diferentes conjuntos de marcadores Y-STR (Powerplex®Y 23, YFiler®, PowerPlex® Y System e Haplótipo mínimo).



Fonte: Este trabalho.

Estes resultados já eram esperados, pois quanto menos marcadores analisados em conjunto, menor será o poder de discriminar um indivíduo, e consequentemente maior a chance de coincidência entre haplótipos (BUTLER, 2005). Um estudo colaborativo (PURPS *et al*., 2014) que analisou 19.630 haplótipos Y-STR com populações oriundas de todos os continentes obteve valores de capacidade de discriminação para os mesmos conjuntos de marcadores citados acima, o que também foi calculado para as populações deste estudo, e estão indicados na tabela 16.

Tabela 16: Capacidade de discriminação de diferentes conjuntos de Y-STR da população estudada em comparação com dados da literatura.

	Referência	Powerplex®Y 23	YFiler®	PowerPlex® Y System	Haplótipo mínimo
Sâo Paulo + Rio de Janeiro	Este Trabalho	0,9975	0,9550	0,8117	0,6532
População Global	PURPS *et al*. (2014)	0,9608	0,8569	0,5831	0,4304

Fonte: Este trabalho

Todos esses dados confirmam grande diversidade genética, baseada em linhagens patrilíneas, tanto das populações estudadas, quanto da população brasileira como um todo, além de que este painel de marcadores (PPY23) se mostrou altamente discriminativo para as populações estudadas, obtendo valores de capacidade de discriminação acima dos obtidos para a população global pelo estudo de PURPS e colaboradores (2014).

Foram incluídos no total 368 novos perfis haplotípicos para o conjunto de marcadores de 23 Y-STR. Sendo 188 perfis referentes a cada uma das populações estudadas.

Todos os dados obtidos no YHRD se referem à versão 61.

5.5 Comparações genéticas entre as populações

Baseando-se na comparação de Y-STR entre duas ou mais populações é possível estimar a distância genética (*Fst*) entre elas. Isto é útil para esclarecer as relações evolutivas entre populações (TAKEZAKI, 1996). *Fst* é uma medida de distância genética, e quanto menor o seu valor maior a similaridade entre as populações.

Com os dados obtidos utilizando o PPY23, foi realizada a análise da distância genética entre as populações dos estados de São Paulo e Rio de Janeiro deste trabalho. Além disso, as mesmas foram comparadas com dados de outras populações brasileiras e estrangeiras disponível na literatura para o mesmo painel de marcadores. Como a população do Brasil é altamente miscigenada, a escolha dessas populações se deu pela influência das mesmas na formação da população brasileira baseando-se em dados históricos, além

da escolha por populações que apresentam proximidade geográfica. Para esta análise (*Fst*) utilizou-se um total de 22 populações, e o valor de *p* como teste de significância foi baseado em 50.000 permutações.

Inicialmente, analisou-se a relação genética entre a população de São Paulo deste estudo, com a população de São Paulo de PURPS e colaboradores (2014). O mesmo foi feito com a população do Rio de Janeiro. O resultado obtido das populações de São Paulo e Rio de Janeiro deste estudo foram unidas às de PURPS e colaboradores (2014), pois averiguou-se não haver diferenciação genética entre estas populações com o cálculo (Tabela 17), formando uma populações únicas para São Paulo (N=320) e Rio de Janeiro (N=324).

Tabela 17: Distâncias genéticas (*Fst*) e valor de *p* correspondente entre as populações de São Paulo 1 e Rio de Janeiro 1 deste trabalho com a população de São Paulo 2 e Rio de Janeiro 2.

Valores de P considerados significantes após correção de Bonferroni: ≤0,008333, considerando as comparações entre as 4 populações.

	São Paulo[1] X São Paulo[2]	**Rio de Janeiro[1] X Rio de Janeiro[2]**
Valor *Fst*	0,00000 (*p*= 0,50905)	0,00419 (*p*= 0,03019)

[1] Este trabalho, [2] PURPS, *et al*. (2014).
Fonte: Este trabalho.

.

Os valores de *Fst* e os respectivos valores de *p* estão apresentados na tabela 18.

Tabela 18: Distâncias genéticas (*Fst*) entre populações (abaixo da diagonal) e valor de *p* correspondente (acima da diagonal).* Valores de P considerados significantes após correção de Bonferroni: ≤0,00022.

	São Paulo[1]	Rio de Janeiro[1]	Manaus[2]	Brasília[3]	São Gabriel da Cachoeira[4]	Argentina[4]	Bolívia[4]	Bolívia_NAM[4]	Paraguai[5]	Peru[4]	Equador[6]	Costa Rica[4]	Panamá[4]	Alemanha[4]	Itália[4]	Espanha[4]	Portugal[4]	Zimbábue[4]	EUA_Africanos[4]	EUA_Europeus[4]	EUA_Asiáticos[4]	EUA_Hispânicos[4]
São Paulo[1]	*	0,36091	0,02641	0,1967	*0,00000**	0,0385	*0,00000**	*0,00000**	0,00008	*0,00000**	*0,00000**	0,07657	0,00289	*0,00000**	0,00034	*0,00000**	0,00923	*0,00000**	*0,00000**	*0,00000**	*0,00000**	0,00417
Rio de Janeiro[1]	0,0001	*	0,01138	0,0401	*0,00000**	0,00283	*0,00000**	*0,00000**	*0,00000**	*0,00000**	*0,00000**	0,04604	0,00181	*0,00000**	0,00022*	*0,00000**	0,00339	*0,00000**	*0,00000**	*0,00000**	*0,00000**	0,00034
Manaus[2]	0,00265	0,00345	*	0,0098	*0,00000**	0,07535	*0,00000**	*0,00000**	0,0001*	*0,00000**	*0,00000**	0,04466	0,0011	*0,00000**	0,00026	*0,00000**	0,00371	*0,00000**	*0,00000**	*0,00000**	*0,00000**	0,02414
Brasília[3]	0,00052	0,00162	0,00349	*	*0,00000**	0,01126	*0,00000**	*0,00000**	0,00002*	*0,00000**	*0,00000**	0,03751	0,00078	*0,00000**	0,00006*	*0,00000**	0,00706	*0,00000**	*0,00000**	*0,00000**	*0,00000**	0,00056
São Gabriel da Cachoeira[4]	0,06581	0,06624	0,05549	0,0691	*	*0,00000**	*0,00000**	*0,00000**	*0,00000**	*0,00000**	*0,00000**	*0,00000**	*0,00000**	*0,00000**	*0,00000**	*0,00000**	*0,00000**	*0,00000**	*0,00000**	*0,00000**	*0,00000**	*0,00000**
Argentina[4]	0,00204	0,00441	0,00171	0,003	0,05954	*	*0,00000**	*0,00000**	0,0178	*0,00000**	*0,00000**	0,21251	0,00008*	*0,00000**	0,0003	0,00002*	0,05481	*0,00000**	*0,00000**	*0,00000**	*0,00000**	0,01935
Bolívia[4]	0,07634	0,07837	0,06902	0,0778	0,06144	0,07047	*	0,03589	*0,00000**	0,00426	0,00008*	*0,00000**	*0,00000**	*0,00000**	*0,00000**	*0,00000**	*0,00000**	*0,00000**	*0,00000**	*0,00000**	*0,00000**	*0,00000**
Bolivia_NAM[4]	0,11496	0,11642	0,10765	0,1176	0,08265	0,11216	0,0108	*	*0,00000**	*0,00000**	*0,00000**	*0,00000**	*0,00000**	*0,00000**	*0,00000**	*0,00000**	*0,00000**	*0,00000**	*0,00000**	*0,00000**	*0,00000**	*0,00000**
Paraguai[5]	0,00674	0,00763	0,00797	0,0067	0,07553	0,0023	0,0875	0,13615	*	*0,00000**	*0,00000**	0,00777	*0,00000**	*0,00000**	*0,00000**	0,00367	0,05674	*0,00000**	*0,00000**	0,00004*	*0,00000**	*0,00000**
Peru[4]	0,03742	0,04047	0,03354	0,0364	0,04938	0,03046	0,019	0,04571	0,0451	*	0,01212	*0,00000**	*0,00000**	*0,00000**	*0,00000**	*0,00000**	*0,00000**	*0,00000**	*0,00000**	*0,00000**	*0,00000**	*0,00000**
Equador[6]	0,01622	0,0185	0,01405	0,0156	0,03887	0,01298	0,0263	0,05863	0,023	0,00616	*	0,00002*	0,00014*	*0,00000**	*0,00000**	*0,00000**	*0,00000**	*0,00000**	*0,00000**	*0,00000**	*0,00000**	0,00008
Costa Rica[4]	0,00187	0,00245	0,00281	0,0026	0,06063	0,0008	0,0696	0,11342	0,0043	0,03083	0,01336	*	0,00147	*0,00000**	0,00147	0,00006*	0,02439	*0,00000**	*0,00000**	0,00006*	*0,00000**	0,06045
Panamá[4]	0,00798	0,00902	0,00969	0,0102	0,05257	0,01503	0,051	0,07905	0,025	0,02838	0,01031	0,01095	*	*0,00000**	0,00002*	*0,00000**	*0,00000**	*0,00000**	*0,00000**	*0,00000**	*0,00000**	0,00108
Alemanha[4]	0,01482	0,01395	0,01912	0,014	0,08125	0,01619	0,0837	0,11866	0,0238	0,04497	0,02765	0,01565	0,02364	*	*0,00000**	*0,00000**	*0,00000**	*0,00000**	*0,00000**	*0,00000**	*0,00000**	*0,00000**
Itália[4]	0,00482	0,00487	0,00664	0,005	0,07057	0,00535	0,0866	0,12224	0,0097	0,04196	0,02264	0,00562	0,01474	0,0148	*	*0,00000**	0,00012*	*0,00000**	*0,00000**	*0,00000**	*0,00000**	*0,00000**
Espanha[4]	0,01384	0,01489	0,01453	0,0113	0,09046	0,00769	0,1044	0,15991	0,0022	0,05746	0,03347	0,0102	0,03733	0,0314	0,01602	*	0,00255	*0,00000**	*0,00000**	0,00006*	*0,00000**	*0,00000**
Portugal[4]	0,00354	0,00466	0,00526	0,0036	0,08196	0,00195	0,0947	0,14274	0,0015	0,04892	0,02571	0,00369	0,02229	0,0186	0,00635	0,00381	*	*0,00000**	*0,00000**	0,00022*	*0,00000**	*0,00000**
Zimbábue[4]	0,175	0,17242	0,19516	0,1833	0,24748	0,21409	0,2472	0,23561	0,2298	0,21871	0,18575	0,20128	0,14914	0,1975	0,19396	0,25671	0,22496	*	*0,00000**	*0,00000**	*0,00000**	*0,00000**
EUA_Africanos[4]	0,05793	0,05651	0,07134	0,0622	0,11606	0,08323	0,115	0,12549	0,0966	0,08785	0,06889	0,07256	0,04138	0,0719	0,07189	0,11516	0,08946	0,04802	*	*0,00000**	*0,00000**	*0,00000**
EUA_Europeus[4]	0,01637	0,01494	0,01512	0,0132	0,09398	0,01023	0,1089	0,16143	0,0054	0,06025	0,03888	0,01012	0,03959	0,0251	0,01542	0,00477	0,00636	0,25789	0,11407	*	*0,00000**	*0,00000**
EUA_Asiáticos[4]	0,06891	0,06845	0,07617	0,0702	0,12019	0,0789	0,1207	0,12977	0,0944	0,09158	0,06934	0,07551	0,05745	0,0803	0,06842	0,11213	0,08983	0,16965	0,07601	0,11674	*	*0,00000**
EUA Hispânicos[4]	0,00323	0,00489	0,00232	0,0043	0,04948	0,00227	0,052	0,09016	0,0089	0,02309	0,00629	0,00189	0,008	0,0184	0,00925	0,01535	0,00868	0,18603	0,06681	0,01946	0,06635	*

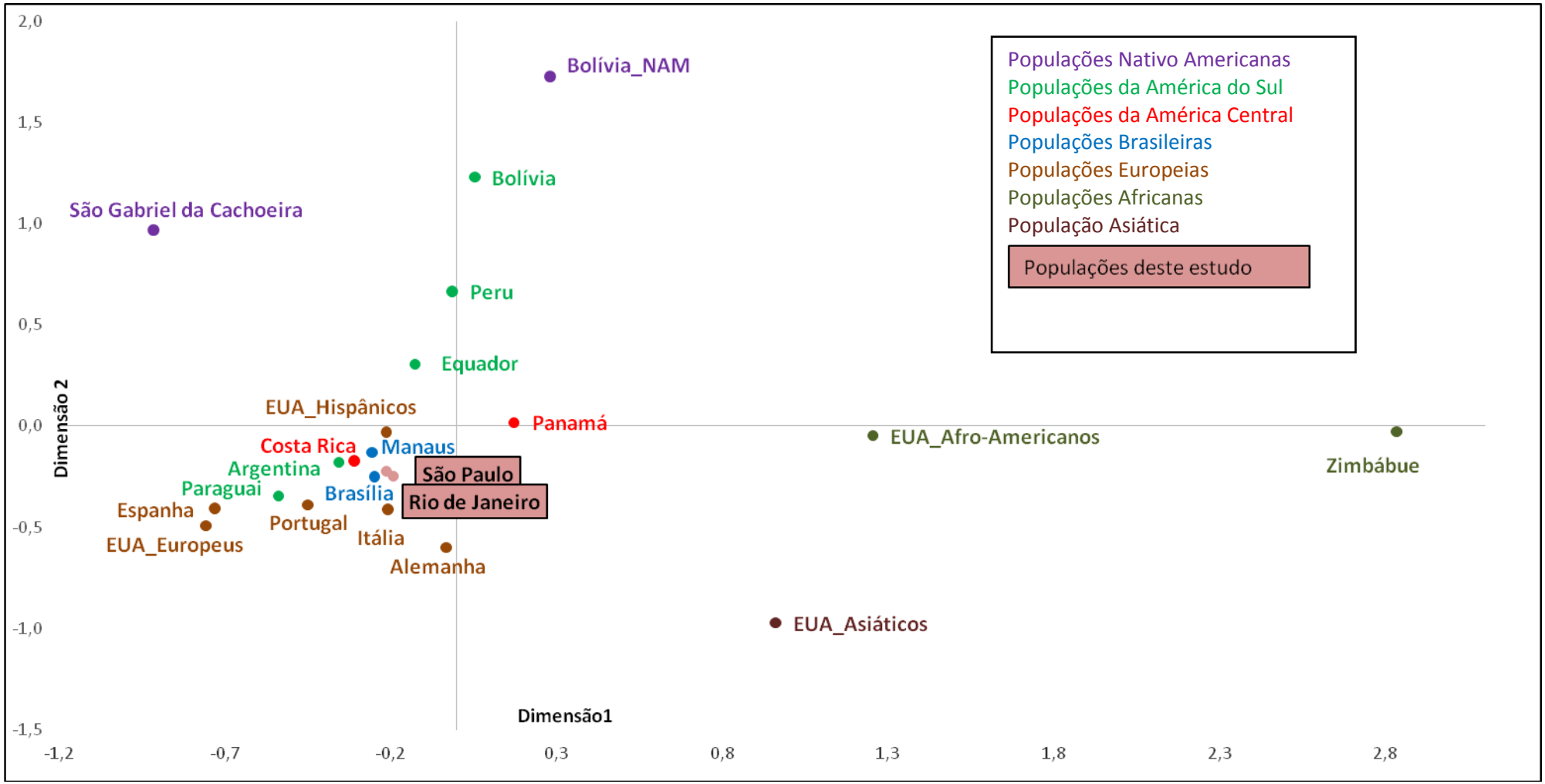
[1], Este Trabalho; [2], Dados colaborativos; [3], Ferreira *et al* (2015); [4], PURPS *et al* (2014), [5], RIBEIRO *et al*. (2018), [6], TOSCANINI *et al*. (2018). **Fonte:** Este trabalho.

A fim de melhorar a visualização dos resultados, com os valores de distância genética realizou-se uma análise em escalonamento multidimensional (MDS), disponibilizado em gráfico de 2 dimensões (Figura 13). A adição de uma terceira dimensão ao gráfico MDS não revelou diferenças importantes (dados não mostrados) e o valor de stress (0,076) foi pouco menor que o obtido para duas dimensões (0,036). Cada ponto neste gráfico refere-se a uma população. A interpretação é feita pela distância entre os pontos, quanto mais próximo um ponto do outro maior a proximidade genética. O gráfico possui quatro quadrantes e a intersecção entre os eixos corresponde ao ponto zero.

Ao visualizar o gráfico, podemos notar a proximidade entre São Paulo e Rio de Janeiro (*Fst*= 0,0001; *p*= 0,36091), mostrando não haver diferença genética significativa entre as mesmas, e que dentre todas as comparações realizadas, foi a que apresentou menor valor de *Fst*. A segunda menor distância foi entre São Paulo e Brasília (*Fst*= 0,00052; *p*= 0,1967). Brasília também não mostrou haver diferença significativa com a população do Rio de Janeiro (*Fst*= 0,0016; *p*= 0,0401). Manaus também se mostrou bem próxima a essas 3 populações brasileiras. No geral, todas as populações brasileiras que apresentam características de misturas, ou seja, com miscigenação genética, se mostraram próximas entre si, e próximas principalmente de populações europeias e da América do Sul que também apresentam forte influência europeia (RIBEIRO *et al*., 2018, TOSCANINI *et al*., 2018).

Dentre as populações europeias, São Paulo mostrou-se mais próxima de hispânicos residentes na América do Norte (*Fst*= 0,00323; *p*= 0,00417) e Portugal (*Fst*= 0,00354; *p*= 0,00923) e Itália (*Fst*= 0,00482; *p*= 0,00034), O mesmo foi visto para a população do Rio de Janeiro (Portugal: *Fst*= 0,00466; *p*= 0,00339; Itália: *Fst*= 0,00487; *p*= 0,00022; Hispânicos: *Fst*= 0,00489; *p*= 0,00034) confirmando a grande influência europeia na formação do Brasil, principalmente na região Sudeste, como mostra os dados históricos, relacionados principalmente à imigração destas populações. Dentre as populações europeias, a que se mostrou menos próxima às populações estudadas foi a população da Alemanha.

Figura 13: Gráfico MDS bidimensional elaborado a partir das distâncias genéticas (*Fst*) entre as populações estudadas (São Paulo e Rio de Janeiro) com populações de outras regiões brasileiras e provenientes de outros países. **Fonte:** Este trabalho.



Dentre as populações europeias, São Paulo mostrou-se mais próxima de hispânicos residentes na América do Norte (*Fst*= 0,00323; *p*= 0,00417) e Portugal (*Fst*= 0,00354; *p*= 0,00923) e Itália (*Fst*= 0,00482; *p*= 0,00034), O mesmo foi visto para a população do Rio de Janeiro (Portugal: *Fst*= 0,00466; *p*= 0,00339; Itália: *Fst*= 0,00487; *p*= 0,00022; Hispânicos: *Fst*= 0,00489; *p*= 0,00034) confirmando a grande influência europeia na formação do Brasil, principalmente na região Sudeste, como mostra os dados históricos, relacionados principalmente à imigração destas populações. Dentre as populações europeias, a que se mostrou menos próxima às populações estudadas foi a população da Alemanha.

A população nativo-americana de São Gabriel da Cachoeira foi a população brasileira que se mostrou mais distante das demais, pelo fato de apresentar ancestralidade nativo americana, mostrando ser mais próxima de outras populações nativas (Bolívia_NAM) e sul americanas (Bolívia e Peru). Sabe-se que a população indígena é a mais antiga a habitar o Brasil, entretanto tratando-se de linhagens masculinas, não apresentam forte influência genética na população brasileira, já que diversos estudos mostram a grande influência europeia em todo território brasileiro, porém em diferentes proporções (de AZEVEDO *et al*., 2009, PALHA *et al*., 2014, OLIVEIRA *et al*., 2014, FIGUEIREDO *et al*., 2015, RESQUE *et al*., 2016, STANGE, *et al*., 2019). Essas características se mostram em acordo com a história de povoamento do Brasil, que após o descobrimento pelos portugueses, em 1500, houve um grande declínio do número de nativo-americanos, devido a doenças e trazidas pelos portugueses, além da forte resistência aos mesmos (IBGE).

A população africana de Zimbábue foi a que mais se dispersou no gráfico perante as demais populações, sendo a que apresentou maiores valores de diferenças genéticas, comparada as demais. O maior valor de *Fst* encontrado foi entre a população de Zimbábue e de europeus residentes na América do Norte (*Fst*= 0,25789; *p*=0,00000)

A população de asiáticos residentes na América do Norte mostrou ser pouco mais distante de populações com maior ancestralidade europeia, sendo que a menor distância genética encontrada para esta população foi com a do Panamá (*Fst*= 0,05745; *p*=0,00000). Historicamente sabe-se que populações asiáticas, principalmente japonesas apresentam influência na população

brasileira, porém pequena, principalmente ao comprar com populações europeias.

Esse padrão de ancestralidade da população brasileira (predominante influência europeia), se mostra um pouco diferente ao serem avaliados mtDNA e cromossomos autossômicos, que mostram maior variação na ancestralidade da população de acordo com a localização geográfica da mesma (RESQUE *et al*., 2010, MOURA *et al*., 2015). No Sudeste, apesar da maior influência ancestral ser europeia, estudos envolvendo cromossomos autossômicos, mostraram maior miscigenação desta população, apresentando maiores proporções da contribuição africana e nativo-americana, quando são analisadas apenas linhagens patrilíneas (MOURA, *et al*., 2015, BRAGANHOLI, 2016). Além de que ao analisar dados de mtDNA que avaliam linhagens maternas, a influência de nativo-americanos e africanos se mostrou maior em relação aos obtidos em europeus (SANCHES, 2014, BERNARDO, *et al*., 2014, BRAGANHOLI, 2016).

5.6 Ancestralidades dos haplogrupos

Com auxílio da ferramenta online Nevgen, foi realizada a inferência dos haplogrupos de cada amostra das populações de São Paulo e Rio de Janeiro, com base em seus haplótipos Y-STR. Apesar desta não ser a forma mais precisa para classificar uma amostra em haplogrupo, baseando-se apenas em seu haplótipo, esta ferramenta permite prever qual o haplogrupo pertencente à amostra em questão. Cada uma das amostras foi classificada em seu respectivo haplogrupo, e a partir dos dados referentes à principal ocorrência de cada haplogrupo e a classificação de outros estudos que também realizaram a distinção dos haplogrupos por ancestralidade (BRIÓN *et al*., 2005a; SILVA *et al*., 2006; DE AZEVEDO *et al*., 2009; SÃO BENTO *et al*., 2009), foi realizada uma estimativa da contribuição continental paterna nas populações de São Paulo e Rio de Janeiro.

Foram identificados 28 haplogrupos diferentes na população de São Paulo (Tabela 18) e 25 na população do Rio de Janeiro (Tabela 19).

Tabela 19: Frequência dos haplogrupos na população de São Paulo (N=200) baseada nos resultados obtidos pelo Nevgen.

Haplogrupos	Origem dos Haplogrupos	Frequência
R1b	Europeu	43,00%
E1b1a	Africano/Europeu	13,50%
E1b1b	Africano	13,00%
J2a1	Europeu	8,00%
I2a1a	Europeu	3,00%
J1a2a1a2	Europeu	3,00%
I1	Europeu	1,50%
G2a2b2a1c	Europeu	0,02
E2	Africano	1,00%
G1	Asiático	1,00%
G2a1	Europeu	1,00%
J2b2a	Europeu	1,00%
Q1a3a	Nativo Americano	1,00%
R1a	Europeu	1,00%
T	Europeu	1,00%
D1b	Asiático	0,50%
G2a2a	Europeu	0,50%
G2a2b1	Europeu	0,50%
G2a2b2	Europeu	0,50%
G2a2b2a1a	Europeu	0,50%
H1a1a	Asiático	0,50%
I2a1	Europeu	0,50%
J1a3	Europeu	0,50%
L	Europeu	0,50%
L1b	Asiático	0,50%
O1b2	Asiático	0,50%
Q1a1	Asiático	0,50%
I2a2a	Europeu	0,50%

Fonte: Este trabalho.

Em ambas as populações o haplogrupo mais frequentemente encontrado foi o R1b, de origem europeia que também é o haplogrupo mais frequente no continente europeu conforme dados do YHRD, além de também ser o haplogrupo mais observado em estudos realizados com populações de outras regiões do país que apresentaram predominante influência europeia (PALHA *et al* 2012; OLIVEIRA *et al*, 2014; FIGUEIREDO, *et al*., 2015, RESQUE *et al*., 2016,). O segundo haplogrupo mais encontrado para a população de São Paulo foi o E1b1a (13,5%) e na população do Rio de Janeiro

foi o E1b1b (13,43%). Ambos haplogrupos são encontrados mais frequentemente em populações africanas.

Tabela 20: Frequência dos haplogrupos na população do Rio de Janeiro (N=201) baseada nos resultados obtidos pelo Nevgen.

Haplogrupos	Origem dos Haplogrupos	Frequência
R1b	Europeu	40,80%
E1b1b	Africano	13,43%
E1b1a	Africano	11,94%
J2a1	Europeu	5,47%
I1	Europeu	3,48%
J2b2a	Europeu	3,48%
G2a2b1	Europeu	2,99%
I2a1a	Europeu	2,99%
B2a1	Africano	1,49%
I2a2a	Europeu	1,49%
Q1a3a	Nativo Americano	1,49%
T	Europeu	1,49%
R1a	Europeu	1,49%
G2a2b2a1b	Europeu	1,00%
H1a1a	Asiático	1,00%
J1a3	Europeu	1,00%
E2	Africano	1,00%
E1a	Africano	0,50%
G2a	Europeu	0,50%
G2a2	Europeu	0,50%
G2a2a	Europeu	0,50%
G2a2b1	Europeu	0,50%
G2a2b2a1c	Europeu	0,50%
I2c1a1	Europeu	0,50%
J1a2a1a2	Europeu	0,50%

Fonte: Este trabalho.

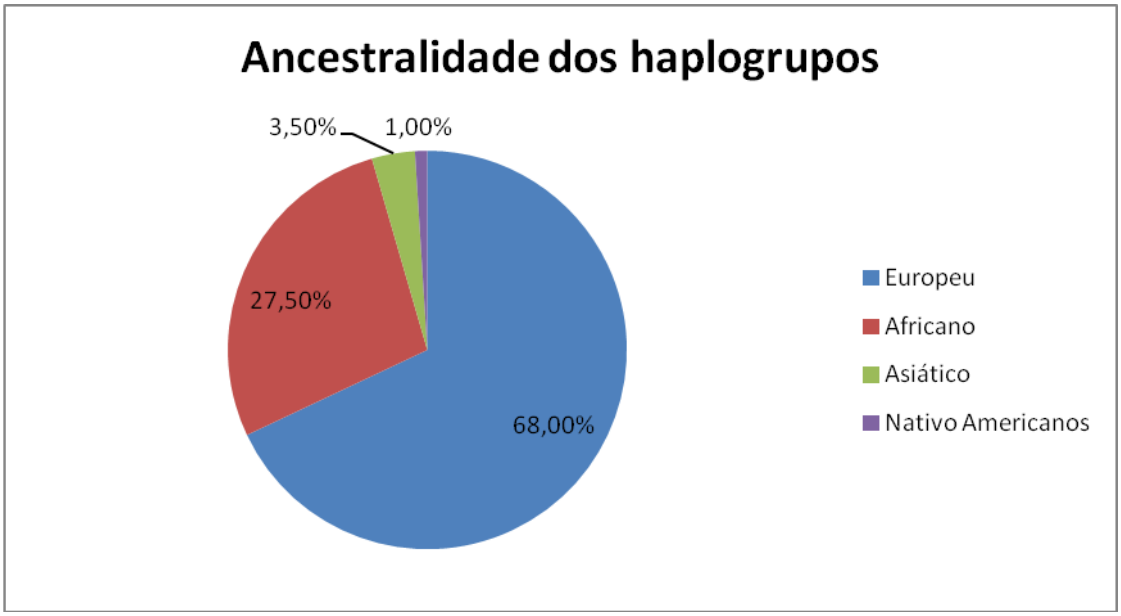
O haplogrupo Q1a3a é a maior linhagem dentre os nativo americanos (BRIÓN *et al*., 2005, UNDERHILL; KIVISILD, 2007), sendo quase completamente restrito às Américas (ZEGURA *et al*., 2004), e foi o único haplogrupo proveniente dessas populações observado neste estudo, mostrando a presença da ancestralidade nativo americana na população estudada.

Haplogrupos europeus foram os que mais se diversificaram perante haplogrupos originários da África e de populações nativo americanas, sendo o

continente que mais apresentou contribuição na ancestralidade das duas populações deste estudo.

Assim, as linhagens paternas da população de São Paulo obtidas neste estudo foram consideradas: 68,0 % com origem europeia, 27,5% origem africana, 3,5% origem asiática e 1,0% de origem nativa americana (Figura 14).

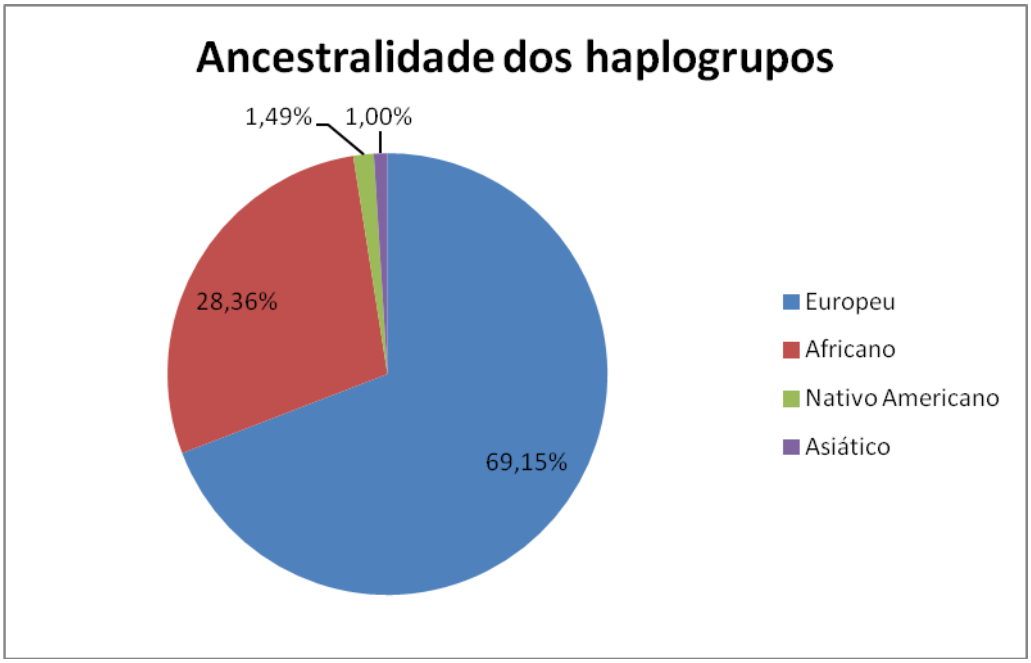
Figura 14: Classificação das amostras de São Paulo segundo a ancestralidade dos haplogrupos.



Fonte: Este trabalho.

A população do Rio de Janeiro foi considerada: 69,15% de origem europeia, 28,36% de origem africana, 1,49% de origem nativa americana e 1% de origem asiática (Figura 15). A ordem de predominância da origem ancestral entre as duas populações se mostrou bastante semelhante. A população de São Paulo apresentou contribuição asiática pouco maior que a população do Rio de Janeiro, o que é explicado historicamente, já que com o início do período republicano, o Brasil recebeu grande imigração em massa e os japoneses fizeram parte desse movimento, e sua grande maioria deslocou-se para a região de São Paulo, sendo o estado que mais recebeu imigrantes japoneses (IBGE), com números que ultrapassaram 250 mil entre os anos de 1940 e 1950.

Figura 15: Classificação das amostras do Rio de Janeiro segundo a ancestralidade dos haplogrupos.



Fonte: Este trabalho.

Esta classificação confirmou a proximidade entre as duas populações estudadas, bem como as aproximou geneticamente das populações europeias, tal como observado no cálculo de distância genética. Os dados referentes à ancestralidade dos haplogrupos encontrados em nossa população, juntamente com a comparação realizada entre diversas populações, se mostram em acordo com a história da formação da população brasileira resultante de cinco séculos de misturas inter-étnica entre, principalmente, nativo americanos, europeus e africanos (PENA *et al*., 2009).

6. CONCLUSÕES

- Este estudo analisou taxas de mutação de 23 Y-STR nas populações de São Paulo e Rio de Janeiro, e comparou-as com dados disponíveis no YHRD, confirmando que essas taxas variam entre os marcadores, sendo necessária a aplicação de taxas de mutação precisas referentes a cada marcador nos cálculos estatísticos. Para isso, há a necessidade de estudos com quantidades volumosas de amostras, o que ainda carece na literatura para alguns marcadores que são utilizados na rotina forense;

- Os dados obtidos neste estudo ajudarão a compreender melhor as ocorrências de mutação nos marcadores Y-STR por meio das características observadas;

- As características das mutações observadas neste trabalho se mostraram semelhantes àquelas de estudos realizados com Y-STRs, confirmando que alelos de maiores comprimentos tendem a sofrer mais mutação que alelos curtos;

- O painel de marcadores utilizado neste estudo mostrou conter alto valor mutacional, podendo haver uma ou mais não coincidências ao avaliar relações de parentesco, e por isso ao utilizá-lo na rotina forense é necessário levar em consideração as frequências alélicas, taxas de mutação, e o grau de parentesco;

- Foram estabelecidas as frequências alélicas e haplotípicas para os 23 marcadores Y-STRs na população brasileira dos Estados de São Paulo e Rio de Janeiro, e a utilização deste painel parece ser uma ferramenta muito interessante para aplicação na rotina de casos forenses da população brasileira, pois esse sistema se mostrou altamente discriminativo nas populações estudadas;

- Por meio da verificação das distâncias genéticas entre as populações, podemos notar a semelhança genética entre algumas populações brasileiras que apresentam característica de miscigenação nas linhagens patrilíneas. Esse fato permite unirmos dados dessas populações disponíveis em bancos de dados para fins de comparação genética com populações estrangeiras, aumentando o número amostral e com isso obter valores cada vez mais precisos;

- A partir da previsão dos haplogrupos, os resultados confirmaram os dados disponíveis na literatura, comprovando a forte influência do cromossomo Y de europeus nas populações estudadas.

7. REFERÊNCIAS

ANTÃO-SOUZA, S. *et al*. Mutation rates and segregation data on 16 Y-STRs: An update to previous GHEP-ISFG studies. **Forensic Science International: Genetic. Suppl. Series,** v. 6, p. e601-e601, 2017.

ATHEY, T. W. Haplogroup Prediction from Y-STR Values Using an Allele-Frequency Approach. **J Genet Geneal**, v.. 1, p. 1-7, 2005.

ATHEY, T.W. Haplogroup Prediction from Y-STR Values Using a Bayesian-Allele Frequency Approach. **J Genet Geneal**, v. 2, p. 34-39, 2006.

BALLANTYNE, K.N. *et al*. Mutability of Y-Chromosomal Microsatellites: Rates, Characteristics, Molecular Bases, and Forensic Implications. **Am J Hum Genet,** v. 87, p. 341-353, 2010.

BALLANTYNE, K.N. *et al*. A new future of forensic Y-chromosome analysis: Rapidly mutating Y-STRs for differentiating male relatives and paternal lineages. **Forensic Sci. Int.: Genet.,** v.6, p. 208-218, 2012.

BARCELOS, R.S.S. *et al*. Male contribution in the constitution of the Brazilian Centro-Oeste population estimated by Y-chromosome binary markers. **Int Congr Ser**., v.1288, p. 228–230, 2006.

BEIGUELMAN, B. Dinâmica dos genes nas famílias e nas populações. 2ª. ed. Ribeirao Preto: Sociedade Brasileira de Genetica, 1996.

BERNARDO, S. *et al*. MtDNA ancestry of Rio de Janeiro population, Brazil. **Mol.Biol. Rep.,** v. 41, p. 1945-1950, 2014.

BODNER M. *et al*. Recommendations of the DNA Commission of the International Society for Forensic Genetics (ISFG) on quality control of autosomal Short Tandem Repe*et al*lele frequency databasing (STRidER). **Forensic Sci. Int. Genet**, v. 24, p. 97-102, 2016.

BOSCH, E. *et al*. Variation in Short Tandem Repeats Is Deeply Structured by Genetic Background on the Human Y Chromosome. **Am. J. Hum. Genet**, v. 65, p. 1623-1838, 1999.

BRINKMANN, B. *et al.* Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. **Am J Hum Genet,** v. 62, p. 1408-1415, 1998.

BRAGANHOLI, D.F. **Análise de polimorfismos INDELs na Identificação Humana.** Tese (Doutorado). Universidade Estadual Júlio de Mesquita Filho, 2016.

BRENNER, C.H. Understanding Y haplotype matching probability. **Forensic Sci Int. Genet.,** v. 8, p. 233-243, 2013.

BRIÓN, M. *et al*. Introduction of an single nucleotide polymorphism-based “Major Y-chromosome haplogroup typing kit” suitable for predicting the geographical origin of male lineages. **Electrophoresis,** v.26, p.4411-4420, 2005.

BUENO, B.P.S. Dilatação dos confins: caminhos, vilas e cidades na formação da Capitania de São Paulo (1532-1822). **An. Mus. Paul,** v. 17, p. 251-94, 2009.

BUTLER, J.M. Forensic DNA Typing: Biology, Technology, and Genetics of STR Markers. 2ª ed. Nova Iorque: Elsevier Academic Press. 2005. 688p.

BUTLER, J.M. Short tandem repeat typing technologies used in human identity testing. **Bio Techniques**, v. 43, p. II-V, 2007.

BUTLER, J.M. Advanced Topics in Forensic DNA Typing: Methodology. 1ª ed. Nova Iorque: Elsevier Academic Press. 2011. 704p

CALABREZ, M.C.T. **Influência do calor na análise de DNA extraído de sangue e tecidos humanos: importância para a identificação de corpos carbonizados.** 1999. 92f. Tese (Doutorado) – Faculdade de Medicina da Universidade de São Paulo, 1999.

CARRACEDO, A, SÁNCHEZ-DIZ, P. Forensic genetics: from classical serological genetic markers to DNA polymorphis analysed by microarray technology Em: Handbook of analytical separations, 1ª ed. Amsterdã: Elsevier, Amsterdam. 2000. Vol. 2

CLIFFORD, R. *et al.* Expression-based genetic physical maps of single nucleotide polymorphisms identified by the cancer genome anatomy project. **Genome Research**, v. 10, p. 1259-1265, 2000.

CHUNG, Y.K.; FUNG, W.K. Identifying contributors of two-person DNA mixtures by familial database search. **Int J Legal Med.**, v. 127, p. 25-33, 2013.

CORACH, D.; PENACINO, G.; SALA, A. Marcadores del cromosoma Y en la identificación de rastros de interés forense y en el rastreo de línea paterna em la prueba del ADN en medicina Forense. Barcelona: Masson S.A. 1999. Cap 6.

CORACH, D. *et al*. Inferring Continental Ancestry of Argentineans from Autosomal, Y-Chromosomal and Mitochondrial DNA. **Ann. Hum. Genet.,** v. 74: p. 65-76, 2010.

DA FRÉ N.N. *et al*. Genetic data and de novo mutation rates in father-son pairs of 23 Y-STR loci in Southern Brazil population. **Int. J. Legal Med.**, v. 129: p. 1221-1223, 2015.

DALTON G.C. **Análise de diferenciação intra-populacional em amostras de diferentes estratos sócio-econômicos do Distrito Federal**. Tese (Doutorado em Biologia Animal) - Universidade de Brasília, 2010.

DAWSON, E. *et al*. A SNP resource for human chromosome 22: extracting dense clusters of SNPs from the genomic sequence. **Genome Research,** v. 11, p. 170-178, 2001.

DE ANDRADE, E.S. *et al*. Mutation rate at 14 STR loci in the population from Pernambuco, Northeast Brazil. **Forensic Sci. Int. Genet.,** v. 3, p. 141-143, 2009.

DE AZEVEDO D.A. *et al*. Analysis of Y chromosome SNPs in Alagoas, Northeastern Brazil. **Forensic Sci Int Genet Suppl Ser,** v. 2, p. 421–422, 2009.

DECKER, A.E. *et al*. Analysis of mutations in father–son pairs with 17 Y-STR loci. **Forensic Sci Int. Genet.,** v. 2, p. e31-e35, 2008.

EXCOFFIER, L. Arlequim 3.5 An integrated software package for population Genetics Data analysis, user manual. http://cmpg.unibe.ch/software/arlequin3, 2005.

FAN, H,; CHU, J.Y. A brief review of short tandem repeat mutation. **Genomics Proteomics Bioinformatics,** v. 5, p. 7-14, 2007.

FERREIRA, S. T. G. *et al.* Routine analysis of sexual assault cases in Brasília, Brazil, using 23 Y chromosomal markers. **Forensic Sci Int Genet Suppl Ser.,** v. 5, p. e619-e621, 2015.

FIGUEIREDO, R.F. *et al*. Male-specific contributions to the Brazilian population of Espirito Santo. **Int. J. Legal Med,** v. 3, p. 679-681, 2015.

FRANÇA, G.V. Medicina Legal. 6. Ed. Rio de Janeiro: Guanabara-Koogan. 2011. p.32-63. Cap 3.

FRANCEZ P.A.C. *et al*. Haplotype diversity of 17 Y-STR loci in an admixed population from the Brazilian Amazon. **Genet. Mol. Biol,** v. 35, p. 45-52, 2012.

GJERTSON, D.W. *et al.* ISFG: Recommendations on biostatistics in paternity testing. **Forensic Sci Int Genet.,** v. 1, p.223-231, 2007.

GODINHO, N.M.O. *et al*. Regional patterns of genetic admixture in South America. **Forensic Sci. Int.**, v. 1, p. 329-330, 2008.

GÓES, A.C.S. *et al*. Population and mutation analysis of 17 Y-STR loci from Rio de Janeiro (Brazil). **Int. J. Legal Med**., v. 119, p. 70-76, 2005.

GUSMÃO L. *et al*. Grouping of Y-STR haplotypes discloses European geographic clines. **Forensic. Sci. Int.,** v. 134, p. 172–179, 2003.

GUSMÃO, L. *et al*. Mutation rates at Y chromosome specific microsatellites. **Human Mutat**., v. 26, p. 529-528, 2005.

GUSMÃO, L.; ALVEZ, C. Marcadores dos cromossomas sexuais. Relevância na identificação Genética. Em Pinheiro, M.F.T. Genética Forense Perspectivas da identificação Genética. 1ª Ed. Porto: Edições Universidade Fernando Pessoa, 2010.

GUSMÃO L.; BRIÓN M.; GONZÁLEZ-NEIRA, A. The human Y-chromosome. Male Specific polymorphism and forensic genetics. M.J. Bogus (Ed) Forensic Science handbook of analytical separations. 2000. V. 2.

HOLMES, C.L., RUSSELL, J.A., WALLEY, K.R. Genetic polymorphisms in sepsis and septic shock: role in prognosis and potential for therapy. **Chest,** v. 124, p. 1103-15. 2003.

IBGE. Instituto Brasileiro de Geografia e Estatística. Censo Demográfico de 2010. Rio de Janeiro: IBGE; 2010.

IBGE. Brasil 500 anos de povoamento. Rio de Janeiro: IBGE, 2000.

JAKOBSSON, M. *et al*. Genotype, haplotype and copy-number variation in worldwide human populations. **Nature,** v. 451, p. 998-1003, 2008.

JARRETA, M.B.M. La prueba del ADN en medicina Forense. Barcelona: Masson S.A. 1999.

JEFFREYS, A.J.; WILSON, V.; THEIN, S.L. Hypervariable “minisatellite” regions in human DNA. **Nature**, v. 314, p. 67-73, 1985.

JEFFREYS, A.J.; BROOKFIELD, J.F.Y.; SEMEONOFF, R. Positive identification of an immigration test-case using human DNA fingerprintings. **Nature,** v.317, p. 818-819, 1986b.

JOBLING, M. A., GILL, P. Encoded evidence: DNA in forensic analysis. **Nature Reviews Genetics,** v. 5, p. 739-751, 2004.

JORDE, L.B.; WATKINS, W.S.; BAMSHAD, M.J. Population genomics: a bridge from evolutionary history to genetic medicine. **Human Mol. Genetics**, v.10, p.2199-2207, 2001.

KARKI, R. *et al*. Defining “mutation” and “polymorphism” in the era of personal genomics. **BMC Med Genomics**. v. 8. p. 37, 2015.

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. of Hum. Genet.,**. v. 66, p. 1580-1588, 2000.

KAYSER, M. *et al*. A comprehensive survey of human Y-chromosomal microsatellites. **Am. J. Hum. Genet. 74**, 1183–1197, 2004.

KAYSER, M.; de KNIJFF, P. Improving human forensics through advances in genetics genomics and molecular biology. **Nat Rev Genet.,** v. 12, p. 179-192, 2011.

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

KAYSHYAP, V.K. *et al*. DNA profiling technologies in forensic analysis. **International Journal of Human Genetics**, v.4, p. 11-30, 2004.

KRAWCZAK, M. Forensic evaluation of Y-STR haplotype matches: a comment. **Forensic Sci Int,** v. 118, p. 114-115, 2001.

LIN, S.W.; LI, C.; IP, S.C.Y. A selection guide for the new generation 6-dye DNA profiling systems. **Forensic Sci. Int. Genetic**, v. 30, p.34-42, 2017.

MARTÍNEZ, H. *et al*. Admixture Estimates for Caracas, Venezuela, Basedon Autosomal, Y-Chromosome, and mtDNA Markers. **Human Biol.,** v.79, n.2, p.201-213, 2007.

MARTINEZ, J. *et al.* Mutation rates for 20 STR loci in a population from São Paulo state, Southeast, Brazil. **Ann. Hum. Biol.,** v. 44, p. 659-662, 2017.

MILLS, R. E. *et al*. An initial map of insertion and deletion (InDel) variation in the human genome. **Genome Res**. v. 16, p. 1182-90, 2006.

MONTEIRO, J.M. Negros da terra: índios e bandeirantes nas origens de São Paulo. 2ª reimpressão. São Paulo: Companhia das Letras. 1994. 300p.

MONTEIRO, E. H. G. **Desenvolvimento e aplicação de polimorfismos de inserção/deleção do cromossomo X em genética forense**. 2007. 162 f. Dissertação (Mestrado em Ciências Genômicas e Biotecnologia) - Universidade Católica de Brasília, Brasília, DF, 2007.

MOTA, C.G. São Paulo: exercício de memória. **Estud. Av.,** v. 17, n. 48, p. 241-63. 2003.

MOURA, R.R. *et al*. Meta-analysis of Brazilian genetic admixture and comparison with other Latin America countries. **Am. J. Hum. Biol.,** v. 27, p. 670-680, 2015.

MULERO, J.J. *et al*. Development and validation of the AmpFlSTR Yfiler PCR amplification kit: a male specific, single amplification 17 Y-STR multiplex system. **Forensic Sci. Int. Genet.,** v. 51, p. 64-75, 2006.

MULLIS, K.B., FALOONA, F.A. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. **Methods Enzymol**, 155:335-350, 1987.

NEI, M. Analysis of Gene Diversity in subdivides populations. **Procedures Natural Academy Sciences**, v. 70, p. 3321-3323, 1973

NIEDERSTÄTTER, H. *et al*. Separate analysis of DYS385a and b vesus conventional DYS385 typing: is there forensic relevance?. **Int. J. Legal. Med.,** v. 119, p. 1-9, 2004.

NUÑEZ, C. *et al*. Highly discriminatory capacity of the PowerPlex1 Y23 System for the study of isolated populations. **Forensic Sci. Int.: Genet,** v. 17, p.104-107, 2015.

OH, N.W. *et al*. Haplotype and mutation analysis for newly suggested Y-STRs in Korean father–son pairs. **Forensic Sci Int: Genet,** v. 15, p. 64-68, 2015.

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

PALHA, T. *et al*. Disclosing the Genetic Structure of Brazil through Analysis of Male Lineages with Highly Discriminating Haplotypes. **Plos One**, v. 7, p. 1-8, 2012.

PARRA, F.C. *et al*. Color and genomic ancestry in Brazilians. **Proc. Natl. Acad. Sci, USA,** v. 100, p. 177-182, 2003.

PASCALLI, VL; DOBOSZ, M.; BRINKIMANN. Coordinating Y- chromosomal STR research of the Courts. **Int. J. Legal Med**, v. 112, p.1, 1999.

PENA, S.D.J. *et al*. Retrato Molecular do Brasil. **Ciência Hoje,** v. 159, p. 16-25, 2000.

PENA, S.D. *et al.* DNA tests probe the genomic ancestry of Brazilians. Braz. **J. Med. Biol. Res**., v.42, p.870-876, 2009.

PEREIRA, R.; GUSMÃO, L. Capillary electrophoresis of 38 noncoding biallelic mini-Indels for degraded samples and as complementary tool in paternity testing. **Methods Mol. Biol.,** v. 830. p. 141- 57, 2012.

PINHEIRO, M.F. *et al*. Allele frequencies of sixteen STRs in the populations of nothern Portugal. **Forensic Sci. Int.**, v. 148, p. 221-223, 2005.

PINTO, N.; GUSMÃO, L.; AMORIM, A. Mutation and mutation rates at Y chromosome specific Short Tandem Repeat Polymorphisms (STRs): A reappraisal. **Forensic Sci. Int.: Genet.,** v. 9, p. 20-24, 2014

PONTES, M.L. *et al*. Allele frequencies and population data for 17 Y-STR loci (AmpFℓSTR® Y-filer™) in a Northern Portuguese population sample. **Forensic Sci Int,** v. 1, p. 61-67, 2007.

PURPS, J. *et al.* A global analysis of Y-chromosomal haplotype diversity for 23 STR loci. **Forensic Sci Int**., v. 12, p. 12-23, 2014.

QUINTANA-MURCI, L.; KRAUSZ, C.; MC ELREAVY, K. The human Y chromosome: function, evolution and disease. **Forensic Sci. Int.,** v. 118, p. 169-181, 2001.

RESQUE, R. *et al*. Male Lineages in Brazil: Intercontinental Admixture and Stratification of the European Background. **Plos One,** v. 11, p. 1-17, 2016.

RIBEIRO, D. O povo brasileiro: A formação e o sentido do Brasil. São Paulo: Cia das Letras. 1995. 480p.

RIBEIRO, J. *et al*. Analysis of 23 Y-STRs in a population sample from eastern Paraguay. **Forensic Sci. Int. Genet.,** v.37, p. e20-e22, 2018.

ROEWER, L. et al. Simple repeat sequences on the human Y chromosome are equally polymorphic as their autosomal counterparts. **Human Genetics,** v.89, p. 389-394, 1993.

ROEWER, L. *et al*. Online reference database of European Y-chromosomal short tandem repeat (STR) haplotypes. **Forensic Sci Int,** v. 118, p. 106-113, 2001.

ROEWER, L. DNA fingerprinting in forensics: past, present, future. **Investigative Genetics**, v. 4, p. 1-10, 2013.

ROEWER, L. WILLUWEIT, S. Y-Chromosomale STR-Analyse in der forensischen Praxis**. Rechtsmedizin,** v. 28, p. 149-64, 2018.

SAFERSTEIN, R. Criminalistics: An introduction to Forensic Science. 9ª ed. New Jersey: Prentice Hall, 2006. 1043p.

SAIKI, R. K. *et al*. Primer-Directed Enzymatic Amplification of DNA with a Thermostable DNA Polymerase. **Science,** v.239, p. 487-491, 1988.

SANCHES, N. M. *et al.* Mitochondrial DNA control region diversity in a population from Espirito Santo State, Brazil. **Molecular Biology Reports**, v. 41, p. 6645-6648, 2014.

SILVA, B.M. *et al*. Allele frequencies of fifteen STR loci in a population from Central Brazil. **Forensic Sci. Int. Genet.,** v. 4, p. 151-152, 2010.

SINGER-SAM, J.; TANGUAY, R.; RIGGS. A.D. Use of Chelex to improve the PCR signal from a small number of cells. **Amplifications.** v. 3. p. 11, 1989

STANGE, V.S. *et al*. Stratification among Europeans descent and admixed Brazilian populations of Espírito Santo for 27 Y-STRs. **Forensic Sci. Int. Genet.,** v. 41, p. e20-e22, 2019.

SKALETSKY, H. *et al*. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. **Nature**, v. 423, p. 825–37, 2003.

SOBRINO, B.; BRION, M.; CARRACEDO, A. SNPs in forensic genetics: a review on SNP typing methodologies. **Forensic Sci. Int.,** v. 154, p. 181-194, 2005.

TAKEZAKI, N., NEI, M. Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. **Genetics,** v. 144, p. 389-99, 1996.

TARAZONA-SANTOS, E. *et al*. Genetic Differentiation in South Amerindians Is Related to Environmental and Cultural Diversity: Evidence from the Y Chromosome. **Am. J. Hum. Genet.**, v. 68, p. 1485-1496, 2001.

THOMSEN, A.R. *et al*. A report of the 2002-2008 paternity testing workshops of the English speaking working group of the International Society for Forensic Genetics. **Forensic Sci. Int. Genet.**, v. 3, p. 214- 21, 2009.

TORRES, S.R.R. *et al*. Population genetic data and forensic parameters of 30 autosomal InDel markers in Santa Catarina State population, Southern Brazil. **Mol Biol Rep.**, v. 41, p. 5429-5433, 2014.

TOSCANINI, U. *et al*. The geographic mosaic of Ecuadorian Y-chromosome ancestry. **Forensic Sci. Int. Genet**., v. 33, p. 59-65, 2018.

TURRINA, S. *et al*. Haplotype data and mutation rates for the 23 Y-STR loci of PowerPlex®Y 23 System in a Northeast Italian population sample. **Int. J. Legal Med.**, v. 129, p. 725-728, 2015.

UNDERHILL, P.A.; KIVISILD, T. Use of Y Chromosome and Mitochondrial DNA Population Structure in Tracing Human Migrations. **Annu. Rev. Genet**., v.41, p.539-564, 2007.

VELDMAN, A. **Evidential strength of Y-STR haplotype matches in forensic DNA casework**. Tese (mestrado em Matemática Forense), Nederlands, Universiteit Leiden, Leiden. 44pp, 2007.

VIEIRA-SILVA, C. *et al*. Y-STR mutational rates determination in South Portugal Caucasian population. **Forensic Sci. Int.: Genet. Suppl. Series**, v. 2, p. 60-61, 2009.

VIEIRA, T.C. *et al*. Y-STR haplotype diversity and population data for Central Brazil: implications for environmental forensics and paternity testing. **Genetics and Molecular Research**, v. 13, p. 3404-3410, 2014.

WATKINS, W.S. *et al*. Genetic variation in South Indian castes: evidence from Y-chromosome, mitochondrial and autosomal polymorphisms. **BMC Genet.**, v. 9, p. 86, 2008.

WEBER, J.L. *et al*. Human diallelic insertion/deletion polymorphisms. **Am. J. Hum. Genet.,** v. 71, p. 854-62, 2002.

WEEDEN, V. W.; SWARNEN, S. L. Exames forenses de identificação por análises do DNA. Em: HENRY, J. B. Diagnósticos clínicos e tratamento por métodos laboratoriais. 19. ed. São Paulo: Manole, 1998. Cap. 63, p. 1427-1438.

WEI, W. *et al*. A comparison of Y-chromosomal lineage dating using either resequencing or Y-SNP plus Y-SNP genotyping. **Forensic Sci. Int.: Genetics,** v. 7, p. 568-572, 2013.

XUE, Y. *et al*. Human Y Chromosome Base-Substitution Mutation Rate Measured by Direct Sequencing in a Deep-Rooting Pedigree. **Current Biology**, v. 19, p. 1453-1457, 2009.

YANG, N. *et al*. Examination of ancestry and ethnic affiliation using highly informative diallelic DNA markers: application to diverse and admixed populations and implications for clinical epidemiology and forensic medicine. **Human Genetics**, v. 118, p. 382-392, 2005.

ZEGURA, S.L *et al.*. High resolution SNPs and microsatellite haplotypes point to a single, recent entry of Native American Y chromosomes into the Americas. **Mol. Biol. Evol.,** v.21, p.164-175, 2004.

ZIDKOVA, A. *et al*. Application of the new insertion– deletion polymorphism kit for forensic identification and parentage testing on the Czech population**. Int J Legal Med**, v. 127, p. 7-10, 2011.

ANEXO 1

Int J Legal Med
DOI 10.1007/s00414-015-1214-2

POPULATION DATA

Male-specific contributions to the Brazilian population of Espirito Santo

Raquel de F. Figueiredo[1] · Isabela B. Ambrosio[1] · Danilo F. Braganholi[1] · Gustavo Chemale[2] · Joyce A. Martins[1] · Veronica Gomes[3] · Leonor Gusmão[3,4] · Regina M. B. Cicarelli[1]

Received: 29 January 2015 / Accepted: 5 June 2015
© Springer-Verlag Berlin Heidelberg 2015

Abstract Y chromosome markers have been widely studied due to their various applications in the fields of forensic and evolutionary genetics. In this study, 35 Y-SNPs and 17 Y-STRs were genotyped in 253 males from the State of Espirito Santo, Brazil. A total of 18 haplogroups and 243 haplotypes were detected; the haplogroup and haplotype diversities were 0.7794 and 0.9997, respectively. Genetic distance analysis using the Y-STR data showed no statistically significant differences between Espirito Santo and other admixed populations from Brazil. The classification of paternal lineages based on haplogroups showed a predominant European contribution (85.88 %), followed by African (11.37 %) and Amerindian (2.75 %) contributions.

Raquel de F. Figueiredo, Isabela B. Ambrosio and Danilo F. Braganholi contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s00414-015-1214-2) contains supplementary material, which is available to authorized users.

✉ Regina M. B. Cicarelli
cicarelli@fcfar.unesp.br

[1] Laboratório de Investigação de Paternidade, Núcleo de Atendimento a Comunidade - Faculdade de Ciências Farmacêuticas, São Paulo State University (UNESP), Rodovia Araraquara-Jaú, km 01, 14801-902 Araraquara, São Paulo, Brazil

[2] Laboratório de Genética Forense, Departamento de Polícia Federal, Instituto Nacional de Criminalística, Diretoria Técnico-Científica, Brasília, Distrito Federal, Brazil

[3] IPATIMUP, Institute of Molecular Pathology and Immunology of the University of Porto, 4200-465 Porto, Portugal

[4] DNA Diagnostic Laboratory (LDD), State University of Rio de Janeiro (UERJ), Rio de Janeiro, Brazil

Published online: 16 June 2015

Keywords Y chromosome · Y-SNP haplogroup · Y-STR haplotype · Brazil

Brazil has a large and heterogeneous population originated from the admixture of Native Americans, Europeans, and Africans. As in most South American countries, the process of admixture in Brazil was very diverse depending on the country region, which created a highly sub-structured population. Therefore, the definition of the limits of a database to be used for forensic purposes is a complex task and implies the need for a full comprehension of all genetic variation that exists across the country.

Many studies of Y chromosome distribution in the diverse regions of the country [1–5] have been performed to fill some gaps in knowledge about the demographic events involved in the formation of Brazilian populations. Nevertheless, none of these studies have focused on the population of Espirito Santo State, making the present work the first report of Y-STRs combined with Y-SNPs in this population.

Blood samples were collected from 253 unrelated males from the State of Espirito Santo, located in the southeastern Brazil.

All samples were typed to 35 Y-SNP markers (see Fig. S1 for markers information and details on the methods used), as well as to a set of 17 Y-STR loci included in the Y-Filer (Life Technologies), following the manual instructions.

This project was approved by the Ethical Committee of FCFAr-UNESP (n. 12/2010) and all participants have signed an informed consent.

The data were deposited in the Y Chromosome Haplotype Reference Database—YHRD (http://www.yhrd.org; [6]) under the accession number YA003849.

Haplotypes and haplogroup diversities and population pairwise genetic distances (F_{ST}) were calculated using ARLEQUIN software version 3.1 [7].

The haplotype data for the 17 Y-STR loci in the population of the State of Espirito Santo (Brazil) are presented in Table S1. Besides those usually present at DYS385, three rare cases of duplicated alleles were detected at DYS19 locus.

A high haplotype diversity was estimated (99.97 %); for a total of 243 different haplotypes identified in our sample, 234 of which being unique and 9 observed more than once in our sample (Table S2). The most frequent haplotype in our sample, present in three individuals from haplogroup R1b1-P25 (ID 13, ID 98, and ID 140), contains the core set of alleles that was previously described having a decreasing W–E gradient in Europe, with its highest frequency in the Iberian Peninsula [8].

Pairwise F_{ST} genetic distance analysis performed between admixed population samples from the five geopolitical states forming the Southeastern region in Brazil [3, 5] showed non-differentiation p values between Espirito Santo and São Paulo (0.00018), Minas Gerais (0.00014) and Rio de Janeiro (0.00024), for a significance level of 0.0083, obtained after applying the Bonferroni's correction. Therefore, the Espirito Santo sample was added to the Southeastern region and compared with populations from other Brazilian regions, accessible through the YHRD [6]: Northern (YP000574, n=400), Northeastern (YP000663, n=205), Central-Western (YP000793, n=445 and YP000754, n=50), and Southern (YP000759, n=243 and YP000757, n=77). The results showed no significant differentiation for Y-STR haplotype data between Southeastern and all the other regions (Table S3).

The frequencies of the 18 Y-SNP haplogroups identified in the 253 Brazilian male individuals in this study (Table S1) are presented in Fig. S1. As expected for an admixed population, high haplogroup diversity was observed (0.7794± 0.0229). This value was greater than those observed in Brazilian populations from Alagoas [2] and Rio de Janeiro [1, 5].

Among the European haplogroups, R1b1-P25 was the most frequently detected haplogroup in our sample, followed by haplogroup I-M170*(xM26) (Fig. S1). These haplogroups, together with lineages within paragroups I-M170 and J-12f2, are the most common in Europe and in the Brazilian populations from Rio de Janeiro [1, 5] and Alagoas [2]. Eight samples were classified in the haplogroup R1-M173. This European haplogroup was identified for the first time in a Brazilian population.

The African haplogroups E1a-M33, E2-M85, E1b1a-M191, and E1b1a-M154 were also detected in our sample (Fig. S1). One sample was classified as Y-A4*, the most basal clade of the Y chromosome phylogenetic tree, which is almost restricted to the African continent.

A low frequency of the main Amerindian haplogroup Q1a2-M3 was detected in this study (Fig. S1; Table S1).

Based on the detected Y-SNP haplogroups, we conclude that Europe is the main source of paternal lineages of the present-day population in the State of Espirito Santo (85.88 %). African and Amerindian sources only account for 11.37 and 2.75 % of the lineages, respectively.

Haplogroup frequencies were used to calculate population pairwise genetic distances (F_{ST}) between the studied sample and other 15 populations from Rio de Janeiro, Brazil [1], Alagoas, Brazil [2], Colombia [9], Argentina [9], Spain [9], Portugal [8], Germany [9], Denmark [9], Turkey [9], Mozambique [9], Somalia [9], Japan [9], Thailand [9], China [9], and Greenland [9] (Supplementary Table S4). The results were represented in a MDS plot (Fig. S2). This figure shows that the population from Espirito Santo is more similar to other Brazilian populations, and to populations from Iberian Peninsula (Spain and Portugal) and Latin America. South American populations tend to cluster together with European populations, which is in accordance with the known sex-biased formation of these populations, in which Europeans has the greatest contribution to paternal lineages.

Acknowledgments The authors wish to thank the "Centro de Hematologia e Hemoterapia do Espírito Santo" for assistance in sample collection. This study was supported by the PADC/FCF - UNESP (proc. 2010/24-I) and Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) (proc. 2010/17220-5). RFF is supported by a fellowship from the "Conselho Nacional de Desenvolvimento Científico e Tecnológico" (CNPq); DFB, IBA, and JAM are supported by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). IPATIMUP is an Associate Laboratory of the Portuguese Ministry of Education and Science and is partially supported by FCT. VG is supported by Fundação para a Ciência e Tecnologia through grant SFRH/BPD/76207/2011.

References

1. Silva DA, Carvalho E, Costa G, Tavares L, Amorim A, Gusmão L (2006) Y-chromosome genetic variation in Rio de Janeiro population. Am J Hum Biol 18:829–837
2. de Azevedo DA, da Silva LAF, Gusmão L, de Carvalho EF (2009) Analysis of Y chromosome SNPs in Alagoas, Northeastern Brazil. Forensic Sci Int Genet Suppl Ser 2:421–422
3. Palha T, Gusmão L, Ribeiro-Rodrigues E, Guerreiro JF, Ribeiro-Dos-Santos A, Santos S (2012) Disclosing the genetic structure of Brazil through analysis of male lineages with highly discriminating haplotypes. PLoS One 7:e40007
4. Chemale G, Freitas JM, Badanaco JL, Rosa LD, Martins JA, Martins AD et al (2013) Y-Chromosomal STR haplotypes in a sample from Brasília, Federal District, Brazil. Forensic Sci Int Genet. doi:10.1016/j.fsigen.2013.06.005
5. Oliveira AM, Domingues PM, Gomes V, Amorim A, Jamauzzi I, Carvalho EF, Gusmão L (2014) Male lineage strata of Brazilian population disclosed by the simultaneous analysis of STRs and SNPs. Forensic Sci Int: Genet 13:264–268
6. Willuweit S, Roewer L (2015) The new Y chromosome haplotype reference database. Forensic Sci Int Genet 15:43–48

7. Excoffier L, Laval G, Schneider S (2005) Arlequin v. 3.0: an integrated software package for population genetics data analysis. Evol Bioinform Online 1:47–50
8. Gusmão L, Sánchez-Diz P, Alves C, Belesa S, Lopes A, Carracedo A, Amorim A (2003) Grouping of Y-STR haplotypes discloses European geographic clines. Forensic Sci Int 134:172–179
9. Brión M, Sanchez JJ, Balogh K, Thacker C, Blanco-Verea A, Borsting C et al (2005) Introduction of an single nucleotide polymorphism-based "Major Y-chromosome haplogroup typing kit" suitable for predicting the geographical origin of male lineages. Electrophoresis 26: 4411–4420

Forensic Science International: Genetics Supplement Series 6 (2017) e296–e297

Contents lists available at ScienceDirect

Forensic Science International: Genetics Supplement Series

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

Allele frequencies of three forensic STR markers (D22S1045, D2S441 and D10S1248) in the population from São Paulo, Brazil

I.B. Ambrosio*, D.F. Braganholi, J. Martinez, F.S. Polverari, R.M.B. Cicarelli

Laboratory of Paternity Investigation, São Paulo State University, Araraquara, Brazil

ARTICLE INFO

Keywords:
STRs
D22S1045
D2S441
D10S1248
São Paulo State
Brazil

ABSTRACT

São Paulo state is one of the Brazilian states with the highest cultural and ethnic miscegenation, mainly of Italian and Portuguese descendants, but also strongly influenced by Amerindians, Africans, Spanish, Germans, Arabs and Japanese, as well as the high number of immigrants from the northeast region. In this work we analyzed statistical data of three STRs markers (D22S1045, D2S441 and D10S1248) that are used in our laboratory's forensic routine for paternity investigations. The use of these data will be extremely important for the laboratory to perform statistical analyzes for the conclusion of the reports. We analyzed 221 samples obtained from unrelated individuals born in the São Paulo state, which the allele frequencies and statistical parameters were estimated with PowerStats version 1.2 (Promega Corp.). The power of discrimination (PD) and power of exclusion (PE) for the D10S1248 marker were 0.914 and 0.576 respectively and the allele 14 showed the highest frequency (0.294) and 8, 10, 18 and 19 alleles presented the lowest frequency (0.002). PD and PE of the D2S441 marker were 0.907 and 0.551 respectively; the highest frequency allele was 11 (0.312) and the lowest frequency alleles were 9 and 12.3 (0.002). The marker D22S1045 showed PD 0.890 and PE 0.489 with a higher frequency allele 16 (0.373) and lowest frequency allele 8 (0.002). Also, significant differences were found between the São Paulo population and the other populations for the three markers analyzed.

1. Introduction

Short Tandem Repeat (STR) polymorphisms are widely used in forensic analysis. Common cases of paternity are analyzed by commercially available multiplex systems, however, for more complex cases, such as complex kinship analysis, additional STRs are required for best results [1].

In order to increase the amount of frequency data and genetic parameters of the Brazilian population, three STRs (D22S1045, D2S441 and D10S1248) were analyzed and used in the forensic routine of our laboratory for paternity investigation. The majority individuals of these cases come from the São Paulo' state, Brazil, which is considered the Brazilian state with the most cultural and ethnic miscegenation, mainly comprising of Italians and Portuguese descendants, but also with strong influence of Amerindians, Africans, Spanish, Germans, Arabs and Japanese populations, as well as a high number of immigrants from the northeast region [2]. Considering this particularity, the use of regional frequency data will be extremely important for the laboratory, facilitating the conclusion of the reports.

2. Materials and methods

A total of 221 samples obtained from unrelated individuals born in the São Paulo state were collected on FTA® classic cards (GE Healthcare, Life Sciences). Samples were amplified with the PowerPlex® Fusion 6C System (Promega), following the User Guide [3]. Capillary electrophoresis and detection were performed on the 3500 Genetic Analyzer using POP-4® polymer (Applied Biosystems). Genotypes were assigned using the GeneMapper ID v1.2.1 software (Applied Biosystems). Allele frequencies and statistical parameters were estimated with PowerStats version 1.2 (Promega Corp.) [4]. Allelic frequencies for each locus in São Paulo (Southeast Brazil) population were compared with other populations: European, African [5], Colombian region [6] and the population of the Rio Grande do Sul' state (Southern Brazil) [7]. Pairwise genetic distances were calculated between populations using Arlequin v3.5 [8] software. Our laboratory successfully participated in proficiency testing provided by the GEP-ISFG Working Group.

* Corresponding author.
E-mail address: isabrunelli@hotmail.com (I.B. Ambrosio).

http://dx.doi.org/10.1016/j.fsigss.2017.09.135
Received 2 September 2017; Accepted 19 September 2017
Available online 23 September 2017
1875-1768/ © 2017 Elsevier B.V. All rights reserved.

Table 1
Allele frequencies and forensic parameters of the three STR loci for São Paulo population, Southeast Brazil.

Allele	D22S1045	D2S441	D10S1248
8	0,0023		0,0023
9		0,0023	
10	0,0113	0,1742	0,0023
11	0,1176	0,3122	0,0113
11.3		0,0566	
12	0,0158	0,0814	0,0588
12.3		0,0023	
13	0,0068	0,0362	0,2579
14	0,0362	0,3009	0,2941
15	0,2986	0,0294	0,1968
16	0,3733	0,0045	0,1471
17	0,1267		0,0249
18	0,0113		0,0023
19			0,0023
N	442	442	442
HO	0,262	0,226	0,213
HE	0,738	0,774	0,787
PIC	0,700	0,735	0,749
PD	0,890	0,907	0,914
PE	0,489	0,551	0,576

N: number of chromosomes; Ho: observed heterozygosity; He: expected heterozygosity; PIC: polymorphism information content; PD: power of discrimination; PE: power of exclusion.

Table 2
Comparison of the allele frequencies for D2S441, D10S1248 and D22S1045 loci between the São Paulo State population and other populations.

Microsatellite markers	Southern Brazil × São Paulo	Europe × São Paulo	África × São Paulo	Colombian × São Paulo
D2S441	0.19992	0.17631	0.17925	–
D10S1248	0.00982	0.13288	0.10796	0.12997
D22S1045	0.13122	0.12334	0.10917	0.15853

– no data available.

3. Results and discussion

The power of discrimination (PD) and power of exclusion (PE) for the marker D10S1248 were 0.914 and 0.576 respectively; the highest frequency allele was 14 (0.294) and the lowest frequency were 8, 10, 18 and 19 (0.002) (Table 1). PD and PE of the marker D2S441 were 0.907 and 0.551, respectively; the allele with the highest frequency was 11 (0.312) and the lowest frequency were 9 and 12.3 (0.002). The D22S1045 marker showed PD 0.890 and PE 0,489, highest frequency in 16 allele (0.373) and lowest allele 8 (0.002).

The analysis of genetic distance showed significant differences between São Paulo population from Europe, Africa, Colombia and Southern Brazilian region. The marker that presented the greatest genetic distance among the populations was D2S441 (Table 2).

4. Conclusion

These 3 STRs loci increase the discriminatory power of forensic analysis and are very useful for forensic purposes. Among the analyzed, the marker with the greatest power of discrimination for the population of São Paulo was D10S1248.

Conflict of interest statement

None.

Acknowledgements

The authors would like to thank Public Defender Service in São Paulo State by agreement signed with UNESP/Laboratory of Paternity. IBA is supported by fellowship garanted by Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq.

References

[1] V. Cortellini, N. Cerri, A. Verzeletti, Genetic variation at 5 new autosomal short tandem repeat markers (D10S1248, D22S1045 D2S441, D1S1656, D12S391) in a population-based sample from Maghreb region, Popul. Genet. 52 (2011) 368–371.
[2] IBGE. Instituto Brasileiro de Geografia e Estatística, Censo Demográfico de 2010, IBGE, Rio de Janeiro, 2010.
[3] M.G. Ensenberger, K.A. Lenz, L.K. Matthies, et al., Developmental validation of the PowerPlex® Fusion 6C System, Forensic Sci. Int. Genet. 21 (2016) 134–144.
[4] PowerStats version 1.2. Promega Corporation. Available on: http://www.promega.com/geneticidtools/powerstats/.
[5] C. Phillips, L. Fernandez-Formoso, M. Garcia-Magariños, et al., Analysis of global variability in 15 established and 5 new European Standard Set (ESS) STRs using the CEPH human genome diversity panel, Forensic Sci. Int. Genet. 5 (2011) 155–169.
[6] G. Burgosa, L. Fernandez-Formoso, M. Garcia-Magariños, et al., Allelic frequencies and forensic parameters for miniSTRs D10S1248, D14S1434 and D22S1045 (NCD1) in a sample from Central Andean Colombian region, Forensic Sci. Int. Genet. Suppl. Ser. 5 (2015) e81–e82.
[7] R. Rodenbusch, S. Schumacher, A.C. Mardini, et al., Allele frequencies of the five new generation forensic STR (D1S1656, D2S441, D10S1248, D12S391 and D22S1045) in the population from Rio Grande do Sul, Southern Brazil, Forensic Sci. Int. Genet. 6 (2012) e55–e57.
[8] L. Excoffier, H. Lischer, Arlequin: an integrated software package for population genetics data analysis (2015).

Forensic Science International: Genetics Supplement Series 6 (2017) e370–e371

Contents lists available at ScienceDirect

Forensic Science International: Genetics Supplement Series

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

500 years later: Understanding the genetic ancestry of the southeast region, brazil

D.F. Braganholi*, I.B. Ambrosio, F.S. Polverari, J. Martinez, R.M.B. Cicarelli

Laboratory of Paternity Investigation, NAC/FCF – UNESP, Araraquara, SP, Brazil

ARTICLE INFO

Keywords:
Ancestry
Southeast region
Brazil
Genetic markers

ABSTRACT

The Southeast region of Brazil is the main economic region, the most populous and with the major cultural and genetic miscegenation of the country. Due to its historical context, it mainly includes descendants of Portuguese and Italians, but also has strong influence of Native americans, Africans, Spaniards and Germans. The influence of Arabs and Asians is very small. In this work, the ancestral proportions of the population of the Southeast region were evaluated through a genetic data compilation with diverse markers (Y-SNPs, AIM-INDELs, mtDNA and X-INDELs) in individuals from the states of São Paulo, Rio de Janeiro, Minas Gerais and Espírito Santo, which form this region. The proportion of paternal lineages (Y-SNPs) with European origin is much higher (86.57%), being 10.86% of African origin and 1.98% native American. Instead, the proportion of African origin (44.1%) predominates in the maternal lineages (mtDNA), followed by Native American with 29.2%, European with 26.2% and only 0.5% for Asian. With the autosomal AIM-INDELs the ancestral proportion obtained was 59.12% European, 28.12% African and 12.76% Native American. The autosomal chromosomes tend to reflect an average between the ancestry proportions of the Y chromosome and mtDNA. No Asian ancestry was identified with these markers. In preliminary analyzes with X-INDELs in samples from São Paulo, we found the following ancestral proportions: 29.4% African, 49.7% European and 20.9% Native American. This genetic miscegenation occurred since the formation of this region, when Portuguese colonizers mated with native women and later with slaves brought from Africa.

1. Introduction

The first inhabitants of the Southeast region of Southeast region were the Native americans belonging to the *macro-jê* and *tupi* groups. From the year 1500, the Portuguese settlers began to arrive [1].

The settlement of Southeastern Brazil began in 1532, with the foundation of the village of São Vicente by the Portuguese Jesuits, supported by the production of sugarcane. From the 17th century, began the phenomenon of "flags", which were expeditions through the interior of Brazil in search of new riches and indigenous to be enslaved [2].

At the end of 17th century, the expeditionary found precious stones in the region currently known of Minas Gerais, beginning the gold cycle.

With mining, the attention of the Portuguese Crown turned to the Southeastern region, considering that sugarcane plantations in the Northeast were in full decay. There has been a great movement of people to the Minas region [2].

In 1763 the capital of the colony is transferred to Rio de Janeiro. In 1808, fleeing from the Napoleonic invasion, a Portuguese Royal Family settled in Rio de Janeiro. The era was marked by several economic changes in the region, with an opening of the ports to friendly nations in 1810 and a rise from Brazil to the United Kingdom of Portugal and the Algarve in 1816 [1].

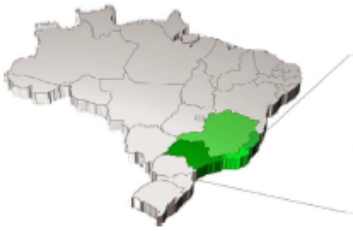
From the 1840s, coffee plantations spread throughout the region and became the basis of the Brazilian economy. It was initially used for slave labor but, with the abolition of slavery in 1888, the lack of labor was filled with the arrival of a large mass of European, mainly Italian, immigrants [1].

Today, the Southeast region is composed of four states: Espírito Santo, Minas Gerais, Rio de Janeiro and São Paulo. It is the most developed region in Brazil, responsible for approximately 60% of the Brazilian Gross Domestic Product (GDP). In it are the most populous municipalities, the highest population density. The Southeast region has a population of approximately 85 million inhabitants, so that 44% of the Brazilian population lives in the Southeast region.

In this work, the ancestral proportions of the population of the Southeast region were evaluated through a genetic data compilation

* Corresponding author.
E-mail address: danilobraganholi@hotmail.com (D.F. Braganholi).

http://dx.doi.org/10.1016/j.fsigss.2017.09.174
Received 22 August 2017; Accepted 19 September 2017
Available online 21 September 2017
1875-1768/ © 2017 Elsevier B.V. All rights reserved.



	EUR	AFR	NAM	ASI
Y-SNPs	85.57	10.86	1.98	NI*
mtDNA	26.2	44.1	29.2	0.5
Uniparental Average	55.9	27.5	15.6	0.25
AIM-INDELs	59.12	28.12	12.76	NI*
X-INDELs	49.7	29.4	20.9	NI*

Fig. 1. In green is illustrated the location of the southeast region in the map of Brazil. The table shows the data (%) of the ancestry analysis obtained with the different markers. EUR: European; AFR: African; NAM: Native American; ASI: Asiatic *NI: Not identified. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

with diverse markers (Y-SNPs, AIM-INDELs, mtDNA and X-INDELs) in individuals from the four states which form this region.

2. Methodology

The compilation was realized from data obtained from several scientific works [4–7] (4, 5, 6 and 7) by the analysis of Y-SNPs, mtDNA control region, AIM-INDELs, and X-INDELs. These markers were analyzed by capillary electrophoresis in samples from about 1000 individuals from the Southeast region.

3. Results and discussion

The coast of São Paulo had already been visited by Portuguese and French navigators in the early 16th century, but its colonization began only in 1532, when Martim Afonso de Souza founded the Vila de São Vicente, located in the present Santos Bay, marking the beginning of the settlement of the Portuguese on the Brazilian coast.

Here it is important to note that there are no records of Portuguese women in Brazil at this time, the Native americans being the first matriarch, which made the predominantly mestizo population of lusos and Native americans (Mameluco). By these characteristics, during the first three centuries of colonization, the number of Native americans and Mameluco surpassed that of Europeans. Since the sixteenth century, there has also been the constant presence of black people, slaves at the time.

The high frequencies of mates among European men with Indigenous and Black women are recognized.

The end of the colonial period and the beginning of the empire in the 18th century was marked by the replacement of sugarcane by coffee plantations and by occurred the mass immigration of Italians, Spaniards, Germans and later Japonese.

All this historical context means that the Southeastern region, as well as the whole of Brazil, presents a great genetic miscegenation as shown in Fig. 1.

4. Conclusion

The genetic data observed are in agreement with the historical knowledge about the formation of the southeastern region of Brazil, where there is great miscegenation. The ancestry of European origin is predominant in the paternal pathway, whereas, Native American and African origins in the maternal pathway are predominant. In general, European ancestry is predominant.

Conflict of interest statement

No one.

References

[1] MOTA, C.G. São Paulo exercício de memória. Estud. av. v. 17, n. 48, p. 241-63. 2003.
[2] BUENO, B.P.S. Dilatação dos confins: caminhos, vilas e cidades na formação da Capitania de São Paulo (1532–1822). An. mus. paul. v. 17, n. 2, p. 251-94. 2009.
[4] R.F. Figueiredo, I.B. Ambrosio, D.F. Braganholi, G. Chemale, J.A. Martins, V. Gomes, L. Gusmão, R.M.B. Cicarelli, Male-specific contributions to the Brazilian population of Espírito Santo, Int. J. Legal Med. (2015), http://dx.doi.org/10.1007/s00414-015-1214-2.
[5] I.B. Ambrosio, D.F. Braganholi, F.S. Polverari, R.M.B. Cicarelli, Mitochondrial DNA 30-SNP data confirm high prevalence of African lineages in the population of Espírito Santo, Brazil, Forensic Sci. Int.: Gene. Suppl. Ser. 5 (2015) e346-e347.
[6] F.S.N. Manta, R. Pereira, R. Vianna, A.R.B. Araújo, D.L.G. Gitaí, D.A. Da Silva, E.V. Wolfgramm, I.M. Pontes, J.I. Aguiar, M.O. Mones, E.F. Carvalho, L. Gusmão, Revisiting the genetic ancestry of brazilians using autosomal AIM-Indels, PLoS One 8 (2012) 2012, http://dx.doi.org/10.1371/journal.pone.0075145.
[7] R. Resque, L. Gusmão, M. Geppert, L. Roewer, T. Palha, L. Alvarez, et al, Male lineages in Brazil: intercontinental admixture and stratification of the european background, PLoS ONE 11 (4) (2016) e0152573, http://dx.doi.org/10.1371/journal.pone.0152573.

ANNALS OF HUMAN BIOLOGY, 2017
https://doi.org/10.1080/03014460.2017.1371222

SHORT REPORT

Mutation rates for 20 STR loci in a population from São Paulo state, Southeast, Brazil

Juliana Martinez*, Danilo Faustino Braganholi*, Isabela Brunelli Ambrósio, Fernanda Silva Polverari and Regina Maria Barretto Cicarelli

Laboratory of Paternity Investigation, São Paulo State University (UNESP), School of Pharmaceutical Sciences, São Paulo, Brazil

ABSTRACT

Short tandem repeats (STRs) are genetic markers largely employed in forensic analysis and paternity investigation cases. When an inconsistency between the parent and child is considered as a possible mutation, the mutation rate should be incorporated into paternity index calculations to give a robust result and to reduce the chance of misinterpretation. The aim of this study was to estimate the mutation rates of 20 autosomal STRs loci used for paternity tests. In these loci we analysed 29,831 parent-child allelic transfers from 929 duo or trio paternity tests carried out during 2012–2016 from São Paulo State, Brazil. We identified 35 mutations in 16 loci, and they were more frequent in the paternal germline compared to the maternal germline. The loci with the highest rate were vWA and FGA and the ones with the lowest rate were PENTA E, PENTA D, D21S11, D7S820 and D6S1043. We did not identified any mutation in D2S1338, TH01, TPOX and D16S539 loci. All mutations consisted of losses or gains of one repeat unit. Mutation rates found in the São Paulo population have peculiarities, which justifies the use of regional databases in laboratories.

ARTICLE HISTORY
Received 18 December 2016
Revised 27 June 2017
Accepted 6 July 2017

KEY WORDS
Short tandem repeats; mutation rates; paternity testing; Brazil

Introduction

Short tandem repeats (STR) are the genetic markers most often used in human identification and paternity tests. In parentage cases, the analysis of 16–20 autosomal STRs can provide a conclusive result in the majority of cases (Ge et al., 2012). To confirm a paternity exclusion, it is necessary to identify at least three inconsistencies between parent and child in the marker set (Gjertson et al., 2007), but sometimes one or two inconsistencies can be identified and, in addition to other factors, mutation should be considered.

Mutation is considered an important evolutionary force and it can be identified by comparison of genotypes between offspring and their parents (Henke & Henke, 2006; Keightley, 2007). If a mutation is identified in paternity tests, the mutation rate should be used in the paternity index calculations.

It is known that the mutation rate can vary among populations, being influenced by their ancestral origin. Brazilian populations present three main ancestral contributions from Native Americans, Europeans and Africans and the proportion of each group may vary among country regions. The ancestry proportion presented by the population of the São Paulo state is ~25% from Africans, 61% from Europeans and 14% from Native Americans. Other Brazilian populations in the North of the country, for example, present a different ancestry proportion of these three groups (Ambrosio et al., 2015; Figueiredo et al., 2015; Manta et al., 2013). Therefore, it is important to know the mutation rates in different Brazilian populations.

Materials and methods

Blood samples from paternity investigation cases were obtained from the Laboratory of Paternity Investigation (UNESP – Araraquara, Brazil) (Martins et al., 2013) during the years 2012 to October 2016.

The participants of this study live in different cities of the São Paulo state and selection criteria, like the ancestry, skin colour, age, diseases and lifestyle, were not used.

The samples were collected in Whatman FTA™ Cards (GE Healthcare, Little Chalfont, UK) and the STR amplifications were performed using the PowerPlex 16™, PowerPlex 18D™, PowerPlex 21™ and PowerPlex Fusion™ (Promega Corporation™, Madison, WI), following the manufacturer's instructions. Capillary electrophoresis was done using ABI 3500 Genetic Analyzer™ (Life Technologies, Carlsbad, CA). The genotypes were analysed using GeneMapper ID-X™ software (Life Technologies).

The mutations were investigated in 29,831 allelic transmissions (17,097 maternal and 12,734 paternal) from 592 trios (father-mother-child), 58 father/child duos and 281 mother/child duos, totalling 931 cases. Null alleles were assumed when a single discrepancy between a homozygous

CONTACT Juliana Martinez jumrtz@hotmail.com Laboratory of Paternity Investigation, São Paulo State University (UNESP), School of Pharmaceutical Sciences, São Paulo, Brazil
*These authors contributed equally to this work.
© 2017 Informa UK Limited, trading as Taylor & Francis Group

Downloaded by [UNESP] at 06:52 18 September 2017

Table 1. Genetic inconsistencies observed at STR loci in São Paulo population, Brazil.

Locus	Genotype			Mutation
	Mother	Child	Father	
D13S317	11–14	12–14	11–14	?[a]
D13S317	9–11	11–12	11–13	?[a]
D13S317	11–13	10–13	11	Paternal
Penta E	7–15	15–16	7–17	Paternal
D18S51	15–21	16–20	16–20	Maternal
D18S51	14–19	15–19	14–16	Paternal
D18S51	14–21	21–28	15–27	Paternal
CSF1PO	11–12	12.3		Maternal
CSF1PO	10–12	10–11	10–12	?[a]
Penta D	9–12	11–12	10–13	Paternal
vWA	16	16–18	16–17	Paternal
vWA	15–17	17–18	16–19	Paternal
vWA	17	17–21	17–20	Paternal
vWA	14–15	14–16	17–19	Paternal
D21S11	24.3–30	29–30	28–30	Paternal
D7S820	10–11	11–13	12	Paternal
D5S818	12–13	13–14	11–13	?[a]
D5S818	11	12–14	12–14	Maternal
D8S1179	13–14	13–15	14–14	Paternal
D8S1179	13–14	13–15	13–14	?[a]
D12S391	16–18.3	18–18.3	19–20	Paternal
D12S391	17–23	17–21	18–22	Paternal
D12S391	19–21	19–21	20–23	Paternal
D19S433	14	14–14.2	13–13.2	Paternal
D19S433	12.2–16	15–16.2	15.2–16.2	Maternal
FGA	21–23	21–26	21–25	Paternal
FGA	23–30	20–31		Maternal
FGA	22–25	21–22	22	?[a]
FGA	20–21	20–28	23–27	Paternal
D3S1358	14–17	14–17	15–20	Paternal
D3S1358	15	15–18	15–19	?[a]
D1S1656	14–17.3	17–17.3	16–18	Paternal
D1S1656	17.3–18.3	12–17.3	11–17.3	Paternal
D1S1656	12–14	14–17	14–16	Paternal
D6S1043	12–22.3	13–22.3	12–20	Paternal

[a]Indicates that it is not possible to identify the origin of the mutation.

parent and homozygous child were found. Mutations were counted after excluding null alleles. Biostatistical analysis was carried out using the Familias software version 3.0 (Egeland et al., 2000) and all cases presented a likelihood ratio favourable to family relationship (LR $\geq$ 10,000).

The maternal, paternal and total mutation rates were calculated with a 95% confidence limit. A statistical hypothesis test was proposed to identify if the mutation rates from other populations were divergent from the confidence limit obtained for São Paulo.

To maintain the quality control of results, the laboratory participated in collaborative studies of the Spanish and Portuguese Speaking Group of the International Society for Forensic Genetics (GHEP-ISFG) and Genes (Colombia) over several years.

Results and discussion

In this study, 35 mutations in 16 loci were observed, of which 23 were paternal, five maternal and seven paternal or maternal mutations. In 17 mutations gains of one step were observed, in 11 there were losses of one step and in seven there were gains or losses of one step. Mutations with gains or losses of two or more steps were not found (Table 1).

In cases where mutation was observed, the paternal mean age at the conception of the child was 34 years old

Table 2. Mutation rate observed at 20 STR loci in the São Paulo population, Brazil.

Locus	Paternal				Maternal				Paternal/Maternal	Total			
	No. of meiosis[a]	No. of mutations	Mutation rate	95% CI	No. of meiosis[b]	No. of mutations	Mutation rate	95% CI	No. of mutations	No. of meiosis[c]	No. of mutations	Mutation rate	95% CI
D13S317	650	1	1.5E−03	0–0.0046	873	0	0.0E+00	0–0	2	1523	3	1.97E−03	0–0.0042
D16S539	650	0	0.0E+00	0–0	873	0	0.0E+00	0–0	0	1523	0	0.0E+00	0–0
Penta E	650	1	1.5E−03	0–0.0046	873	0	0.0E+00	0–0	0	1523	1	6.57E−04	0–0.0019
D18S51	650	2	3.1E−03	0–0.0073	873	1	1.1E−03	0–0.0034	0	1523	3	1.97E−03	0–0.0042
CSF1PO	650	0	0.0E+00	0–0	873	1	1.1E−03	0–0.0034	1	1523	2	1.31E−03	0–0.0031
TH01	650	0	0.0E+00	0–0	873	0	0.0E+00	0–0	0	1523	0	0.0E+00	0–0
Penta D	650	1	1.5E−03	0–0.0046	873	0	0.0E+00	0–0	0	1523	1	6.57E−04	0–0.0019
vWA	650	4	6.2E−03	0.0001–0.0122	873	0	0.0E+00	0–0	0	1523	4	2.63E−03	0.0001–0.0052
D21S11	650	1	1.5E−03	0–0.0046	873	0	0.0E+00	0–0	0	1523	1	6.57E−04	0–0.0019
TPOX	650	0	0.0E+00	0–0	873	0	0.0E+00	0–0	0	1523	0	0.0E+00	0–0
D7S820	650	1	1.5E−03	0–0.0046	873	0	0.0E+00	0–0	0	1523	1	6.57E−04	0–0.0019
D5S818	650	0	0.0E+00	0–0	873	1	1.1E−03	0–0.0034	1	1523	2	1.31E−03	0–0.0031
D8S1179	650	1	1.5E−03	0–0.0046	873	0	0.0E+00	0–0	1	1523	2	1.31E−03	0–0.0031
D12S391	580	3	5.2E−03	0–0.0110	776	0	0.0E+00	0–0	0	1356	3	2.21E−03	0–0.0047
D19S433	622	1	1.6E−03	0–0.0048	837	1	1.2E−03	0–0.0035	0	1459	2	1.37E−03	0–0.0033
D2S1338	622	0	0.0E+00	0–0	837	0	0.0E+00	0–0	0	1459	0	0.0E+00	0–0
FGA	650	2	3.1E−03	0–0.0073	873	1	1.1E−03	0–0.0035	1	1523	4	2.63E−03	0.0001–0.0052
D3S1358	650	1	1.5E−03	0–0.0046	873	0	0.0E+00	0–0	1	1523	2	1.31E−03	0–0.0031
D1S1656	580	3	5.2E−03	0–0.0110	776	0	0.0E+00	0–0	0	1356	3	2.21E−03	0–0.0047
D6S1043	580	1	1.7E−03	0–0.051	776	0	0.0E+00	0–0	0	1356	1	7.37E−04	0–0.0022

[a]Paternal allelic transmission.
[b]Maternal allelic transmission.
[c]Total paternal and maternal allelic transmission.

Downloaded by [UNESP] at 06:52 18 September 2017

(range = 18–56) and maternal mean age was 24 years old (range = 17–37). Kong et al. (2012) propose that paternal age explains ~94% of the mutations identified in a population. Studies have confirmed that age may cause mutations in allelic transfer, especially from ~30 years of age, given that there is an increase of one to two mutations per year per paternal age (Francioli et al., 2015; Kong et al., 2012).

The highest mutation rate was observed for vWA and FGA markers (2.6×10^{-3}). PENTA E, PENTA D, D21S11, D7S820 and D6S1043 presented the lowest mutation rate (6.6×10^{-4}). At the loci D2S1338, TH01, TPOX and D16S539 no mutations were found (Table 2).

The vWA and FGA markers also had the highest mutation rate in another study from a São Paulo population (Sun et al., 2014) and in a Brazilian mixed population (Whittle et al., 2004); however, this was not found in Portugal (Geada et al., 2003).

There are few studies in the literature about Brazilian mutation rates. Data from the São Paulo population obtained in this study were compared with other Brazilian populations. Similarities between several rates were observed (e.g. D13S357, D18S51, CSF1PO, vWA, D21S11, D7S820 and D8S1179), but also some divergences were confirmed by the hypothesis test (e.g. D16S539, TH01 and TPOX) (Andrade et al., 2009; Chula et al., 2009; Mardini et al., 2013; Sun et al., 2014; Whittle et al., 2004).

For data reported in the STR DNA Internet DataBase (Butler & Reeder, 2017) and in Portugal (Geada et al., 2003), Nigeria (Hohoff et al., 2009) and China (Sun et al., 2014), several values were also in agreement with the São Paulo data, while others were not (e.g. D16S539, PENTA E, D2S1338).

One factor that influences the mutation on STR markers is the number of unit repeats, suggesting that the mutation rate increases in longer alleles (Brinkmann et al., 1998; Fan & Chu, 2007). Considering that allele frequencies can vary among populations, it is expected that populations with a high occurrence of longer alleles will have a different mutation rate from those with low occurrence. Thus, the mutation rate for the same marker can vary among populations with different allele frequencies.

Conclusions

For some loci there are discrepancies in the mutation rates among populations, which justifies the use of regional data in laboratories. Thus, the results of this study are important to obtain a correct likelihood ratio in the São Paulo population if the mutation rates need to be used.

Acknowledgements

This study was supported by the PADC/FCF-UNESP (#2015/02-I). We acknowledge the financial support provided by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and National Counsel of Technological and Scientific Development (CNPq).

Disclosure statement

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

Funding

This study was supported by the PADC/FCF-UNESP (#2015/02-I). We acknowledge the financial support provided by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and National Counsel of Technological and Scientific Development (CNPq).

ORCID

Juliana Martinez http://orcid.org/0000-0003-3496-1961

References

Ambrosio IB, Braganholi DF, Polveiari FS, Cicarelli RMB. 2015. Mitochondrial DNA 30-SNP data confirm high prevalence of African lineages in the population of Espirito Santo, Brazil. Forensic Sci Int 5:e346–e347.

Andrade ES, Gomes AV, Raposo G, Mauricio-da-Silva L, Silva Rdos S. 2009. Mutation rates at 14 STR loci in the population from Pernambuco, Northeast Brazil. Forensic Sci Int Genet 3:e141–e143.

Brinkmann B, Klintschar M, Neuhuber F, Hühne J, Rolf B. 1998. Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. Am J Hum Genet 62:1408–1415.

Butler JM, Reeder DJ. 2017. STR DNA Internet database. Available online at: http://www.cstl.nist.gov/strbase/ (accessed April 2017).

Chula FG, Rodenbusch R, Schumacher S, Grandi T, Michelon CT, Gastaldo AZ, Costi C, et al. 2009. 15 STR loci frequencies with mutation rates in the population from Rio Grande do Sul, Southern Brazil. Forensic Sci Int Genet 3:e35–e38.

Egeland T, Mostad P, Bente M, Margurethe S. 2000. Beyond traditional paternity and identification cases. Selecting the most probable pedigree. Forensic Sci Int Genet 110:47–59.

Fan H, Chu JY. 2007. A brief review of short tandem repeat mutation. Genomics Proteomics Bioinformatics 5:7–14.

Figueiredo RF, Ambrosio IB, Braganholi DF, Chemale G, Martins JA, Gomes V, Gusmão L, Cicarelli RMB. 2015. Male-specific contributions to the Brazilian population of Espirito Santo. Int J Legal Med 130:679.

Francioli LC, Polak PP, Koren A, Menelaou A, Chun S, Renkens I, Genome of the Netherlands Consortium, et al. 2015. Genome-wide patterns and properties of de novo mutations in humans. Nature Genetics 47:822–826.

Ge J, Eisenberg A, Budowle B. 2012. Developing criteria and data to determine best options for expanding the core CODIS loci. Investig Genet 3:1.

Geada H, Viriato L, Vieira-Silva C, Cruz C, Lucas I, Ribeiro T, Espinheira R. 2003. STR mutations in paternity investigations: a study of 1-year consecutive cases. Int Congr Ser 1239:657–660.

Gjertson DW, Brenner CH, Baur MP, Carracedo A, Guidet F, Luque JA, Lessig R, et al. 2007. ISFG: recommendations on biostatistics in paternity testing. Forensic Sci Int Genet 1:223–231.

Henke L, Henke J. 2006. Supplemented data on mutation rates in 33 autosomal short tandem repeat polymorphisms. J Forensic Sci 51:446–447.

Hohoff C, Schürenkamp M, Brinkmann B. 2009. Meiosis study in a population sample from Nigeria: allele frequencies and mutation rates of 16 STR loci. Int J Legal Med 123:259–261.

Keightley P. 2007. Rates and fitness consequences of new mutations in humans. Genetics 190:295–304.

Kong A, Frigge ML, Masson G, Besenbacher S, Sulem P, Magnusson G, Gudjonsson SA, et al. 2012. Rate of de novo mutations, father's age, and disease risk. Nature 488:471–475.

Manta FSN, Pereira R, Vianna R, Rodolfo Beuttenmüller de Araújo A, Leite Góes Gitaí D, Aparecida da Silva D, de Vargas Wolfgramm E, et al. 2013. Revisiting the genetic ancestry of Brazilians using autosomal AIM-Indels. PLoS One 8:e75145.

Mardini AC, Rodenbusch R, Schumacher S, Chula FG, Michelon CT, Gastaldo AZ, Maciel LP, et al. 2013. Mutation rate estimates for 13 STR

loci in a large population from Rio Grande do Sul, Southern Brazil. Int J Legal Med 127:45–47.
Martins JA, Braganholi DF, Felix JHP, de Almeida CF, Brunelli I, Alvarenga VLS, Cicarelli RMB. 2013. DNA human identification for the family rights: one year of agreement between UNESP/Laboratory of Paternity and Public Defender Service in São Paulo State. Forensic Sci Int Genet 4:113–114.
Sun H, Liu S, Zhang Y, Whittle MR. 2014. Comparison of southern Chinese Han and Brazilian Caucasian mutation rates at autosomal short tandem repeat loci used in human forensic genetics. Int J Legal Med 128:1–9.
Whittle MR, Romano NL, Negreiros VA. 2004. Updated Brazilian genetic data, together with mutation rates, on 19 STR loci, including D10S1237. Forensic Sci Int 139:207–210.

Downloaded by [UNESP] at 06:52 18 September 2017

Forensic Science International: Genetics Supplement Series 6 (2017) e82–e83

Contents lists available at ScienceDirect

Forensic Science International: Genetics Supplement Series

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

Genetic characterization of four Brazilian states with 25 Yfiler®Plus markers

J. Jannuzzi[a,*], C.S. Alho[b], C. Fridman[c], S.R.B. Ferreira[d], D.F. Braganholi[e], I.B. Ambrosio[e], R.M.B. Cicarelli[e], V. Gomes[f,g], E.F. Carvalho[a], L. Gusmão[a,f,g]

[a] Laboratório de Diagnósticos por DNA (LDD), Universidade do Estado do Rio de Janeiro (UERJ), Rio de Janeiro, Brazil
[b] Laboratório de Genética e Biologia Molecular, Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS), Rio Grande do Sul, Brazil
[c] Departamento de Medicina Legal, Ética Médica e Medicina Social e do Trabalho da Faculdade de Medicina da USP, Universidade de São Paulo (USP), São Paulo, Brazil
[d] Faculdade Pitágoras, São Luís, Brazil
[e] Faculdade de Ciências Farmacêuticas (FCFAr), Universidade Estadual Paulista Júlio de Mesquita Filho (UNESP), São Paulo, Brazil
[f] IPATIMUP-Institute of Pathology and Molecular Immunology from the University of Porto, Portugal
[g] I3S-Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Porto, Portugal

ARTICLE INFO

Keywords:
Y-STRs
Brazil
Haplotype diversity
Admixed populations
South America

ABSTRACT

The Y-STRs have high mutation rates, being useful to discriminate unrelated males. They are widely used in paternity and forensic investigations to study the recent history and migration movements of populations. Recently, the YfilerPlus was released to increase discrimination inside populations, by adding 9 Y-STRs to the previous Yfiler version. The genetic composition of Brazil is known to vary through 5 geopolitical regions: South, Southeast, North, Northeast and Central-west. Therefore, for this study, samples from Maranhão, Espírito Santo, São Paulo and Rio Grande do Sul were characterized for the Y-STR markers present in YfilerPlus kit, to evaluate if diversity increases with the enlargement of the Yfiler set, and if significant differences exists among populations. Genetic differentiation analysis did not reveal statistically significant differences in the YfilerPlus haplotype composition of the studied samples. High diversities were observed in all samples for both Yfiler and YfilerPlus marker sets. Inside populations, 20 haplotypes were shared by two individuals for the Yfiler-STRs. This number decreases to 14 when using the 25 Y-STRs from the YfilerPlus. The YfilerPlus demonstrated an increased discrimination power in comparison to the Yfiler kit, being suitable for forensic applications in the studied Brazilian populations, for which data were not yet available.

1. Introduction

The Y short tandem repeats (Y-STRs) have high mutation rates ($\approx 10^{-3}$), and the set of their alleles characterize haplotypes. Due to their mutation rates, Y-STRs can discriminate male individuals, being widely used in paternity tests, forensic investigations, genealogical researches, besides the analysis of recent history of populations and its migration movements [1–4].

There are diverse sets of Y-STRs being used in forensic and population genetics. To increase discrimination inside populations, new commercial kits have been released displaying additional markers with high mutation rates. The YfilerPlus is a recent Y-STR forensic kit, with 25 markers, including the 16 from the previous Yfiler version.

Brazilian populations are very heterogeneous, presenting a high genetic diversity spread through the 5 country regions, named South, Southeast, North, Northeast and Central-west [5–7].

Since YfilerPlus was recently released and there is no data available for Brazilian populations, the aims of this study were to (i) genotype samples from four states of three geopolitical regions, (ii) evaluate if the addition of Y-STR markers would increase the diversity values in comparison to the observed with Yfiler and (iii) evaluate if there is population substructure between these Brazilian states.

2. Materials and methods

A total of 518 unrelated male individuals from Maranhão (n = 114), Espírito Santo (n = 76), São Paulo (n = 126) and Rio Grande do Sul (n = 202) provided their samples for this study, after signing a written informed consent for cooperation under strictly confidential conditions.

DNA was extracted with standard chelex and phenol/chloroform methodologies. Samples were genotyped with the Yfiler®Plus kit (Applied Biosystems), following the User Guide [8]. Capillary electrophoresis and detection were performed on a 3500 Genetic Analyzer using POP-7™ polymer (Applied Biosystems). The genotypes were assigned using the software GeneMapper ID v4.1 (Applied Biosystems).

* Corresponding author.
E-mail address: jannuzzi.juliana@gmail.com (J. Jannuzzi).

Table 1
Population pairwise R_{ST} genetic distances between the four Brazilian states.

	Espírito Santo	São Paulo	Maranhão	Rio Grande do Sul
Espírito Santo	–	0.7058 ± 0.0041	0.4119 ± 0.0045	0.5358 ± 0.0047
São Paulo	−0.0015	–	0.3699 ± 0.0045	0.5982 ± 0.0051
Maranhão	0.0001	0.0002	–	0.0939 ± 0.0030
Rio Grande do Sul	−0.0005	−0.0006	0.0023	–

Legend: Below diagonal shows the R_{ST} values; Above diagonal shows non-differentiation probabilities calculated for 10.100 permutations. Significance level was equal 0.0083.

The R_{ST} genetic distances were calculated with Arlequin v3.5 [9]. This software was also used to calculate haplotype diversities (HD) and the number of shared haplotypes for the 25 Y-STRs and the 16 STRs from the Yfiler (after removing extra markers), in the same population samples. For the analysis of genetic distances between the four population samples, the number of repetitions in DYS398I was subtracted from DYS398II and DYS385 and DYF387S1 were excluded. Null, microvariant or duplicated alleles were represented as missing data.

3. Results and discussion

High values of HD were found with the 25 Y-STRs in all Brazilian samples analyzed (Maranhão, HD = 0.9992 ± 0.0013; Espírito Santo, HD = 0.9996 ± 0.0022; São Paulo, HD = 1.0000 ± 0.0010; Rio Grande do Sul, HD = 0.9996 ± 0.0005), due to a low number of shared haplotypes inside populations. In total, there were 14 haplotypes shared by two individuals: 1 in Espírito Santo, 5 in Maranhão and 8 in Rio Grande do Sul. When just considering the 16 STRs from the Yfiler, a slight decrease in the HD was observed in Maranhão (0.9989 ± 0.0013), Espírito Santo (0.9998 ± 0.0023) and Rio Grande do Sul (0.9995 ± 0.0006), while remained the same in São Paulo, indicating that both marker sets could equally discriminate individuals from this population sample. After the reduction of the number of Y-STRs to 16, there was an increase in the number of shared haplotypes inside populations: 2 from Espírito Santo, 7 from Maranhão, 11 from Rio Grande do Sul. The improvement in discrimination inside populations that was obtained with the new kit, was also observed in previous studies on other populations [10–12].

Genetic distance analysis indicated a homogeneity between the four populations, since R_{ST} values were low and the non-differentiation probabilities high, far above the significance level (Table 1). A previous study in other Brazilian population samples with 23 Y-STRs showed similar results, in which no significant genetic distances were found between most populations among the five regions of Brazil [6]. This genetic homogeneity can be explained by the fact that the set and the number of markers used provides high discrimination among individuals, being more suitable to detect differences between individuals than between populations.

4. Conclusion

It is possible to conclude that the addition of markers in the YfilerPlus kit increased the discrimination inside populations in comparison with the just 16 markers present in Yfiler kit. Therefore, the 25 STRs from the YfilerPlus demonstrated to be suitable for forensic applications in the studied populations from Brazil, for which data were not yet available.

However, it is still necessary to analyze whether this applies to larger population samples and to generate YfilerPlus data for other Brazilian population samples that were not yet studied.

Conflict of interest statement

None.

Acknowledgements

Financial support was granted by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES, Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq. VG is supported by Fundação para a Ciência e Tecnologia (FCT) and Programa Operacional Potencial Humano (POCH), through the grant SFRH/BPD/76207/2011. IPATIMUP integrates the i3S Research Unit, which is partially supported by FCT.

References

[1] J.M. Butler, Fundamentals of Forensic DNA Typing, 1st edition, (2009).
[2] M. Kayser, R. Kittler, A. Erler, M. Hedman, A.C. Lee, A. Mohyuddin, et al., A comprehensive survey of human Y-chromosomal microsatellites, Am. J. Hum. Genet. 74 (2004) 1183–1197.
[3] K.N. Ballantyne, M. Goedbloed, R. Fang, O. Schaap, O. Lao, A. Wollstein, et al., Mutability of Y-chromosomal microsatellites: rates, characteristics, molecular bases, and forensic implications, Am. J. Hum. Genet. 87 (2010) 341–353.
[4] K.N. Ballantyne, V. Keerl, A. Wollstein, Y. Choi, S.B. Zuniga, A. Ralf, 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. 6 (2012) 209–218.
[5] F. Marta, R. Pereira, R. Vianna, A. Rodolfo Barrantes Iller de Avelar, D. Leite Góes, D. Giuf, D. Aparecida da Silva, et al., Revisiting the genetic ancestry of Brazilians using autosomal AIM-Indels, PLoS One 8 (2013) 1–11.
[6] T. Palha, L. Guerreiro, E. Ribeiro-Rodrigues, J.F. Guerreiro, Â. Ribeiro-dos-Santos, S. Santos, Disclosing the genetic structure of Brazil through analysis of male lineages with highly discriminating haplotypes, PLoS One 7 (2012) 1–8.
[7] R. Scapoli, L. Guerreiro, M. Geppert, L. Roewer, T. Palha, L. Alvarez, et al., Male lineages in Brazil: intercontinental admixture and stratification of the european background, PLoS One 11 (2016) e0152573.
[8] Yfiler® Plus PCR Amplification Kit User Guide, Thermo Fisher Scientific Inc., 2014.
[9] L. Excoffier, H. Lischer, Arlequin: An Integrated Software Package for Population Genetic Data Analysis, (2015).
[10] J.K. Olofsson, H.S. Mogensen, A. Buchard, C. Banning, N. Morling, Forensic and population genetic analyses of Danes, Greenlanders and Somalis typed with the Yfiler® Plus PCR amplification kit, Forensic Sci. Int. Genet. 16 (2015) 232–236.
[11] C. Saporta, E. D'Atanasio, A. Agostino, M. Madano, M.T. Papaluca, F. Cruciani, et al., Forensic genetic value of a 27 Y-STR loci multiplex (Yfiler® Plus kit) in an Italian population sample, Forensic Sci. Int. Genet. 21 (2015) 1–5.
[12] O. García, I. Yurrebaso, I.D. Mendizábal, S. López, S. Alonso, L. Gusmão, Data for 27 Y-chromosome STR loci in the basque country autochthonous population, Forensic Sci. Int. Genet. 20 (2016) e10–e12.

3

BIOTECNOLOGIA APLICADA À SAÚDE

FUNDAMENTOS E APLICAÇÕES



organizador
RODRIGO RIBEIRO RESENDE

colaboradores
SILVIA GUATIMOSIM
MARCUS VINICIUS GOMEZ
CARLOS R. SOCCOL

Blucher

FAPEMIG

CAPÍTULO 20

APLICAÇÕES FORENSES DO DNA: FUNDAMENTOS, MÉTODOS E LIMITAÇÕES

Isabela Brunelli Ambrosio
Danilo Faustino Braganholi
Regina Maria Barretto Cicarelli

20.1 INTRODUÇÃO

A identificação humana é o processo científico por meio do qual se determina a identidade de uma pessoa, através de um conjunto de caracteres que a individualize, fazendo-a igual apenas a si mesma[1]. Esta identificação por meio do DNA constitui um dos produtos mais revolucionários da genética moderna.

A genética forense, também conhecida como DNA Forense, é a área que utiliza os conhecimentos e técnicas de genética e de biologia molecular no auxílio à justiça, sendo aceita rotineiramente em processos judiciais em todo o mundo.

Apesar de ser a identificação humana pelo DNA o ramo mais desenvolvido da genética forense, bem como sua aplicação mais popular ser o teste de paternidade, ela não se limita a isso, e pode ser aplicada na identificação ou individualização de animais, plantas e micro-organismos.

A genética forense iniciou-se quando foram utilizadas pela primeira vez características genéticas para a definição de paternidade, ajudando a justiça na solução de um caso criminal[2, 3]. Sua fase moderna teve início na década de 1980, quando pesquisadores descobriram regiões altamente variáveis do DNA, capazes de individualizar uma pessoa[2, 3, 4]. Em 1985, sir Alec Jeffreys[3] apelidou as características únicas do DNA de uma pessoa de "impressões digitais do DNA". O perfil genético de um indivíduo, comumente utilizado na identificação humana, é baseado na combinação de diversos marcadores que são herdados de seus progenitores. Esses marcadores referem-se geralmente a diferenças nas sequências de DNA entre os indivíduos (polimorfismos).

Os marcadores genéticos utilizados em identificação humana podem ser biparentais (metade dos cromossomos herdados da mãe e a outra metade herdada do pai) ou uniparentais (cromossomo Y herdado apenas do pai em indivíduos do sexo masculino e DNA mitocondrial herdado apenas da mãe em todos os indivíduos). Os marcadores biparentais autossômicos permitem uma identificação individual porque um determinado perfil genético é específico de um só indivíduo, sendo apenas partilhado por gêmeos monozigóticos. Os marcadores uniparentais não permitem a identificação individual, pois todos os indivíduos aparentados por via masculina partilharão a mesma linhagem de cromossomo Y, enquanto todos os indivíduos aparentados por via feminina partilharão a mesma linhagem de DNA mitocondrial, salvo nos casos de ocorrência de mutações durante a transmissão.

20.2 POLIMORFISMOS STRS (AUTOSSÔMICOS E DOS CROMOSSOMOS SEXUAIS, X E Y)

Desde que se desenvolveu a tecnologia do DNA recombinante, em 1973, dividindo em pedaços o DNA genômico para estudar os genes, houve um avanço na biologia molecular, e diversos genes foram descritos e isolados, muitos deles relacionados com a identificação humana[5]. Com o uso de enzimas de restrição e sondas de DNA, foram exploradas as variabilidades ou polimorfismos do DNA que variavam de indivíduo para indivíduo. Descobriram-se, também, regiões constituídas por vários pares de bases, repetidas inúmeras vezes, com alelos variando quanto ao número de sequências, denominados minissatélites (*variable number tandem repeats* – VNTR). Inicialmente, os VNTRs foram utilizados como marcadores genéticos informativos, pois estes são compostos por cem ou mais alelos com diferentes comprimentos na população humana

Outros marcadores altamente polimórficos foram encontrados no genoma humano: os microssatélites (*short tandem repeats,* STR) (Figura 20.1). Os STRs são repetições curtas de 1 a 5 pares de bases encontrados em eucariotos. Tais repetições são encontradas tanto em regiões codificantes quanto não codificantes do DNA, cuja função ainda é desconhecida[5].

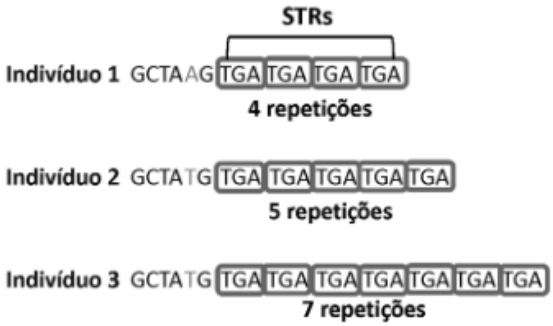


Figura 20.1 Esquema representando variações dos polimorfismos *short tandem repeats* (STRs). O número de repetições das pequenas sequências configuram um alelo que pode variar entre indivíduos.

Inovação tecnológica revolucionária na biologia molecular surgida em 1985, a reação da polimerase em cadeia ou *polymerase chain reaction* (PCR), tornou-se de uso universal. Por meio dessa reação, quantidades muito pequenas de DNA podem ser amplificadas para obtenção de praticamente qualquer quantidade desejada[6, 7, 8]. A técnica utiliza o mesmo princípio pelo qual o DNA é normalmente copiado na célula, exceto pelo fato de que somente um pequeno segmento do DNA no cromossomo será amplificado. Isso possibilitou processar quantidades muito pequenas de DNA frequentemente deixadas como evidência de um crime e aumentou enormemente a sensibilidade dos sistemas disponíveis usados nas análises forenses, auxiliando o sistema judiciário no julgamento de crimes. Graças à PCR, diminutas quantidades de DNA podem ser extraídas de fios de cabelo, selos postais, pontas de cigarro, xícaras de café e outras fontes de DNA.

Os STRs substituíram os polimorfismos de comprimento dos fragmentos de restrição (do inglês, *restriction fragment length polymorphisms* – RFLPs) como principal método de identificação genômica e tornaram-se popular na tipagem de DNA para análises forenses. São sequências curtas que variam de

duas a quatro bases, podendo repetir-se até dezessete vezes (de sete a quinze alelos por lócus), comumente amplificadas por PCR. Os métodos para tipagem de STRs estão atualmente automatizados, com detecção envolvendo fluoróforos[9].

Com o desenvolvimento de reações da PCR em multiplex (Figura 20.2), ou seja, amplificação de várias regiões do DNA em um único tubo, empresas internacionais iniciaram a comercialização de kits contendo um *mix* ou mistura com diversos iniciadores (*primers*), que permitem amplificar até 24 regiões do DNA genômico, sendo 23 *loci* STR mais a amelogenina para determinação do sexo. Entretanto, para aplicação e uso nos diferentes laboratórios forenses, há a necessidade de se conhecer as frequências dos respectivos alelos presentes na população daquela região geográfica em que os testes estão sendo realizados. Tais frequências devem ser publicadas em revistas indexadas para que possam ser utilizadas nas análises estatísticas dos resultados dos perfis que farão parte da conclusão do laudo pericial.

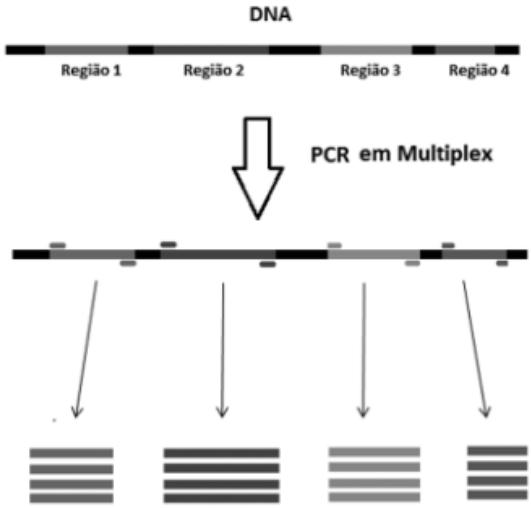


Figura 20.2 Esquema representando uma reação de PCR multiplex, na qual são amplificadas diferentes regiões do DNA em uma única reação.

Tratando-se de amostras obtidas em cenas de crimes, casos de violação sexual (estupro), desastres em massa ou de restos mortais de exumação de cadáver, uma questão fundamental, concernente ao uso do DNA como evidência, está na validação científica dos métodos de análise dos STRs. Em outras palavras, é preciso ter garantias científicas de que os testes podem, inequivocamente, identificar coincidências ou não coincidências para cada marcador genético utilizado, na comparação das amostras questionada e de referência. Inicialmente, a credibilidade dos testes deve partir da natureza das amostras biológicas utilizadas[2]. Muito frequentemente as amostras são encontradas em superfícies não esterilizadas, podendo sofrer danos após contato com a luz solar, micro-organismos e solventes. Existem procedimentos que podem minimizar a ação desses fatores de degradação do DNA. Entretanto, muitos cuidados devem ser tomados para evitar equívocos na interpretação dos resultados dos perfis alélicos. Ressalte-se que a cadeia de custódia das amostras deve ser mantida até o momento de serem processadas no laboratório.

A amplificação por PCR pode produzir falhas e artefatos devido à baixa qualidade do material biológico. Amostras de DNA parcialmente degradadas podem proporcionar, por exemplo, a amplificação preferencial de alguns alelos e/ou o surgimento das bandas fantasmas (*stutter bands*) (Figura 20.3)[2]. No primeiro caso, pode-se ter a amplificação de um alelo em detrimento do outro, o que pode gerar a falsa impressão de se tratar de um indivíduo homozigoto em vez de heterozigoto para o lócus em estudo. As bandas fantasmas ocorrem em virtude de falhas no processo de amplificação, gerando bandas com uma unidade de repetição a menos que a do alelo original. Desse modo, pode-se equivocadamente interpretar o resultado como um falso heterozigoto ou identificar um alelo erroneamente.

Apesar dessas dificuldades, a genética é a área da ciência forense que mais tem avançado na atualidade. Há alguns anos, a Secretaria Nacional de Segurança Pública do Brasil está implementando o Banco de Dados Nacional Criminal de Perfis Genéticos, similarmente ao americano CODIS (Combined DNA Index System), que armazena dados de criminosos condenados, e o europeu Fenix, que contém o perfil genético de milhares de pessoas desaparecidas. Tais ferramentas tornam mais ágil a troca de informações entre as instituições espalhadas pelo vasto território no país, como no caso do CODIS nos Estados Unidos da América[10], facilitando a resolução de diversos casos.

No Brasil, a implementação desse banco de dados trouxe o aumento da demanda nos laboratórios de perícia, uma vez que ele permite, por exemplo,

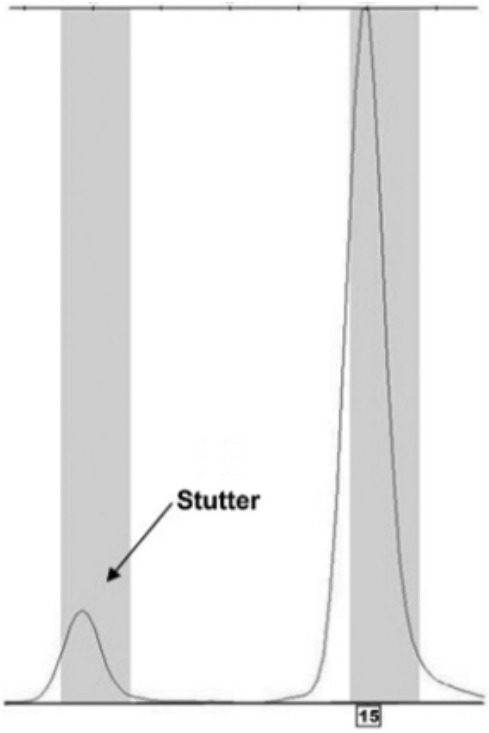


Figura 20.3 Esquema representando um alelo e um *stutter*. A seta indica o *stfuter*. Nota-se a grande diferença entre as intensidades de sinais entre *stutter* e alelo verdadeiro.

identificar o criminoso pela análise de uma única gota de sangue encontrada no local do crime.

Em 2006, alguns peritos judiciais publicaram artigos[11, 12] revelando as possibilidades de erros em exames de DNA, o que acarretou um convite da revista *Scientific American Brasil* para escrever um artigo sobre o estado da arte da genética forense no Brasil[13]. Em decorrência dessas publicações, vários laboratórios brasileiros começaram a ser condenados por erros em exames de DNA.

Na falta de um órgão nacional credenciado para promover a acreditação de laboratórios de DNA forense no Brasil, como é o caso, por exemplo, da American Association of Blood Banks (AABB) nos Estados Unidos, o

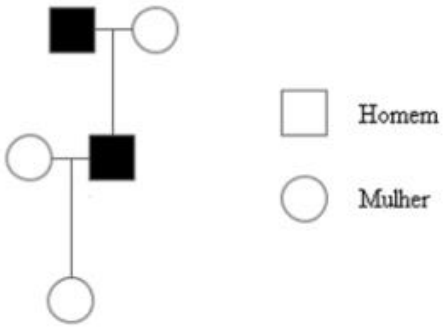
Ministério da Justiça solicitou que o Instituto Nacional de Metrologia, Qualidade e Tecnologia (INMETRO) viesse a padronizar os exames de DNA realizados por laboratórios da polícia científica[14]. Mais recentemente, a Agência Nacional de Vigilância Sanitária (ANVISA) publicou a Resolução RDC n. 11, de 16 de fevereiro de 2012, estabelecendo normas para o funcionamento de laboratórios que realizam testes e análises técnicas, o que abrange também os laboratórios que realizam investigação de paternidade.

Em associação aos STRs autossômicos, que muito contribuem para a identificação do indivíduo tanto quanto para a investigação de paternidade, os Y-STRs relacionados com o cromossomo Y, bem como os X-STRs do cromossomo X, são também ferramentas importantes na elucidação das análises de DNA forense ou casos de relação biológica. Os primeiros são importantes na determinação de linhagem patrilinear, mas não servem para incluir o indivíduo como o suposto pai da criança do sexo masculino, uma vez que o cromossomo Y é transmitido de pai para filho e deste para seu filho de maneira haplotípica, à semelhança da transmissão do DNA mitocondrial (herança materna), levando a que toda a descendência do sexo masculino, daquele primeiro indivíduo compartilhe os mesmos haplotipos Y-STRs; ou seja, todos os descendentes do sexo masculino apresentarão o mesmo padrão haplotípico do seu ancestral. Por essa razão, não se pode associar (ou multiplicar) os valores obtidos a partir das frequências dos alelos do perfil dos STRs autossômicos, denominados de índice de verossimilhança, com aquele da frequência do haplotipo dos Y-STRs. Entretanto, da análise do perfil Y-STR é possível excluir um indivíduo como pai biológico de uma criança do sexo masculino ou como doador de uma mancha de sangue ou sêmen.

Em relação ao cromossomo X (X-STRs), este é transmitido pelo pai para todas as filhas do sexo feminino, que recebem o outro cromossomo X da mãe biológica. Como neste caso o padrão de transmissão é muito semelhante aos STRs autossômicos, em geral os índices obtidos através de cálculos estatísticos utilizando as respectivas frequências dos alelos de ambos, STRs e X-STRs, podem ser multiplicados, aumentando, assim, as chances de favorecimento de paternidade, quando se trata de filha questionada na ausência do suposto pai, por exemplo (Figura 20.4).

A equipe da autora deste capítulo elaborou um banco de dados muito interessante e bastante completo para estudo de vários parâmetros de análise do cromossomo X*, os quais foram obtidos a partir das publicações nacionais sobre esse tema e de dados do próprio laboratório. Esforços estão sendo

* Ver http://bgbx.com.br. Número de registro de programa de *software*: 12166-1.



STRs autossômicos
Probabilidade de Relação Biológica: 71,79757%

X

X-STRs
Probabilidade de Relação Biológica: 99,9813421%

STRs autossômicos e X-STRs
Probabilidade de Relação Biológica: 99,9926702 %

Figura 20.4 Heredograma de um caso de investigação de paternidade em que foi utilizada a análise de X-STRs. Os indivíduos em preto não foram analisados. Foi utilizada neste caso a análise da mãe biológica do suposto pai, uma vez que este estava ausente. O esquema mostra a probabilidade de relação biológica com os STRs autossômicos e X-STRs; a multiplicação dos índices aumenta a probabilidade de relação biológica entre os indivíduos analisados.

investidos no sentido de disponibilizar também as fórmulas para realizar os cálculos estatísticos mais importantes a serem incluídos nos laudos periciais em que há a necessidade da associação dos STRs autossômicos e X-STRs, garantindo, desse modo, um índice de paternidade ou de verossimilhança mais robusto.

A utilização das análises dos marcadores de Y e X-STRs é importante em casos de confirmação de exclusão de paternidade e, especialmente, em casos

de reconstrução, quando o suposto pai está ausente. Em casos criminais, estas também são ferramentas importantes para elucidar a origem de manchas de sangue e/ou sêmen.

A técnica de amplificação de STRs (autossômicos, Y e X) baseia-se na realização de uma reação da PCR multiplex, isto é, na qual vários *primers* são adicionados a um mesmo tubo de 0,2 mL contendo um mix da PCR (tampão, cloreto de magnésio, dNTPs e enzima Taq polimerase) e acrescentando-se à mistura o DNA em estudo.

Duas reações para controle das condições de amplificação são necessárias e imprescindíveis: um controle positivo, contendo DNA humano conhecido e convenientemente diluído (concentração de uso 0,5 ng/uL), e outro negativo, geralmente realizado com água destilada esterilizada. Os tubos são levados ao termociclador para a amplificação em ciclagem adequada e compatível com a reação multiplex.

Atualmente, o uso de kits comerciais são os mais recomendáveis para as análises forenses e de paternidade por serem validados nacional e internacionalmente. A seguir, a Figura 20.5 apresenta um esquema dos procedimentos básicos para a reação.

O sucesso da análise dos STRs está diretamente relacionado com a qualidade do DNA obtido. O DNA está presente em todas as células nucleadas e, portanto, pode ser extraído de diferentes materiais biológicos deixados na cena de um crime. Em geral, as amostras testadas nos laboratórios forenses são tipicamente sangue e manchas de sêmen, mas podem existir muitas outras fontes de DNA.

Para a comparação dos perfis de DNA, sempre há a necessidade de amostras de referência (sangue ou *swab* oral). Todas as amostras devem ser cuidadosamente manuseadas para a extração de DNA, de modo a evitar qualquer tipo de contaminação. Por essa razão, muitos laboratórios processam as amostras de evidência (aquelas obtidas na cena do crime) em momentos diferentes ou mesmo em locais diferentes das amostras de referência. Existem diversos métodos de extração de DNA; entretanto, os mais utilizados atualmente são orgânico, Chelex e papel FTA. Todos eles permitem a obtenção de DNA de boa qualidade para análise de STRs. A Figura 20.6 ilustra um perfil alélico amplificado a partir de uma amostra de DNA humano.

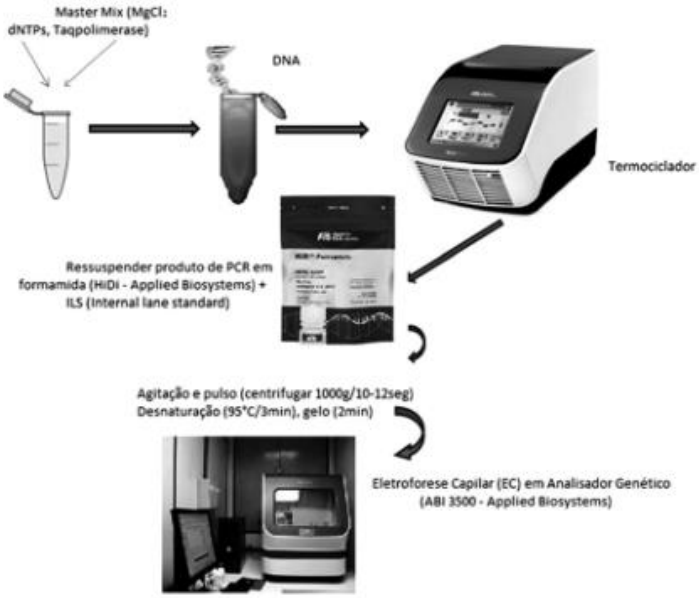


Figura 20.5 Esquema ilustrativo da reação de PCR multiplex para amplificação de STRs (autossômicos, Y ou X-STRs).

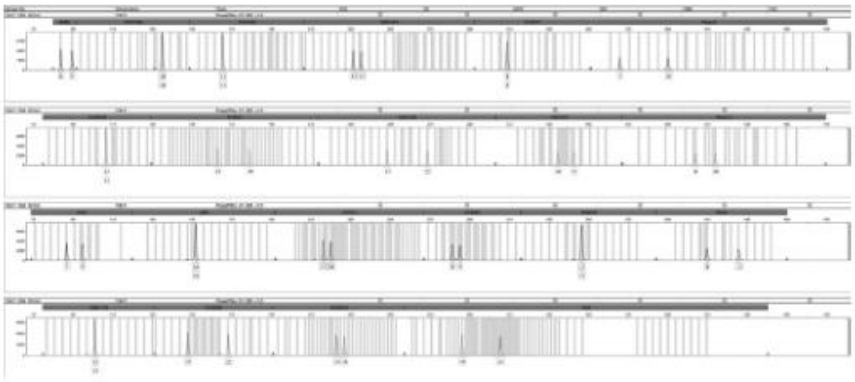


Figura 20.6 Eletroferograma obtido no analisador genético (ABI 3500 – Applied Biosystems) e analisado no GeneMapper ID-X apresentando os alelos amplificados na reação multiplex de 21 regiões do DNA humano (STRs autossômicos). A amostra pertence a um indivíduo do sexo masculino (X/Y). Os picos representam os alelos obtidos em cada marcador, que são repetições de pequenas sequências de bases, como mostrado na Figura 20.1.

20.3 POLIMORFISMOS DE INSERÇÃO/ DELEÇÃO (INDELS) E SUAS APLICAÇÕES

O estudo das variações genéticas, utilizando os polimorfismos do DNA, permite uma melhor compreensão da história e da diversidade das populações humanas, além de proporcionar um sistema para a identificação genética de indivíduos. A aplicação do DNA decorre do seu alto poder de discriminação gerado por seus polimorfismos, e além dos STRs já estudados anteriormente, destacam-se outros dois: *single nucleotide polymorphisms* (SNPs) e *insertion-deletion polymorphisms* (INDELs)[15].

Weber e colaboradores (2002)[16] foram pioneiros, pois identificaram e caracterizaram aproximadamente 2 mil INDELs no genoma humano, com diferentes tamanhos e frequências alélicas em europeus, africanos, japoneses e nativo-americanos, e devido à abundância e facilidade de análise, destacaram a sua utilidade para estudos genéticos. INDELs chegam a representar cerca de 16% de todos os polimorfismos do DNA humano e se apresentam muito distribuídos pelo genoma, em média, um INDEL a cada 7,2 Kb[17].

INDELs são polimorfismos de comprimento, caracterizados pela inserção ou deleção de um ou mais nucleotídeos em uma determinada região do genoma; por exemplo, o polimorfismo rs16363 é identificado pela deleção da sequência "TGTTT" na localização cromossômica 22q13.1. O polimorfismo ao ser inserido no National Center for Biotechnology Information (NCBI) recebe um código único descrito com as letras rs (*reference sequence*), seguido da numeração específica para sua identificação correspondente à sua posição no genoma. Na localização cromossômica, o primeiro número refere-se ao cromossomo; no exemplo citado acima, o polimorfismo está localizado no cromossomo 22; a letra refere-se aos braços do cromossomo; a letra *p* indica o braço curto e *q* o braço longo; o número seguinte representa a posição no braço do cromossomo, no exemplo, banda 13, sub-banda 1.

A Figura 20.7 a seguir ilustra a identificação do INDEL "ATCG" no eletroferograma.

20.4 UTILIZAÇÃO DOS INDELS NA ANÁLISE DE DNA DEGRADADO

O DNA pode ser danificado ou destruído devido a condições ambientais adversas. A exposição ao ambiente degrada as moléculas de DNA, dividindo-o aleatoriamente em pequenos pedaços. Água, oxigênio, radiação

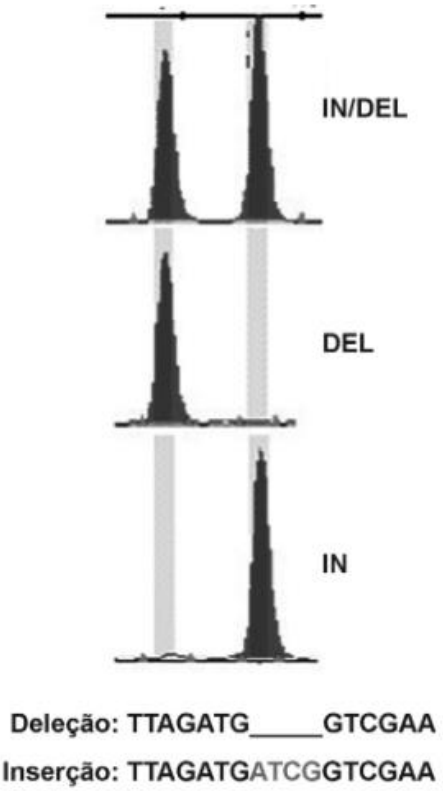


Figura 20.7 Eletroferograma identificando o INDEL (In/Del) com o alelo curto (deleção-Del) e o alelo longo (inserção-In-ATCG).

ultravioleta e enzimas nucleases são alguns dos agentes naturais que atuam nessa degradação[18].

A habilidade de recuperar e analisar sequências de DNA de amostras como ossos e dentes, expostos ao longo do tempo a uma variedade de condições ambientais, tornou-se uma ferramenta valiosa para a identificação de indivíduos ou amostras desconhecidas[19].

Pouco DNA endógeno, ação do meio ambiente, micro-organismos e presença de inibidores são os principais fatores que fazem com que a

recuperação e análise do DNA deste tipo de amostra sejam um desafio na área de identificação humana[20].

Ainda que a análise de STR (*short tandem repeats*) seja o principal polimorfismo utilizado como marcador genético em identificação humana[18], conforme já foi mencionado anteriormente, apresenta dificuldades quando as amostras possuem DNA degradado e/ou em pouca quantidade, devido ao tamanho relativamente grande dos fragmentos a serem amplificados (150 pb a 500 pb)[21].

Por essa razão, aumentou de modo considerável recentemente o interesse na utilização de INDELs como marcadores genéticos nas áreas de identificação humana e genética forense, devido à facilidade de análise, ampla distribuição no genoma, por apresentarem fragmento de pequeno tamanho, cerca de 100 pb e poderem ser detectados por diferentes técnicas. As frequências alélicas apresentam diferenças significativas entre populações geograficamente distintas, sendo potenciais marcadores para estudo de ancestralidade. Apresentam menor taxa de mutação em comparação com os STRs (Figura 20.8)[22].



Figura 20.8 Esquema comparando STRs com INDELs e mostrando os respectivos tamanhos em pares de bases (pb) obtidos na amplificação.

O fato de o fragmento de DNA amplificado ser pequeno é muito interessante, facilitando a análise em amostras biológicas não mantidas em boas condições de preservação, por exemplo, ossos e dentes coletados em casos de exumação de cadáveres ou de acidentes em massa. Nessas situações, a alternativa é a combinação de marcadores STRs e INDELs para a obtenção de uma análise mais conclusiva e robusta.

Em testes de paternidade ou outra relação biológica, a maioria dos laboratórios utiliza um conjunto de dezesseis marcadores STRs que oferecem um poder discriminatório alto e geram um resultado de exclusão de paternidade (entre 0,9999983 e 0,9999998) ou fortemente indicativo de probabilidade

de inclusão (índices de paternidade entre 522.000 e 4.110.000). No entanto, alguns casos apresentam resultados ambíguos ou inconclusivos, em que a relação reivindicada não pode ser confirmada por uma probabilidade suficientemente alta, ou uma exclusão é sugerida por incompatibilidade alélica em apenas um ou dois destes marcadores, sendo que para uma análise resultar em exclusão de paternidade há a necessidade de não coincidência entre o alelo paterno obrigatório (filho/a) e o alelo presente no suposto pai em, no mínimo, três marcadores autossômicos[23]. Na Figura 20.9, estão representados os eletroferogramas parciais de três regiões de STRs autossômicos, mostrando duas exclusões. É importante notar que incompatibilidades podem também ser identificadas por ação de mutações, uma vez que os STRs apresentam elevada taxa de mutação (cerca de 2×10^{-3}).

Resultados ambíguos surgem frequentemente, quando, por exemplo, o irmão biológico de primeiro grau do pai biológico é indicado como suposto pai de uma criança, de modo que a taxa de exclusão torna-se muito reduzida (poucas não coincidências nos alelos) e o índice de paternidade, usando uma razão de probabilidade de um homem aleatório na população, não se aplica; estes resultados são difíceis de serem interpretados. Um recurso encontrado é a análise de STRs adicionais para melhorar a probabilidade ou fornecer claras e inequívocas exclusões; no entanto, além daqueles que já são utilizados nos kits comerciais, poucos STRs são validados para que possam ser aplicáveis[24].

Estes resultados também são observados nos casos em que o suposto pai é falecido e há a necessidade de analisar amostras provenientes de exumação, como ossos e dentes, uma vez que esse tipo de amostra pode apresentar baixa quantidade de DNA e alto nível de degradação e contaminação, o que na grande maioria das vezes dificulta a análise de um número adequado de STRs. Nesses casos, analisando-se um número relativamente grande de INDELs (de 30 a 40), que apresentam menor taxa de mutação em comparação aos STRs, pode-se encontrar novas incompatibilidades, o que favorece a hipótese de exclusão, ou, ao contrário, caso não sejam encontradas incompatibilidades, aumentar o índice de paternidade a valores que suportem a confirmação de uma relação biológica.

20.5 TÉCNICAS PARA ANÁLISE DE INDELS AUTOSSÔMICOS

A técnica para análise de INDELs também é uma reação da PCR em multiplex, conforme mencionado anteriormente para amplificação de STRs.

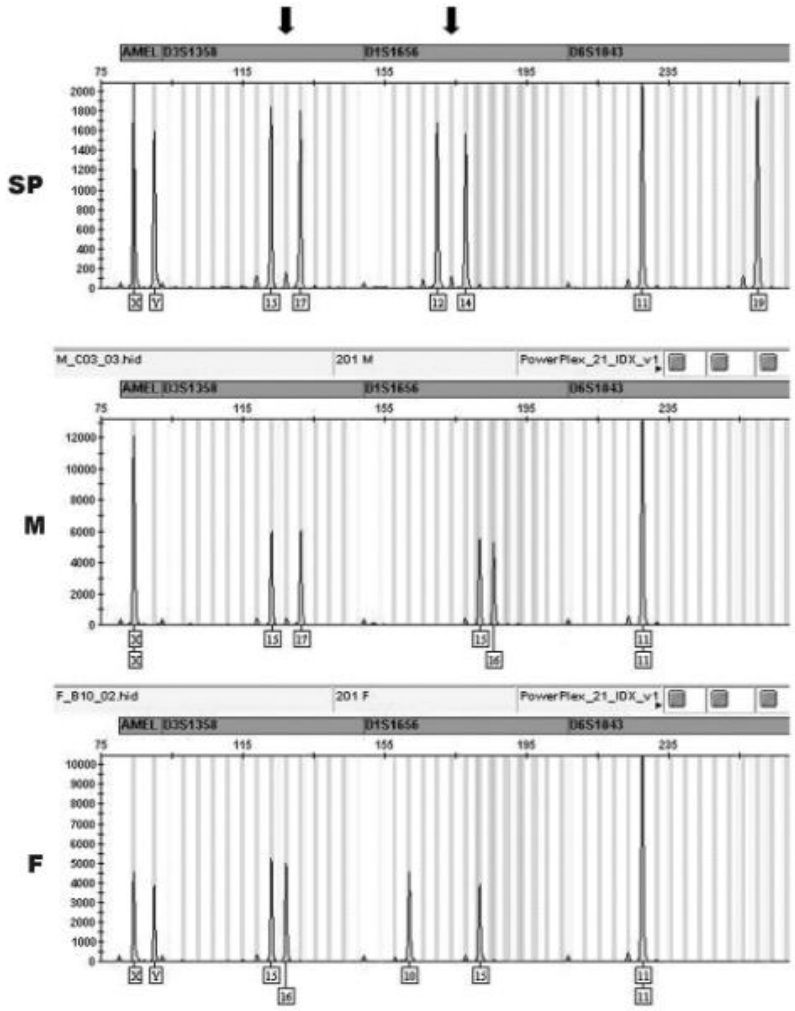


Figura 20.9 Comparação de três eletroferogramas parciais no *software* GenneMapper ID-X para verificar compatibilidade de alelos em um caso de investigação de paternidade. SP, suposto pai; M, mãe; F, filho. As setas indicam dois marcadores onde não houve compatibilidade, ou seja, o F não herdou o alelo obrigatório do SP.

Podem ser analisados trinta marcadores utilizando-se o kit *InvestigatorDIPplex* (QIAGEN), que também amplifica a amelogenina em uma única reação multiplex[25] (Figura 20.10).

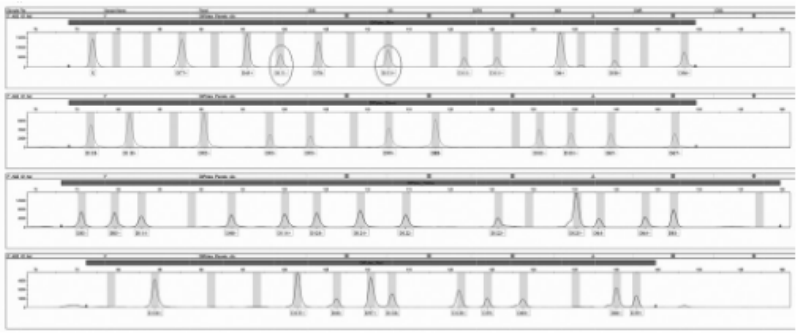


Figura 20.10 Eletroferograma da reação com o kit *InvestigatorDIPplex* (QIAGEN). (Trata-se de uma amostra feminina.) Os picos representam os alelos obtidos para cada marcador, sendo o alelo curto o que apresenta a deleção e o alelo longo o que apresenta a inserção, conforme exemplificado na Figura 20.3. Neste eletroferograma, os alelos são encontrados em regiões separadas, sendo divididos entre as quatro cores de fluoróforos. Estão identificados dois alelos de um mesmo marcador, separados no eletroferograma.

Devido ao fato de que os INDELs apresentam menor taxa de recombinação em comparação com os STRs, o número de INDELs analisados deverá ser maior para atingir um poder de discriminação aceitável; neste caso, os trinta polimorfismos são suficientes segundo o fabricante.

Outra reação foi padronizada por Pereira e colaboradores (2009)[15] e permite analisar 38 INDELs autossômicos em multiplex (Figura 20.11). Segundo os autores, esses 38 polimorfismos são suficientes para atingir um poder de descriminação aceitável.

Comparando-se as duas reações, verifica-se que a reação de Pereira e colaboradores (2009)[15] apresenta oito polimorfismos a mais do que aquela do kit *InvestigatorDIPplex* (QIAGEN), não havendo nenhum polimorfismo em comum. Em contrapartida, a reação de Pereira e colaboradores (2009)[15] não amplifica a amelogenina.

Outra avaliação interessante nos eletroferogramas demonstra que, nos 38-INDELs, os marcadores podem apresentar os alelos na mesma região, enquanto no *InvestigatorDIPplex* (QIAGEN), os alelos dos marcadores se

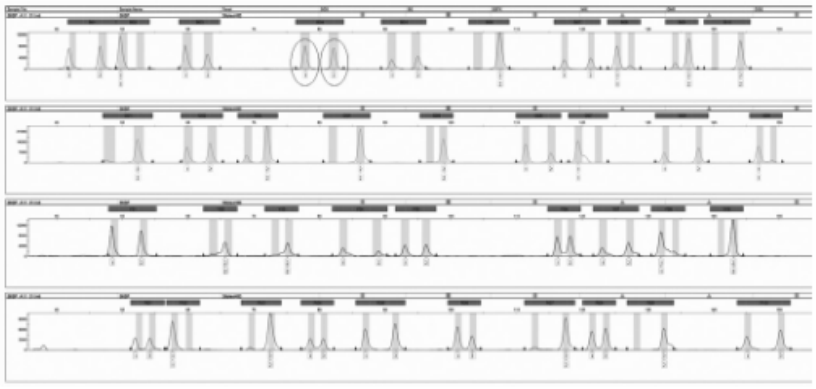


Figura 20.11 Eletroferograma da reação padronizada por Pereira e colaboradores (2009)[15]. (Amostra diferente da apresentada na Figura 20.10.) Os picos representam os alelos obtidos para cada marcador, sendo o alelo curto o que apresenta a deleção e o alelo longo o que apresenta a inserção, conforme exemplificado na Figura 20.3. Ao contrário da Figura 20.10, neste eletroferograma os alelos podem ser identificados na mesma região.

intercalam na mesma cor de fluoróforo. Isto dificulta a análise e a comparação entre os alelos de um mesmo marcador. A análise de amostras é realizada pelo *software* GeneMapper.

Como apresentado na Figura 20.12, os mesmos passos são realizados para análise de INDELs por ambas as reações multiplex, sendo que o passo de purificação pode ser opcional conforme a necessidade (quando o eletroferograma não discrimina adequadamente os alelos). De maneira geral, a metodologia para análise de INDELs é prática e rápida e muito similar à metodologia para análise de STRs, o que agrega praticidade ao trabalho em laboratórios de identificação humana.

20.6 APLICAÇÕES DO DNA MITOCONDRIAL

As mitocôndrias são organelas citoplasmáticas que apresentam em sua composição um genoma extracromossômico separado e distinto do genoma nuclear, o DNA mitocondrial (DNAmt), composto por 16.569 nucleotídeos e apresentando-se como uma dupla fita circular. Sua composição química



Figura 20.12 Fluxograma para análise de INDELs por ambas as reações. (O passo de purificação é opcional.)

em nada difere da composição do DNA nuclear; entretanto, ele possui um código genético próprio.

O DNAmt evoluiu pelo acúmulo de mutações nas linhagens maternas e acredita-se fixar novas mutações dez vezes mais rápido do que o DNA nuclear[26]. Este fato ocorre devido à baixa fidelidade da DNA polimerase mitocondrial, juntamente com a falta de mecanismos de reparo durante sua replicação, o que acaba sendo de grande interesse no campo forense.

Em alguns casos pode não ser possível a análise do DNA nuclear de algumas amostras (casos em que o DNA apresente degradações ou quando o material biológico não apresentar DNA nuclear), e então a melhor alternativa é a análise do DNA mitocondrial. Assim, o DNAmt é utilizado na rotina forense para analisar ossos, dentes, fios de cabelos e outros tipos de amostras biológicas, sempre que se necessite identificar o indivíduo.

A principal vantagem do DNAmt é que ele está presente em uma média de quinhentas a 2 mil cópias por célula. Esta abundância aumenta significantemente as chances de que algumas cópias de DNAmt permanecerão estáveis em ambientes que apresentam condições não favoráveis à sua conservação, além do que a natureza circular do DNAmt o torna menos suscetível à degradação por exonucleases[27].

O genoma mitocondrial possui herança estritamente materna, permitindo investigar relações de parentesco nesta linhagem, sendo, portanto, haploide e, por isso, não é submetido a processos de recombinação. Por essas duas razões, o DNAmt não pode ser utilizado na investigação de paternidade. Vale ressaltar que tanto os filhos como as filhas compartilham o DNAmt da mãe biológica.

Este genoma possui natureza monoclonal, ou seja, todo o DNAmt de um indivíduo apresenta a mesma sequência, o que simplifica a interpretação

dos resultados durante o sequenciamento do DNAmt. Porém, em alguns casos pode haver indivíduos que apresentam uma condição denominada de heteroplasmia, na qual uma mesma pessoa apresenta mais de um tipo de DNAmt, podendo ser heteroplasmia de sequência, quando há diferença em apenas um nucleotídeo, ou heteroplasmia de comprimento, quando há deleção ou inserção de uma citosina em regiões com repetições mononucleotídicas (conhecidas como poli-C). Nos dois casos, as análises devem ser muito cuidadosas para a interpretação da heteroplasmia. Portanto, há necessidade de treinamento do perito para análise do DNAmt em casos forenses.

20.7 UTILIZAÇÃO DO DNA MITOCONDRIAL EM ANÁLISES FORENSES

O genoma mitocondrial é composto de duas regiões (Figura 20.13): a região codificadora e a região controle, esta também conhecida como região não codificadora, região hipervariável ou *D-loop*, que controla a replicação e transcrição do DNAmt, a qual, devido ao acúmulo de mutações pontuais, apresenta uma maior taxa de mutação quando comparada à região codificadora.

A região controle é formada por aproximadamente 1.200 pares de base e, por ser altamente polimórfica, é muito utilizada para o propósito da genética forense. Esta região do DNAmt inclui três sub-regiões denominadas HV1 (formada por 342 pb), HV2 (268 pb) e HV3 (137 pb), sendo esta última região um segmento que pode apresentar um tamanho variável de pares de base, pelo fato de conter repetições dinucleotídicas "CA", cujo número pode variar de um indivíduo para outro[28]. Entre essas regiões existem outras duas denominadas inter-regiões, a primeira entre HV1 e HV2 e a segunda entre HV2 e HV3. A maior parte da variação de sequência entre indivíduos é encontrada nos segmentos HV1 e HV2. Por isso, estes são os mais utilizados na rotina forense para análise do DNAmt. Entretanto, a região HV3 também pode auxiliar as análises[29].

A identificação baseia-se na análise dos polimorfismos presentes nesses segmentos (SNPs, inserções e deleções/*indels*), que são comparados a uma sequência de referência (*revised Cambridge Reference Sequence – rCRS*), sequenciada em 1981 por Anderson e colaboradores[30], reanalisada e revisada por Andrews e colaboradores em 1999[31]. No entanto, tais regiões oferecem um limitado poder de discriminação em um contexto forense[32]. Assim, a região HV3, junto com as inter-regiões, vem sendo explorada nos últimos

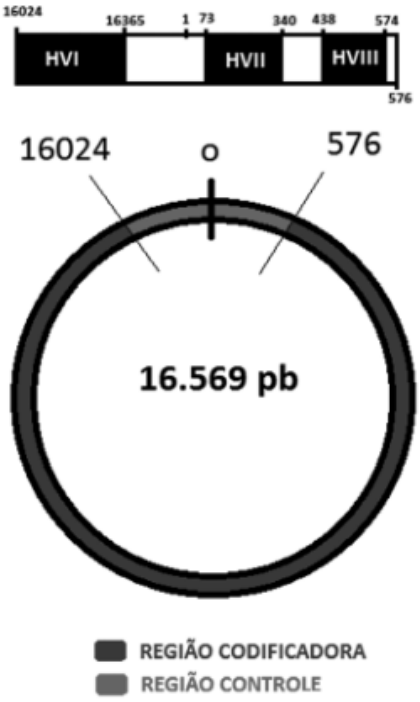


Figura 20.13 Esquema do DNA mitocondrial: região controle, com três sub-regiões (HV1, HV2 e HV3) e região codificadora.

anos, pois apesar de apresentarem baixos números de polimorfismos, informações complementares são obtidas quando adicionadas às regiões HV1 e HV2, aumentando o poder discriminativo, o que é importante na solução de casos forenses e, principalmente, em estudos de frequência populacional[33, 34, 35]. Essas características fazem com que o sequenciamento completo da região não codificadora do DNAmt seja mais adequado tanto para testes de identidade forense como para estudos evolucionários.

O sequenciamento da região controle do DNAmt pode ser feito a partir de DNA obtido de diversos materiais biológicos, como ossos, dentes, tecidos, sangue, sêmen e fios de cabelo. A partir do DNA extraído, realiza-se uma PCR utilizando os *primers* (*forward* e *reverse*) que flanqueiam as regiões de

interesse, podendo ser toda a região não codificadora ou cada uma das suas sub-regiões.

Os produtos da PCR são, então, purificados para retirada do excesso de *primers* e nucleotídeos que não foram incorporados durante a PCR, e submetidos à reação de sequenciamento. Os produtos obtidos a partir do sequenciamento também deverão ser purificados previamente à eletroforese capilar realizada em analisador genético (ABI3500 – Applied Biosystems). A análise dos resultados é realizada com a ajuda de *softwares* de bioinformática, e as sequências obtidas são comparadas com a rCRS.

Outra técnica de análise do DNAmt é a técnica de SNaPshot. Essa reação baseia-se na análise de polimorfimos de base única (SNPs), que representam a maior classe de polimorfismos humanos. Essa técnica está em uso crescente devido à detecção rápida de vários polimorfismos em um único ensaio. A capacidade multiplex é particularmente importante, especialmente no que diz respeito à análise de DNA forense, uma vez que pode reduzir o consumo de amostras, enquanto aumenta o rendimento do processamento das amostras e análise de dados. Esta técnica é realizada a partir da utilização de um kit comercial (ABI PRISM SNaPshot multiplex kit – Applied Biosystems) produzido pela empresa Life Technologies. A análise de SNPs do DNAmt, tanto da região não codificadora como da região codificadora, permite a classificação de amostras em haplogrupos.

20.8 APLICAÇÕES DO SEQUENCIAMENTO DO DNAMT

Alguns fatos importantes comprovam como a análise do DNAmt é de grande importância no contexto forense: foi a primeira ferramenta utilizada na identificação de ossadas de soldados americanos que lutaram na Guerra do Vietnã[36] e na identificação das vítimas da tragédia de 11 de setembro de 2001 no World Trade Center (EUA).

Outro exemplo interessante da utilização do DNA mitocondrial foi realizado na década de 1990 em um estudo de parentesco, no qual cientistas americanos foram convidados a examinar ossos e dentes de oito esqueletos de uma vala na Sibéria, onde se acreditava estar enterrada a família real russa Romanov, que fora executada pelos bolcheviques. Neste caso, os resultados do sequenciamento do DNAmt elucidou que havia vínculo genético entre as ossadas[37].

A utilização do DNAmt estendeu-se para estudos da evolução humana, migrações e formação da história de diversas populações, não permanecendo

apenas na rotina forense. O emprego do DNAmt como ferramenta para investigações desta natureza surgiu com os estudos pioneiros do grupo liderado por Allan Wilson nos anos 1980[38].

20.9 BANCO DE DADOS

A análise do DNAmt é rotineiramente utilizada em diversos laboratórios como ferramenta em análises forenses e estudos genéticos evolutivos. No entanto, é preciso um conhecimento aprofundado das várias características do DNAmt a fim de analisar e interpretar corretamente os resultados. A forma correta de interpretar um resultado de DNAmt ainda não foi padronizada, existindo apenas recomendações, como, por exemplo, aquelas adotadas pelo Federal Bureau of Investigation (FBI)[39] ou publicadas pelo Grupo de Habla Española y Portuguesa – International Society for Forensic Genetics (GHEP-ISFG)*.

A maioria dos laboratórios que utilizam tipagem do DNAmt baseiam-se nos polimorfismos (SNPs, deleções e inserções) presentes na sequência de nucleotídeos nas regiões HV1 e HV2, comparando a amostra questionada com a sequência referência (rCRS) para a anotação das diferenças (polimorfismos)[29].

De acordo com o FBI, se essas sequências diferirem em dois ou mais polimorfismos, uma exclusão inequívoca pode ser feita. Se as sequências corresponderem uma à outra, existe a possibilidade da amostra questionada pertencer ao indivíduo analisado ou demonstrar vínculo materno, podendo, no entanto, pertencer a outro indivíduo não aparentado; por isso, torna-se necessário verificar a frequência com que um conjunto de polimorfismo ocorre em uma população (frequência do haplótipo), gerando a necessidade da criação de um banco de dados populacional de DNAmt.

O termo "haplogrupo" refere-se ao conjunto de haplótipos (conjunto de alelos), que são designados por letras do alfabeto e derivados por descendência de uma mesma molécula ancestral, as quais apresentam um padrão basal de mutações. Os haplogrupos são definidos por um conjunto de haplótipos que são constituídos de nucleotídeos particulares presentes nas regiões controle e codificadora do DNAmt. A relação entre haplótipos e haplogrupos do DNAmt humano nos fornece informações mais detalhadas sobre a estrutura da variação genética. No site PhyloTree** encontra-se a árvore filogené-

* Ver http://www.gep-isfg.org.
** Ver www.phylotree.org.

tica mitocondrial humana que está representada de modo simplificado na Figura 20.14.

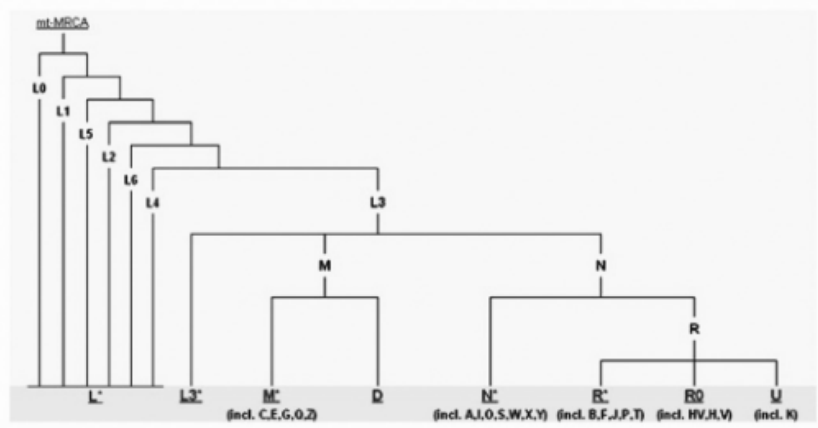


Figura 20.14 Árvore filogenética humana simplificada do DNA mitocondrial, mostrando os principais haplogrupos.

A criação de banco de dados de DNAmt para estimar a frequência dos haplótipos em determinada população é indispensável para permitir sua utilização. Entretanto, pode estar sujeito a vários tipos de erros, principalmente humanos (na leitura e digitação de dados, por exemplo), havendo a necessidade de se estabelecerem regras de controle de qualidade tanto na geração quanto na compilação de resultados[29]. Atualmente, existem ferramentas que podem ser utilizadas para que esses erros humanos sejam evitados como, por exemplo, MITOMAP (*mtDNA control region sequence polimorphisms*)* e IAN LOGAN**, as quais permitem comparar se todos os polimorfismos encontrados já foram descritos na literatura.

Diversos pesquisadores geram milhares de dados de genomas mitocondriais completos de seres humanos em todos os continentes, bem como uma nomenclatura detalhada dos tipos de DNAmt ou haplogrupos em todo o mundo e uma enorme árvore filogenética dos haplogrupos observados, sub-haplogrupos e macro-haplogrupos[40]. No campo da identificação humana,

* Ver http://www.ianlogan.co.uk/mtDNA.htm.
** Ver http://www.ianlogan.co.uk/mtDNA.htm.

as revistas científicas de maior impacto já exigem que os dados de trabalhos envolvendo análises de DNAmt ou outros tipos de análises gerado por pesquisadores sejam adicionadas aos bancos de dados para serem publicados.

Um exemplo de banco de dados que foi desenvolvido com rigorosos parâmetros de qualidade é o banco de dados mundial de DNAmt denominado EMPOP. Para manter um padrão de qualidade, é necessário ter uma padronização da geração, análise e transferência de dados, além do uso de *softwares* para a visualização de dados, interpretação e detecção de erros.

20.10 TÉCNICA PASSO A PASSO

Protocolo para a análise de 38 INDELs autossômicos

Extração de DNA

A extração do DNA do interior das células é o primeiro passo para sua análise. Descreveremos a seguir um protocolo[41] para a extração de DNA de sangue armazenado em papel de filtro FTA (Whatman), utilizando a resina Chelex 100 (Biorad) a 5%:

µ

1) Colocar em um tubo de 1,5 mL 2 discos (3 mm cada) de papel de filtro com sangue.
2) Adicionar 50 µL de H_2O, vortexar rapidamente e descartar a H_2O.
3) Adicionar 50 µL de H_2O e incubar a temperatura ambiente por 30 minutos.
4) Descartar a H_2O e adicionar 100 µL da solução homogeneizada da resina chelex 100 (Biorad) a 5%.
5) Incubar por 1 hora a 56 °C.
6) Vortexar rapidamente e incubar por 8 miunutos a 100 °C.
7) Centrifugar a 14.000 RPM por 3 minutos.
8) Armazenar o sobrenadante.

Reação em cadeia da polimerase (PCR)

Segue a descrição do preparo da PCR para amplificação dos 38 INDELs padronizado por Pereira et al. (2009)[15], utilizando o *Multiplex PCR kit* (QIAGEN):

H_2O	3 µL
Qiagen *Multiplex PCR master mix* (2×)	5 µL
Primer mix (10×)	1 µL
DNA 0,3 a 5 ng/µ L	1 µL
Volume total	10 µL

Em seguida, é utilizada a ciclagem em termociclador:

95 °C – 15 min.	
94 °C – 30 seg.	
60 °C – 90 seg.	10 ciclos
72 °C – 60 seg.	
94 °C – 30 seg.	
58 °C – 90 seg.	20 ciclos
72 °C – 60 seg.	
72 °C – 60 min.	
4 °C – até a retirada do termociclador.	

Observação: um passo adicional de purificação do produto da reação de PCR pode ser utilizado se necessário.

Análise do produto da reação de PCR

O produto da PCR deve ser preparado para eletroforese capilar pela adição de 1 µL do produto amplificado em 9,7 µL de formamida HI-DI (Applied Biosystems) e 0,3 µL de GeneScan 600 LIZ (Applied Biosystems). A separação e detectção são realizadas por eletroforese capilar.

20.11 CONCLUSÕES

As ferramentas para identificação humana utilizando o DNA são, atualmente, bastante eficazes na discriminação entre indivíduos e/ou vestígios e na investigação de paternidade, gerando índices de verossimilhança bastante elevados. Entretanto, cabe ressaltar que o exame de DNA na genética

forense constitui-se em uma evidência a mais na elucidação dos casos do Poder Judiciário, favorecendo ou não a hipótese de acusação ou de defesa. O perito deve analisar os resultados obtidos de maneira cuidadosa, isenta e ética, relatando as evidências científicas encontradas através das análises dos perfis genéticos de maneira clara e objetiva, valorando os dados com os cálculos estatísticos adequados, mas deixando ao juiz a tarefa de avaliação final da importância de tais evidências no julgamento do referido caso.

REFERÊNCIAS

1. França GV. Medicina Legal. 6. ed. Rio de Janeiro: Guanabara-Koogan; 2001. 579 p.
2. Jeffreys AJ, Wilson V, Thein SL. Hypervariable minisatellite regions in human DNA. Nature. 1985;314:67-72.
3. Jeffreys AJ, Wilson V, Thein SL. Individual specific "fingerprints" of human DNA. Nature. 1985;316:75-79.
4. Wyman AR, White R. A highly polymorphic locus in human DNA. Proc Natl Acad. Sci USA. 1980;77:6754-8.
5. Jeffreys AJ. Genetic fingerprinting. Nature Medicine. 2005;11(10):XIV-XVIII.
6. Saiki RK, Scharf S, Faloona F, Mullis K, Horn G, Erlich HA, et al. Enzymatic amplification of ß-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science. 1985;230:1350-4.
7. Saiki RK, Gelfand DH, Stoffel S, Scharf S, Higuchi RH, Horn GT, et al. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science. 1988;239:487-91.
8. Mullis K, Faloona F. Specific synthesis of DNA in vitro via a polymerase catalyzed chain reaction. Method Enzymol. 1987;155: 335-50.
9. Watson JD. DNA: o segredo da vida. São Paulo: Companhia das Letras; 2005. 470 p.
10. The FBI Federal Bureau of Investigation. Combined DNA Index System (Codis) [Internet]. Available from: http://www.fbi.gov/about-us/lab/biometric-analysis/codis.
11. Consultor Jurídico. Técnica usada na tipagem genética não está livre de erros [Internet]. 2006 Jun 13. Available from: http://www.conjur.com.br/2006-jun-13/tecnica_usada_tipagem_genetica_nao_livre_erros.
12. Paradela ER, Figueiredo AL dos S; Smarra ALS. A identificação humana por DNA: aplicações e limites [Internet] [Cited Nov 2013]. In: Âmbito Jurídico, Rio Grande, IX, n. 30, jun 2006. Available from: http://www.ambito-juridico.com.br/site/index.php?artigo_id=1175&n_link=revista_artigos_leitura.
13. Houck MM. A realidade do CSI. Scientific American Brasil. Available from: http://www.uol.com.br/sciam/reportagens/a_realidade_do_csi.html.
14. Bassette F. País padroniza técnicas para perícia. Folha de S. Paulo. 27 set 2010:Cotidiano.
15. Pereira R, Phillips C, Alves C, Amorim A, Carracedo A, Gusmao L. A new multiplex for human identification using insertion/deletion polymorphisms. Electrophoresis. 2009;30:3682-90.
16. Weber JL, David D, Heil J, Fan Y, Zhao C, Marth G. Human diallelic insertion/deletion polymorphisms. Am J Hum Genet. 2002;71(4):854-862.

17. Mills RE, Luttig CT, Larkins CE, Beauchamp A, Tsui C, Pittard WS, Devine SE. An initial map of insertion and deletion (INDEL) variation in the human genome. Genome Res. 2006;16(9):1182-90.
18. Butler J. Advanced Topics in Forensic DNA Typing: Methodology. 1st ed. Waltham/San Diego/London:Academic Press; 2011. 704 p.
19. Hochmeister MN, Budowle B, Borer UV, Eggmann U, Comey CT, Dirnhofer R. Typing of deoxyribonucleic acid (DNA) extracted from compact bone from human remains. J Forensic Sci. 1991;36:1649-61.
20. Loreille OM, Diegoli TM, Irwin JA, Coble MD, Parsons TJ. High efficiency DNA extraction from bone by total demineralization. Forensic Sc International. 2007;1:191-5.
21. Zidkova A, Horinek A, Kebrdlova V, Korabecna M. Application of the new insertion–deletion polymorphism kit for forensic identification and parentage testing on the Czech population. Int J Legal Méd. 2011;127(1):7-10.
22. Phillips C, Fondevila M, Garcia-Margarinos M, Rodriguez A, Salas A, Carracedo A, Lareu MV. Resolving relationship tests that show ambiguous STR results using autosomal SNPs as supplementary markers. Forensic Science International Genetics. 2008;2:198-204.
23. Mullaney JM, Mills RE, Pittard WS, Devine SE. Small insertions and deletions (INDELs) in human genomes. Hum Mol Genet. 2010;19(2):131-6.
24. Yang N, Li H, Criswell LA, Gregersen PK, Alarcon-Riquelme ME, Kittles R, et al. Examination of ancestry and ethnic affiliation using highly informative diallelic DNA markers: application to diverse and admixed populations and implications for clinical epidemiology and forensic medicine. Hum Genet. 2005;118(3-4):382-92.
25. Larue BL, Ge J, King JL, Budowle B. A validation study of the Qiagen Investigator DIPplex® kit; an INDEL-based assay for human identification. Int J Legal Med. 2012;126(4):533-40.
26. Wallace DC, Ye J, Neckelmann SN, et al. Sequence Analysis of cDNAs for the Human and Bovine ATP Synthase Beta Subunit: Mitochondrial Genes Sustain Seventeen Times more Mutations. Current Genetics. 1987;12:81-90.
27. Kashyap VK, Sitalaximil T, Chattopadhyay P, Trivedi R. DNA Profiling Technologies in Forensic Analysis. International Journal of Human Genetics. 2004;3: 11-30.
28. Lutz S, Wittig H, Weisser HJ, Heizmann J, Jungle A, Simonin ND, et al. Is it possible to differentiate mtDNA by means of HVIII in samples that cannot be distinguished by sequencing the HVI and HVII regions? Forensic Science International. 2000;113:97-101.
29. Paneto GG, et al. Heteroplasmy in Hair: Study of Mitochondrial DNA Third Hypervariable Region in Hair and Blood Samples. J Forensic Sci. 2010;55:715-7.
30. Anderson S, Bankier AT, Barrel BG, et al. Sequence and organization of the human mitochondrial genome. Nature. 1981;290:457-65.

31. Andrews R, et al. Reanalysis and revision of Cambridge reference sequence for human mitochondrial DNA. Nat Genet. 1999;23:147.
32. Huang D, Gi C, Yi S, et al. Typing of 24 mtDNA SNPs in Chinese Population Using SNaPshot Minisequencing. Med Sci. 2010;30:291-98.
33. Bini C, et al. Different informativeness of the three hypervariable mitochondrial DNA regions in the population of Bologna (Italy). Forensic Sci Int. 2003;135:48-52.
34. Vanecek T, Vorel F, Sip M. Mitochondrial DNA D-loop hypervariable regions: Czech population data. Int J Legal Med. 2004;118:14-8.
35. Zhang YJ, et al. Haplotype diversity in mitochondrial DNA hypervariable region I, II and III in northeast China Han. Forensic Sci Int. 2005;149:267-9.
36. Holland MM, Fisher DL, Mitchell LG. Mitochondrial DNA sequence analysis of human skeletal remains: identification of remains from the Vietnam war. J Forensic Sci. 1993;38:542-53.
37. Farah SB. DNA segredos e mistérios. 2 ed. São Paulo: Sarvier; 2007.
38. Wilson AC, et al. Mitochondrial DNA and two perspectives on evolutionary genetics. Biol J Linn. 1985;6:375-400.
39. Scientific Working Group on DNA Analysis Methods (Swgdam). Guidelines for Mitochondrial DNA (mtDNA) Nucleotide Sequence Interpretation. Forensic Science Communications. 2003. 5 p.
40. Bandelt HJ, Macaulay V, Richards M. Human mitochondrial DNA and the evolution of Homo sapiens. 1st ed. Heidelberg: Springer; 2006. p.153.
41. Singer-Sam J, Tanguay R, Riggs AD. Use of chelex to improve the PCR signal from a small number of cells. Amplifications: A Forum for PCR Users. 1989;3:11.

3

BIOTECNOLOGIA APLICADA À SAÚDE

FUNDAMENTOS E APLICAÇÕES



organizador
RODRIGO RIBEIRO RESENDE

colaboradores
SILVIA GUATIMOSIM
MARCUS VINICIUS GOMEZ
CARLOS R. SOCCOL

Blucher

FAPEMIG

CAPÍTULO 20

APLICAÇÕES FORENSES DO DNA: FUNDAMENTOS, MÉTODOS E LIMITAÇÕES

Isabela Brunelli Ambrosio
Danilo Faustino Braganholi
Regina Maria Barretto Cicarelli

20.1 INTRODUÇÃO

A identificação humana é o processo científico por meio do qual se determina a identidade de uma pessoa, através de um conjunto de caracteres que a individualize, fazendo-a igual apenas a si mesma[1]. Esta identificação por meio do DNA constitui um dos produtos mais revolucionários da genética moderna.

A genética forense, também conhecida como DNA Forense, é a área que utiliza os conhecimentos e técnicas de genética e de biologia molecular no auxílio à justiça, sendo aceita rotineiramente em processos judiciais em todo o mundo.

Apesar de ser a identificação humana pelo DNA o ramo mais desenvolvido da genética forense, bem como sua aplicação mais popular ser o teste de paternidade, ela não se limita a isso, e pode ser aplicada na identificação ou individualização de animais, plantas e micro-organismos.

A genética forense iniciou-se quando foram utilizadas pela primeira vez características genéticas para a definição de paternidade, ajudando a justiça na solução de um caso criminal[2, 3]. Sua fase moderna teve início na década de 1980, quando pesquisadores descobriram regiões altamente variáveis do DNA, capazes de individualizar uma pessoa[2, 3, 4]. Em 1985, sir Alec Jeffreys[3] apelidou as características únicas do DNA de uma pessoa de "impressões digitais do DNA". O perfil genético de um indivíduo, comumente utilizado na identificação humana, é baseado na combinação de diversos marcadores que são herdados de seus progenitores. Esses marcadores referem-se geralmente a diferenças nas sequências de DNA entre os indivíduos (polimorfismos).

Os marcadores genéticos utilizados em identificação humana podem ser biparentais (metade dos cromossomos herdados da mãe e a outra metade herdada do pai) ou uniparentais (cromossomo Y herdado apenas do pai em indivíduos do sexo masculino e DNA mitocondrial herdado apenas da mãe em todos os indivíduos). Os marcadores biparentais autossômicos permitem uma identificação individual porque um determinado perfil genético é específico de um só indivíduo, sendo apenas partilhado por gêmeos monozigóticos. Os marcadores uniparentais não permitem a identificação individual, pois todos os indivíduos aparentados por via masculina partilharão a mesma linhagem de cromossomo Y, enquanto todos os indivíduos aparentados por via feminina partilharão a mesma linhagem de DNA mitocondrial, salvo nos casos de ocorrência de mutações durante a transmissão.

20.2 POLIMORFISMOS STRS (AUTOSSÔMICOS E DOS CROMOSSOMOS SEXUAIS, X E Y)

Desde que se desenvolveu a tecnologia do DNA recombinante, em 1973, dividindo em pedaços o DNA genômico para estudar os genes, houve um avanço na biologia molecular, e diversos genes foram descritos e isolados, muitos deles relacionados com a identificação humana[5]. Com o uso de enzimas de restrição e sondas de DNA, foram exploradas as variabilidades ou polimorfismos do DNA que variavam de indivíduo para indivíduo. Descobriram-se, também, regiões constituídas por vários pares de bases, repetidas inúmeras vezes, com alelos variando quanto ao número de sequências, denominados minissatélites (*variable number tandem repeats* – VNTR). Inicialmente, os VNTRs foram utilizados como marcadores genéticos informativos, pois estes são compostos por cem ou mais alelos com diferentes comprimentos na população humana

Outros marcadores altamente polimórficos foram encontrados no genoma humano: os microssatélites (*short tandem repeats,* STR) (Figura 20.1). Os STRs são repetições curtas de 1 a 5 pares de bases encontrados em eucariotos. Tais repetições são encontradas tanto em regiões codificantes quanto não codificantes do DNA, cuja função ainda é desconhecida[5].

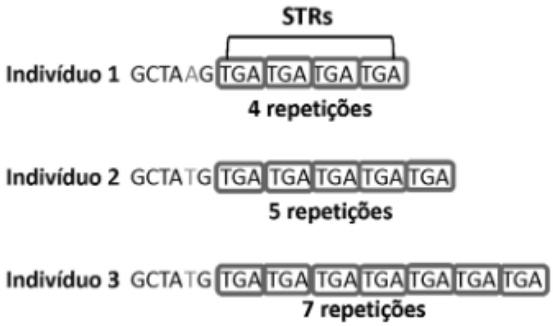


Figura 20.1 Esquema representando variações dos polimorfismos *short tandem repeats* (STRs). O número de repetições das pequenas sequências configuram um alelo que pode variar entre indivíduos.

Inovação tecnológica revolucionária na biologia molecular surgida em 1985, a reação da polimerase em cadeia ou *polymerase chain reaction* (PCR), tornou-se de uso universal. Por meio dessa reação, quantidades muito pequenas de DNA podem ser amplificadas para obtenção de praticamente qualquer quantidade desejada[6, 7, 8]. A técnica utiliza o mesmo princípio pelo qual o DNA é normalmente copiado na célula, exceto pelo fato de que somente um pequeno segmento do DNA no cromossomo será amplificado. Isso possibilitou processar quantidades muito pequenas de DNA frequentemente deixadas como evidência de um crime e aumentou enormemente a sensibilidade dos sistemas disponíveis usados nas análises forenses, auxiliando o sistema judiciário no julgamento de crimes. Graças à PCR, diminutas quantidades de DNA podem ser extraídas de fios de cabelo, selos postais, pontas de cigarro, xícaras de café e outras fontes de DNA.

Os STRs substituíram os polimorfismos de comprimento dos fragmentos de restrição (do inglês, *restriction fragment length polymorphisms* – RFLPs) como principal método de identificação genômica e tornaram-se popular na tipagem de DNA para análises forenses. São sequências curtas que variam de

duas a quatro bases, podendo repetir-se até dezessete vezes (de sete a quinze alelos por lócus), comumente amplificadas por PCR. Os métodos para tipagem de STRs estão atualmente automatizados, com detecção envolvendo fluoróforos[9].

Com o desenvolvimento de reações da PCR em multiplex (Figura 20.2), ou seja, amplificação de várias regiões do DNA em um único tubo, empresas internacionais iniciaram a comercialização de kits contendo um *mix* ou mistura com diversos iniciadores (*primers*), que permitem amplificar até 24 regiões do DNA genômico, sendo 23 *loci* STR mais a amelogenina para determinação do sexo. Entretanto, para aplicação e uso nos diferentes laboratórios forenses, há a necessidade de se conhecer as frequências dos respectivos alelos presentes na população daquela região geográfica em que os testes estão sendo realizados. Tais frequências devem ser publicadas em revistas indexadas para que possam ser utilizadas nas análises estatísticas dos resultados dos perfis que farão parte da conclusão do laudo pericial.

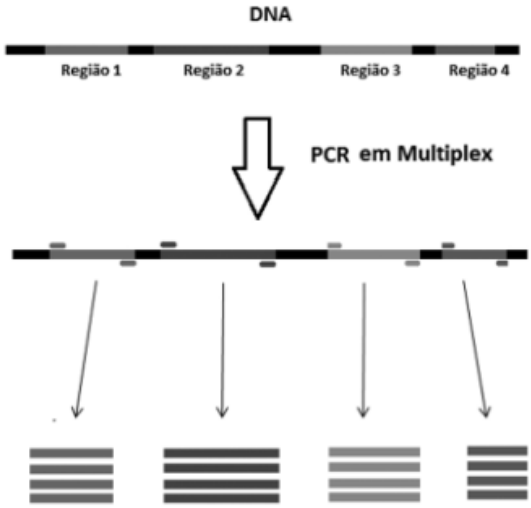


Figura 20.2 Esquema representando uma reação de PCR multiplex, na qual são amplificadas diferentes regiões do DNA em uma única reação.

Tratando-se de amostras obtidas em cenas de crimes, casos de violação sexual (estupro), desastres em massa ou de restos mortais de exumação de cadáver, uma questão fundamental, concernente ao uso do DNA como evidência, está na validação científica dos métodos de análise dos STRs. Em outras palavras, é preciso ter garantias científicas de que os testes podem, inequivocamente, identificar coincidências ou não coincidências para cada marcador genético utilizado, na comparação das amostras questionada e de referência. Inicialmente, a credibilidade dos testes deve partir da natureza das amostras biológicas utilizadas[2]. Muito frequentemente as amostras são encontradas em superfícies não esterilizadas, podendo sofrer danos após contato com a luz solar, micro-organismos e solventes. Existem procedimentos que podem minimizar a ação desses fatores de degradação do DNA. Entretanto, muitos cuidados devem ser tomados para evitar equívocos na interpretação dos resultados dos perfis alélicos. Ressalte-se que a cadeia de custódia das amostras deve ser mantida até o momento de serem processadas no laboratório.

A amplificação por PCR pode produzir falhas e artefatos devido à baixa qualidade do material biológico. Amostras de DNA parcialmente degradadas podem proporcionar, por exemplo, a amplificação preferencial de alguns alelos e/ou o surgimento das bandas fantasmas (*stutter bands*) (Figura 20.3)[2]. No primeiro caso, pode-se ter a amplificação de um alelo em detrimento do outro, o que pode gerar a falsa impressão de se tratar de um indivíduo homozigoto em vez de heterozigoto para o lócus em estudo. As bandas fantasmas ocorrem em virtude de falhas no processo de amplificação, gerando bandas com uma unidade de repetição a menos que a do alelo original. Desse modo, pode-se equivocadamente interpretar o resultado como um falso heterozigoto ou identificar um alelo erroneamente.

Apesar dessas dificuldades, a genética é a área da ciência forense que mais tem avançado na atualidade. Há alguns anos, a Secretaria Nacional de Segurança Pública do Brasil está implementando o Banco de Dados Nacional Criminal de Perfis Genéticos, similarmente ao americano CODIS (Combined DNA Index System), que armazena dados de criminosos condenados, e o europeu Fenix, que contém o perfil genético de milhares de pessoas desaparecidas. Tais ferramentas tornam mais ágil a troca de informações entre as instituições espalhadas pelo vasto território no país, como no caso do CODIS nos Estados Unidos da América[10], facilitando a resolução de diversos casos.

No Brasil, a implementação desse banco de dados trouxe o aumento da demanda nos laboratórios de perícia, uma vez que ele permite, por exemplo,

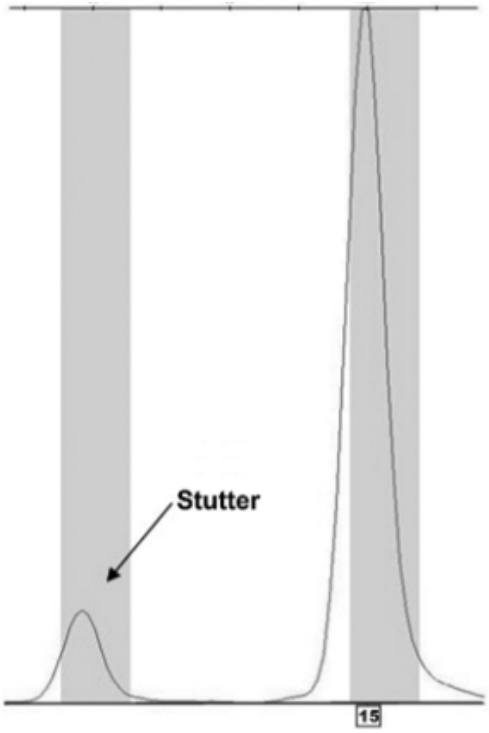


Figura 20.3 Esquema representando um alelo e um *stutter*. A seta indica o *stfuter*. Nota-se a grande diferença entre as intensidades de sinais entre *stutter* e alelo verdadeiro.

identificar o criminoso pela análise de uma única gota de sangue encontrada no local do crime.

Em 2006, alguns peritos judiciais publicaram artigos[11, 12] revelando as possibilidades de erros em exames de DNA, o que acarretou um convite da revista *Scientific American Brasil* para escrever um artigo sobre o estado da arte da genética forense no Brasil[13]. Em decorrência dessas publicações, vários laboratórios brasileiros começaram a ser condenados por erros em exames de DNA.

Na falta de um órgão nacional credenciado para promover a acreditação de laboratórios de DNA forense no Brasil, como é o caso, por exemplo, da American Association of Blood Banks (AABB) nos Estados Unidos, o

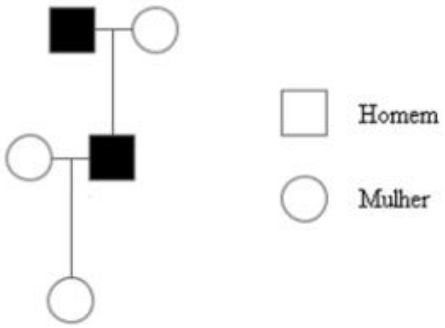
Ministério da Justiça solicitou que o Instituto Nacional de Metrologia, Qualidade e Tecnologia (INMETRO) viesse a padronizar os exames de DNA realizados por laboratórios da polícia científica[14]. Mais recentemente, a Agência Nacional de Vigilância Sanitária (ANVISA) publicou a Resolução RDC n. 11, de 16 de fevereiro de 2012, estabelecendo normas para o funcionamento de laboratórios que realizam testes e análises técnicas, o que abrange também os laboratórios que realizam investigação de paternidade.

Em associação aos STRs autossômicos, que muito contribuem para a identificação do indivíduo tanto quanto para a investigação de paternidade, os Y-STRs relacionados com o cromossomo Y, bem como os X-STRs do cromossomo X, são também ferramentas importantes na elucidação das análises de DNA forense ou casos de relação biológica. Os primeiros são importantes na determinação de linhagem patrilinear, mas não servem para incluir o indivíduo como o suposto pai da criança do sexo masculino, uma vez que o cromossomo Y é transmitido de pai para filho e deste para seu filho de maneira haplotípica, à semelhança da transmissão do DNA mitocondrial (herança materna), levando a que toda a descendência do sexo masculino, daquele primeiro indivíduo compartilhe os mesmos haplotipos Y-STRs; ou seja, todos os descendentes do sexo masculino apresentarão o mesmo padrão haplotípico do seu ancestral. Por essa razão, não se pode associar (ou multiplicar) os valores obtidos a partir das frequências dos alelos do perfil dos STRs autossômicos, denominados de índice de verossimilhança, com aquele da frequência do haplotipo dos Y-STRs. Entretanto, da análise do perfil Y-STR é possível excluir um indivíduo como pai biológico de uma criança do sexo masculino ou como doador de uma mancha de sangue ou sêmen.

Em relação ao cromossomo X (X-STRs), este é transmitido pelo pai para todas as filhas do sexo feminino, que recebem o outro cromossomo X da mãe biológica. Como neste caso o padrão de transmissão é muito semelhante aos STRs autossômicos, em geral os índices obtidos através de cálculos estatísticos utilizando as respectivas frequências dos alelos de ambos, STRs e X-STRs, podem ser multiplicados, aumentando, assim, as chances de favorecimento de paternidade, quando se trata de filha questionada na ausência do suposto pai, por exemplo (Figura 20.4).

A equipe da autora deste capítulo elaborou um banco de dados muito interessante e bastante completo para estudo de vários parâmetros de análise do cromossomo X*, os quais foram obtidos a partir das publicações nacionais sobre esse tema e de dados do próprio laboratório. Esforços estão sendo

* Ver http://bgbx.com.br. Número de registro de programa de *software*: 12166-1.



STRs autossômicos
Probabilidade de Relação Biológica: 71,79757%

X

X-STRs
Probabilidade de Relação Biológica: 99,9813421%

STRs autossômicos e X-STRs
Probabilidade de Relação Biológica: 99,9926702 %

Figura 20.4 Heredograma de um caso de investigação de paternidade em que foi utilizada a análise de X-STRs. Os indivíduos em preto não foram analisados. Foi utilizada neste caso a análise da mãe biológica do suposto pai, uma vez que este estava ausente. O esquema mostra a probabilidade de relação biológica com os STRs autossômicos e X-STRs; a multiplicação dos índices aumenta a probabilidade de relação biológica entre os indivíduos analisados.

investidos no sentido de disponibilizar também as fórmulas para realizar os cálculos estatísticos mais importantes a serem incluídos nos laudos periciais em que há a necessidade da associação dos STRs autossômicos e X-STRs, garantindo, desse modo, um índice de paternidade ou de verossimilhança mais robusto.

A utilização das análises dos marcadores de Y e X-STRs é importante em casos de confirmação de exclusão de paternidade e, especialmente, em casos

de reconstrução, quando o suposto pai está ausente. Em casos criminais, estas também são ferramentas importantes para elucidar a origem de manchas de sangue e/ou sêmen.

A técnica de amplificação de STRs (autossômicos, Y e X) baseia-se na realização de uma reação da PCR multiplex, isto é, na qual vários *primers* são adicionados a um mesmo tubo de 0,2 mL contendo um mix da PCR (tampão, cloreto de magnésio, dNTPs e enzima Taq polimerase) e acrescentando-se à mistura o DNA em estudo.

Duas reações para controle das condições de amplificação são necessárias e imprescindíveis: um controle positivo, contendo DNA humano conhecido e convenientemente diluído (concentração de uso 0,5 ng/uL), e outro negativo, geralmente realizado com água destilada esterilizada. Os tubos são levados ao termociclador para a amplificação em ciclagem adequada e compatível com a reação multiplex.

Atualmente, o uso de kits comerciais são os mais recomendáveis para as análises forenses e de paternidade por serem validados nacional e internacionalmente. A seguir, a Figura 20.5 apresenta um esquema dos procedimentos básicos para a reação.

O sucesso da análise dos STRs está diretamente relacionado com a qualidade do DNA obtido. O DNA está presente em todas as células nucleadas e, portanto, pode ser extraído de diferentes materiais biológicos deixados na cena de um crime. Em geral, as amostras testadas nos laboratórios forenses são tipicamente sangue e manchas de sêmen, mas podem existir muitas outras fontes de DNA.

Para a comparação dos perfis de DNA, sempre há a necessidade de amostras de referência (sangue ou *swab* oral). Todas as amostras devem ser cuidadosamente manuseadas para a extração de DNA, de modo a evitar qualquer tipo de contaminação. Por essa razão, muitos laboratórios processam as amostras de evidência (aquelas obtidas na cena do crime) em momentos diferentes ou mesmo em locais diferentes das amostras de referência. Existem diversos métodos de extração de DNA; entretanto, os mais utilizados atualmente são orgânico, Chelex e papel FTA. Todos eles permitem a obtenção de DNA de boa qualidade para análise de STRs. A Figura 20.6 ilustra um perfil alélico amplificado a partir de uma amostra de DNA humano.

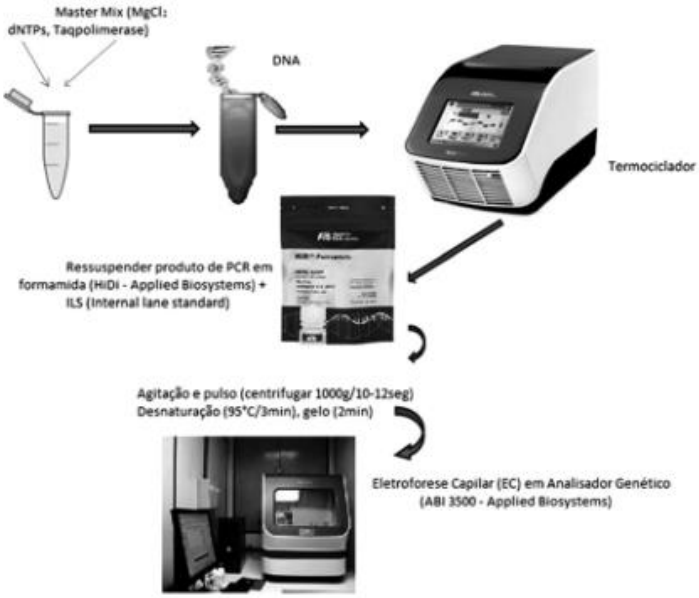


Figura 20.5 Esquema ilustrativo da reação de PCR multiplex para amplificação de STRs (autossômicos, Y ou X-STRs).

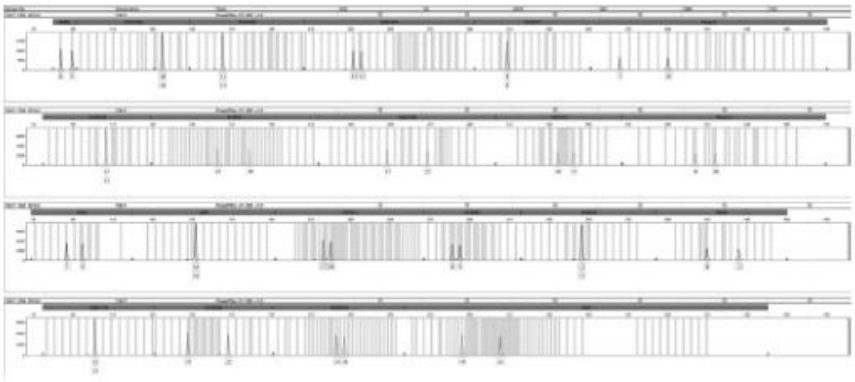


Figura 20.6 Eletroferograma obtido no analisador genético (ABI 3500 – Applied Biosystems) e analisado no GeneMapper ID-X apresentando os alelos amplificados na reação multiplex de 21 regiões do DNA humano (STRs autossômicos). A amostra pertence a um indivíduo do sexo masculino (X/Y). Os picos representam os alelos obtidos em cada marcador, que são repetições de pequenas sequências de bases, como mostrado na Figura 20.1.

20.3 POLIMORFISMOS DE INSERÇÃO/ DELEÇÃO (INDELS) E SUAS APLICAÇÕES

O estudo das variações genéticas, utilizando os polimorfismos do DNA, permite uma melhor compreensão da história e da diversidade das populações humanas, além de proporcionar um sistema para a identificação genética de indivíduos. A aplicação do DNA decorre do seu alto poder de discriminação gerado por seus polimorfismos, e além dos STRs já estudados anteriormente, destacam-se outros dois: *single nucleotide polymorphisms* (SNPs) e *insertion-deletion polymorphisms* (INDELs)[15].

Weber e colaboradores (2002)[16] foram pioneiros, pois identificaram e caracterizaram aproximadamente 2 mil INDELs no genoma humano, com diferentes tamanhos e frequências alélicas em europeus, africanos, japoneses e nativo-americanos, e devido à abundância e facilidade de análise, destacaram a sua utilidade para estudos genéticos. INDELs chegam a representar cerca de 16% de todos os polimorfismos do DNA humano e se apresentam muito distribuídos pelo genoma, em média, um INDEL a cada 7,2 Kb[17].

INDELs são polimorfismos de comprimento, caracterizados pela inserção ou deleção de um ou mais nucleotídeos em uma determinada região do genoma; por exemplo, o polimorfismo rs16363 é identificado pela deleção da sequência "TGTTT" na localização cromossômica 22q13.1. O polimorfismo ao ser inserido no National Center for Biotechnology Information (NCBI) recebe um código único descrito com as letras rs (*reference sequence*), seguido da numeração específica para sua identificação correspondente à sua posição no genoma. Na localização cromossômica, o primeiro número refere-se ao cromossomo; no exemplo citado acima, o polimorfismo está localizado no cromossomo 22; a letra refere-se aos braços do cromossomo; a letra *p* indica o braço curto e *q* o braço longo; o número seguinte representa a posição no braço do cromossomo, no exemplo, banda 13, sub-banda 1.

A Figura 20.7 a seguir ilustra a identificação do INDEL "ATCG" no eletroferograma.

20.4 UTILIZAÇÃO DOS INDELS NA ANÁLISE DE DNA DEGRADADO

O DNA pode ser danificado ou destruído devido a condições ambientais adversas. A exposição ao ambiente degrada as moléculas de DNA, dividindo-o aleatoriamente em pequenos pedaços. Água, oxigênio, radiação

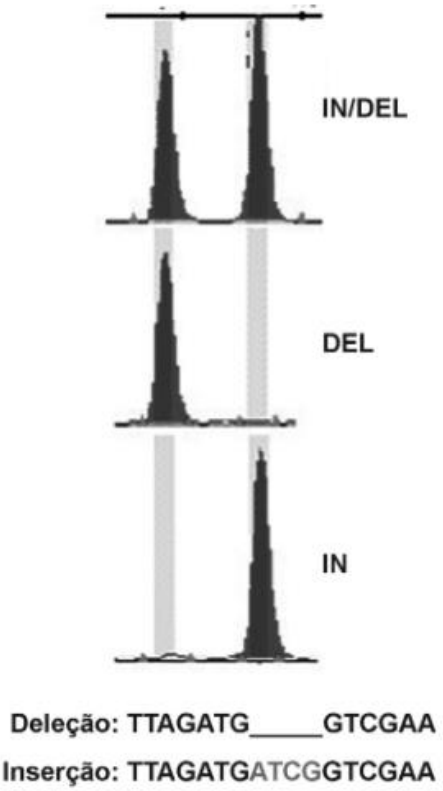


Figura 20.7 Eletroferograma identificando o INDEL (In/Del) com o alelo curto (deleção-Del) e o alelo longo (inserção-In-ATCG).

ultravioleta e enzimas nucleases são alguns dos agentes naturais que atuam nessa degradação[18].

A habilidade de recuperar e analisar sequências de DNA de amostras como ossos e dentes, expostos ao longo do tempo a uma variedade de condições ambientais, tornou-se uma ferramenta valiosa para a identificação de indivíduos ou amostras desconhecidas[19].

Pouco DNA endógeno, ação do meio ambiente, micro-organismos e presença de inibidores são os principais fatores que fazem com que a

recuperação e análise do DNA deste tipo de amostra sejam um desafio na área de identificação humana[20].

Ainda que a análise de STR (*short tandem repeats*) seja o principal polimorfismo utilizado como marcador genético em identificação humana[18], conforme já foi mencionado anteriormente, apresenta dificuldades quando as amostras possuem DNA degradado e/ou em pouca quantidade, devido ao tamanho relativamente grande dos fragmentos a serem amplificados (150 pb a 500 pb)[21].

Por essa razão, aumentou de modo considerável recentemente o interesse na utilização de INDELs como marcadores genéticos nas áreas de identificação humana e genética forense, devido à facilidade de análise, ampla distribuição no genoma, por apresentarem fragmento de pequeno tamanho, cerca de 100 pb e poderem ser detectados por diferentes técnicas. As frequências alélicas apresentam diferenças significativas entre populações geograficamente distintas, sendo potenciais marcadores para estudo de ancestralidade. Apresentam menor taxa de mutação em comparação com os STRs (Figura 20.8)[22].



Figura 20.8 Esquema comparando STRs com INDELs e mostrando os respectivos tamanhos em pares de bases (pb) obtidos na amplificação.

O fato de o fragmento de DNA amplificado ser pequeno é muito interessante, facilitando a análise em amostras biológicas não mantidas em boas condições de preservação, por exemplo, ossos e dentes coletados em casos de exumação de cadáveres ou de acidentes em massa. Nessas situações, a alternativa é a combinação de marcadores STRs e INDELs para a obtenção de uma análise mais conclusiva e robusta.

Em testes de paternidade ou outra relação biológica, a maioria dos laboratórios utiliza um conjunto de dezesseis marcadores STRs que oferecem um poder discriminatório alto e geram um resultado de exclusão de paternidade (entre 0,9999983 e 0,9999998) ou fortemente indicativo de probabilidade

de inclusão (índices de paternidade entre 522.000 e 4.110.000). No entanto, alguns casos apresentam resultados ambíguos ou inconclusivos, em que a relação reivindicada não pode ser confirmada por uma probabilidade suficientemente alta, ou uma exclusão é sugerida por incompatibilidade alélica em apenas um ou dois destes marcadores, sendo que para uma análise resultar em exclusão de paternidade há a necessidade de não coincidência entre o alelo paterno obrigatório (filho/a) e o alelo presente no suposto pai em, no mínimo, três marcadores autossômicos[23]. Na Figura 20.9, estão representados os eletroferogramas parciais de três regiões de STRs autossômicos, mostrando duas exclusões. É importante notar que incompatibilidades podem também ser identificadas por ação de mutações, uma vez que os STRs apresentam elevada taxa de mutação (cerca de 2×10^{-3}).

Resultados ambíguos surgem frequentemente, quando, por exemplo, o irmão biológico de primeiro grau do pai biológico é indicado como suposto pai de uma criança, de modo que a taxa de exclusão torna-se muito reduzida (poucas não coincidências nos alelos) e o índice de paternidade, usando uma razão de probabilidade de um homem aleatório na população, não se aplica; estes resultados são difíceis de serem interpretados. Um recurso encontrado é a análise de STRs adicionais para melhorar a probabilidade ou fornecer claras e inequívocas exclusões; no entanto, além daqueles que já são utilizados nos kits comerciais, poucos STRs são validados para que possam ser aplicáveis[24].

Estes resultados também são observados nos casos em que o suposto pai é falecido e há a necessidade de analisar amostras provenientes de exumação, como ossos e dentes, uma vez que esse tipo de amostra pode apresentar baixa quantidade de DNA e alto nível de degradação e contaminação, o que na grande maioria das vezes dificulta a análise de um número adequado de STRs. Nesses casos, analisando-se um número relativamente grande de INDELs (de 30 a 40), que apresentam menor taxa de mutação em comparação aos STRs, pode-se encontrar novas incompatibilidades, o que favorece a hipótese de exclusão, ou, ao contrário, caso não sejam encontradas incompatibilidades, aumentar o índice de paternidade a valores que suportem a confirmação de uma relação biológica.

20.5 TÉCNICAS PARA ANÁLISE DE INDELS AUTOSSÔMICOS

A técnica para análise de INDELs também é uma reação da PCR em multiplex, conforme mencionado anteriormente para amplificação de STRs.

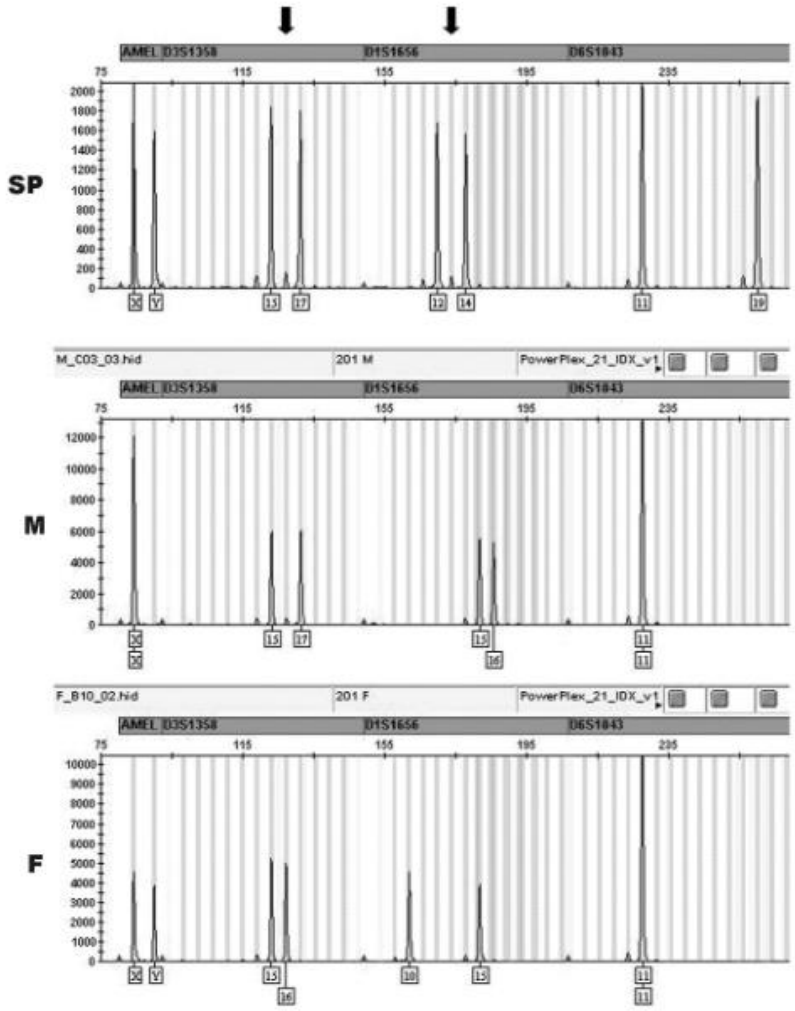


Figura 20.9 Comparação de três eletroferogramas parciais no *software* GenneMapper ID-X para verificar compatibilidade de alelos em um caso de investigação de paternidade. SP, suposto pai; M, mãe; F, filho. As setas indicam dois marcadores onde não houve compatibilidade, ou seja, o F não herdou o alelo obrigatório do SP.

Podem ser analisados trinta marcadores utilizando-se o kit *InvestigatorDIPplex* (QIAGEN), que também amplifica a amelogenina em uma única reação multiplex[25] (Figura 20.10).

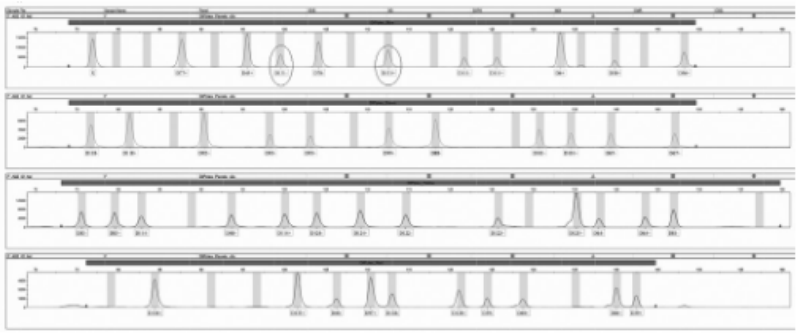


Figura 20.10 Eletroferograma da reação com o kit *InvestigatorDIPplex* (QIAGEN). (Trata-se de uma amostra feminina.) Os picos representam os alelos obtidos para cada marcador, sendo o alelo curto o que apresenta a deleção e o alelo longo o que apresenta a inserção, conforme exemplificado na Figura 20.3. Neste eletroferograma, os alelos são encontrados em regiões separadas, sendo divididos entre as quatro cores de fluoróforos. Estão identificados dois alelos de um mesmo marcador, separados no eletroferograma.

Devido ao fato de que os INDELs apresentam menor taxa de recombinação em comparação com os STRs, o número de INDELs analisados deverá ser maior para atingir um poder de discriminação aceitável; neste caso, os trinta polimorfismos são suficientes segundo o fabricante.

Outra reação foi padronizada por Pereira e colaboradores (2009)[15] e permite analisar 38 INDELs autossômicos em multiplex (Figura 20.11). Segundo os autores, esses 38 polimorfismos são suficientes para atingir um poder de descriminação aceitável.

Comparando-se as duas reações, verifica-se que a reação de Pereira e colaboradores (2009)[15] apresenta oito polimorfismos a mais do que aquela do kit *InvestigatorDIPplex* (QIAGEN), não havendo nenhum polimorfismo em comum. Em contrapartida, a reação de Pereira e colaboradores (2009)[15] não amplifica a amelogenina.

Outra avaliação interessante nos eletroferogramas demonstra que, nos 38-INDELs, os marcadores podem apresentar os alelos na mesma região, enquanto no *InvestigatorDIPplex* (QIAGEN), os alelos dos marcadores se

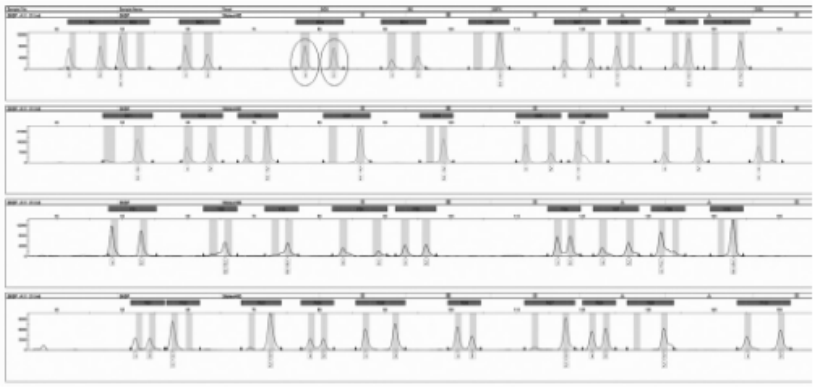


Figura 20.11 Eletroferograma da reação padronizada por Pereira e colaboradores (2009)[15]. (Amostra diferente da apresentada na Figura 20.10.) Os picos representam os alelos obtidos para cada marcador, sendo o alelo curto o que apresenta a deleção e o alelo longo o que apresenta a inserção, conforme exemplificado na Figura 20.3. Ao contrário da Figura 20.10, neste eletroferograma os alelos podem ser identificados na mesma região.

intercalam na mesma cor de fluoróforo. Isto dificulta a análise e a comparação entre os alelos de um mesmo marcador. A análise de amostras é realizada pelo *software* GeneMapper.

Como apresentado na Figura 20.12, os mesmos passos são realizados para análise de INDELs por ambas as reações multiplex, sendo que o passo de purificação pode ser opcional conforme a necessidade (quando o eletroferograma não discrimina adequadamente os alelos). De maneira geral, a metodologia para análise de INDELs é prática e rápida e muito similar à metodologia para análise de STRs, o que agrega praticidade ao trabalho em laboratórios de identificação humana.

20.6 APLICAÇÕES DO DNA MITOCONDRIAL

As mitocôndrias são organelas citoplasmáticas que apresentam em sua composição um genoma extracromossômico separado e distinto do genoma nuclear, o DNA mitocondrial (DNAmt), composto por 16.569 nucleotídeos e apresentando-se como uma dupla fita circular. Sua composição química



Figura 20.12 Fluxograma para análise de INDELs por ambas as reações. (O passo de purificação é opcional.)

em nada difere da composição do DNA nuclear; entretanto, ele possui um código genético próprio.

O DNAmt evoluiu pelo acúmulo de mutações nas linhagens maternas e acredita-se fixar novas mutações dez vezes mais rápido do que o DNA nuclear[26]. Este fato ocorre devido à baixa fidelidade da DNA polimerase mitocondrial, juntamente com a falta de mecanismos de reparo durante sua replicação, o que acaba sendo de grande interesse no campo forense.

Em alguns casos pode não ser possível a análise do DNA nuclear de algumas amostras (casos em que o DNA apresente degradações ou quando o material biológico não apresentar DNA nuclear), e então a melhor alternativa é a análise do DNA mitocondrial. Assim, o DNAmt é utilizado na rotina forense para analisar ossos, dentes, fios de cabelos e outros tipos de amostras biológicas, sempre que se necessite identificar o indivíduo.

A principal vantagem do DNAmt é que ele está presente em uma média de quinhentas a 2 mil cópias por célula. Esta abundância aumenta significantemente as chances de que algumas cópias de DNAmt permanecerão estáveis em ambientes que apresentam condições não favoráveis à sua conservação, além do que a natureza circular do DNAmt o torna menos suscetível à degradação por exonucleases[27].

O genoma mitocondrial possui herança estritamente materna, permitindo investigar relações de parentesco nesta linhagem, sendo, portanto, haploide e, por isso, não é submetido a processos de recombinação. Por essas duas razões, o DNAmt não pode ser utilizado na investigação de paternidade. Vale ressaltar que tanto os filhos como as filhas compartilham o DNAmt da mãe biológica.

Este genoma possui natureza monoclonal, ou seja, todo o DNAmt de um indivíduo apresenta a mesma sequência, o que simplifica a interpretação

dos resultados durante o sequenciamento do DNAmt. Porém, em alguns casos pode haver indivíduos que apresentam uma condição denominada de heteroplasmia, na qual uma mesma pessoa apresenta mais de um tipo de DNAmt, podendo ser heteroplasmia de sequência, quando há diferença em apenas um nucleotídeo, ou heteroplasmia de comprimento, quando há deleção ou inserção de uma citosina em regiões com repetições mononucleotídicas (conhecidas como poli-C). Nos dois casos, as análises devem ser muito cuidadosas para a interpretação da heteroplasmia. Portanto, há necessidade de treinamento do perito para análise do DNAmt em casos forenses.

20.7 UTILIZAÇÃO DO DNA MITOCONDRIAL EM ANÁLISES FORENSES

O genoma mitocondrial é composto de duas regiões (Figura 20.13): a região codificadora e a região controle, esta também conhecida como região não codificadora, região hipervariável ou *D-loop*, que controla a replicação e transcrição do DNAmt, a qual, devido ao acúmulo de mutações pontuais, apresenta uma maior taxa de mutação quando comparada à região codificadora.

A região controle é formada por aproximadamente 1.200 pares de base e, por ser altamente polimórfica, é muito utilizada para o propósito da genética forense. Esta região do DNAmt inclui três sub-regiões denominadas HV1 (formada por 342 pb), HV2 (268 pb) e HV3 (137 pb), sendo esta última região um segmento que pode apresentar um tamanho variável de pares de base, pelo fato de conter repetições dinucleotídicas "CA", cujo número pode variar de um indivíduo para outro[28]. Entre essas regiões existem outras duas denominadas inter-regiões, a primeira entre HV1 e HV2 e a segunda entre HV2 e HV3. A maior parte da variação de sequência entre indivíduos é encontrada nos segmentos HV1 e HV2. Por isso, estes são os mais utilizados na rotina forense para análise do DNAmt. Entretanto, a região HV3 também pode auxiliar as análises[29].

A identificação baseia-se na análise dos polimorfismos presentes nesses segmentos (SNPs, inserções e deleções/*indels*), que são comparados a uma sequência de referência (*revised Cambridge Reference Sequence – rCRS*), sequenciada em 1981 por Anderson e colaboradores[30], reanalisada e revisada por Andrews e colaboradores em 1999[31]. No entanto, tais regiões oferecem um limitado poder de discriminação em um contexto forense[32]. Assim, a região HV3, junto com as inter-regiões, vem sendo explorada nos últimos

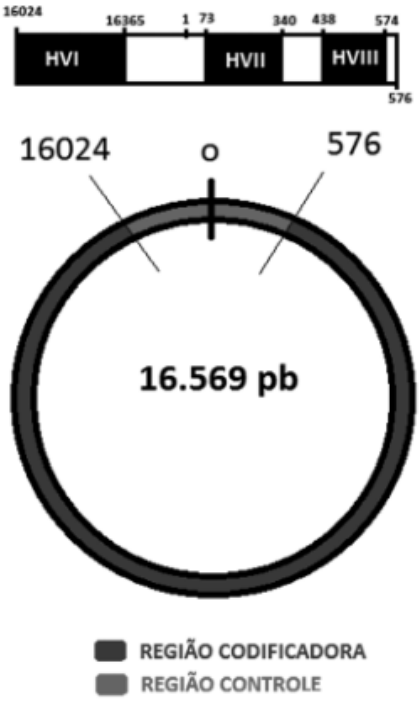


Figura 20.13 Esquema do DNA mitocondrial: região controle, com três sub-regiões (HV1, HV2 e HV3) e região codificadora.

anos, pois apesar de apresentarem baixos números de polimorfismos, informações complementares são obtidas quando adicionadas às regiões HV1 e HV2, aumentando o poder discriminativo, o que é importante na solução de casos forenses e, principalmente, em estudos de frequência populacional[33, 34, 35]. Essas características fazem com que o sequenciamento completo da região não codificadora do DNAmt seja mais adequado tanto para testes de identidade forense como para estudos evolucionários.

O sequenciamento da região controle do DNAmt pode ser feito a partir de DNA obtido de diversos materiais biológicos, como ossos, dentes, tecidos, sangue, sêmen e fios de cabelo. A partir do DNA extraído, realiza-se uma PCR utilizando os *primers* (*forward* e *reverse*) que flanqueiam as regiões de

interesse, podendo ser toda a região não codificadora ou cada uma das suas sub-regiões.

Os produtos da PCR são, então, purificados para retirada do excesso de *primers* e nucleotídeos que não foram incorporados durante a PCR, e submetidos à reação de sequenciamento. Os produtos obtidos a partir do sequenciamento também deverão ser purificados previamente à eletroforese capilar realizada em analisador genético (ABI3500 – Applied Biosystems). A análise dos resultados é realizada com a ajuda de *softwares* de bioinformática, e as sequências obtidas são comparadas com a rCRS.

Outra técnica de análise do DNAmt é a técnica de SNaPshot. Essa reação baseia-se na análise de polimorfimos de base única (SNPs), que representam a maior classe de polimorfismos humanos. Essa técnica está em uso crescente devido à detecção rápida de vários polimorfismos em um único ensaio. A capacidade multiplex é particularmente importante, especialmente no que diz respeito à análise de DNA forense, uma vez que pode reduzir o consumo de amostras, enquanto aumenta o rendimento do processamento das amostras e análise de dados. Esta técnica é realizada a partir da utilização de um kit comercial (ABI PRISM SNaPshot multiplex kit – Applied Biosystems) produzido pela empresa Life Technologies. A análise de SNPs do DNAmt, tanto da região não codificadora como da região codificadora, permite a classificação de amostras em haplogrupos.

20.8 APLICAÇÕES DO SEQUENCIAMENTO DO DNAMT

Alguns fatos importantes comprovam como a análise do DNAmt é de grande importância no contexto forense: foi a primeira ferramenta utilizada na identificação de ossadas de soldados americanos que lutaram na Guerra do Vietnã[36] e na identificação das vítimas da tragédia de 11 de setembro de 2001 no World Trade Center (EUA).

Outro exemplo interessante da utilização do DNA mitocondrial foi realizado na década de 1990 em um estudo de parentesco, no qual cientistas americanos foram convidados a examinar ossos e dentes de oito esqueletos de uma vala na Sibéria, onde se acreditava estar enterrada a família real russa Romanov, que fora executada pelos bolcheviques. Neste caso, os resultados do sequenciamento do DNAmt elucidou que havia vínculo genético entre as ossadas[37].

A utilização do DNAmt estendeu-se para estudos da evolução humana, migrações e formação da história de diversas populações, não permanecendo

apenas na rotina forense. O emprego do DNAmt como ferramenta para investigações desta natureza surgiu com os estudos pioneiros do grupo liderado por Allan Wilson nos anos 1980[38].

20.9 BANCO DE DADOS

A análise do DNAmt é rotineiramente utilizada em diversos laboratórios como ferramenta em análises forenses e estudos genéticos evolutivos. No entanto, é preciso um conhecimento aprofundado das várias características do DNAmt a fim de analisar e interpretar corretamente os resultados. A forma correta de interpretar um resultado de DNAmt ainda não foi padronizada, existindo apenas recomendações, como, por exemplo, aquelas adotadas pelo Federal Bureau of Investigation (FBI)[39] ou publicadas pelo Grupo de Habla Española y Portuguesa – International Society for Forensic Genetics (GHEP-ISFG)*.

A maioria dos laboratórios que utilizam tipagem do DNAmt baseiam-se nos polimorfismos (SNPs, deleções e inserções) presentes na sequência de nucleotídeos nas regiões HV1 e HV2, comparando a amostra questionada com a sequência referência (rCRS) para a anotação das diferenças (polimorfismos)[29].

De acordo com o FBI, se essas sequências diferirem em dois ou mais polimorfismos, uma exclusão inequívoca pode ser feita. Se as sequências corresponderem uma à outra, existe a possibilidade da amostra questionada pertencer ao indivíduo analisado ou demonstrar vínculo materno, podendo, no entanto, pertencer a outro indivíduo não aparentado; por isso, torna-se necessário verificar a frequência com que um conjunto de polimorfismo ocorre em uma população (frequência do haplótipo), gerando a necessidade da criação de um banco de dados populacional de DNAmt.

O termo "haplogrupo" refere-se ao conjunto de haplótipos (conjunto de alelos), que são designados por letras do alfabeto e derivados por descendência de uma mesma molécula ancestral, as quais apresentam um padrão basal de mutações. Os haplogrupos são definidos por um conjunto de haplótipos que são constituídos de nucleotídeos particulares presentes nas regiões controle e codificadora do DNAmt. A relação entre haplótipos e haplogrupos do DNAmt humano nos fornece informações mais detalhadas sobre a estrutura da variação genética. No site PhyloTree** encontra-se a árvore filogené-

* Ver http://www.gep-isfg.org.
** Ver www.phylotree.org.

tica mitocondrial humana que está representada de modo simplificado na Figura 20.14.

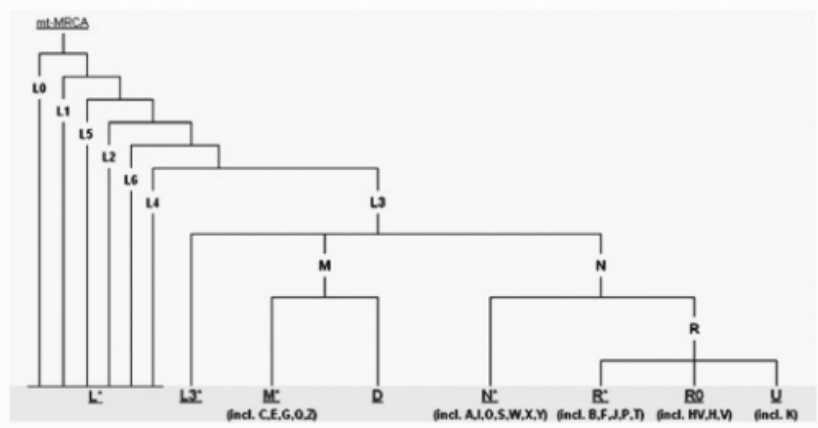


Figura 20.14 Árvore filogenética humana simplificada do DNA mitocondrial, mostrando os principais haplogrupos.

A criação de banco de dados de DNAmt para estimar a frequência dos haplótipos em determinada população é indispensável para permitir sua utilização. Entretanto, pode estar sujeito a vários tipos de erros, principalmente humanos (na leitura e digitação de dados, por exemplo), havendo a necessidade de se estabelecerem regras de controle de qualidade tanto na geração quanto na compilação de resultados[29]. Atualmente, existem ferramentas que podem ser utilizadas para que esses erros humanos sejam evitados como, por exemplo, MITOMAP (*mtDNA control region sequence polimorphisms*)* e IAN LOGAN**, as quais permitem comparar se todos os polimorfismos encontrados já foram descritos na literatura.

Diversos pesquisadores geram milhares de dados de genomas mitocondriais completos de seres humanos em todos os continentes, bem como uma nomenclatura detalhada dos tipos de DNAmt ou haplogrupos em todo o mundo e uma enorme árvore filogenética dos haplogrupos observados, sub-haplogrupos e macro-haplogrupos[40]. No campo da identificação humana,

* Ver http://www.ianlogan.co.uk/mtDNA.htm.
** Ver http://www.ianlogan.co.uk/mtDNA.htm.

as revistas científicas de maior impacto já exigem que os dados de trabalhos envolvendo análises de DNAmt ou outros tipos de análises gerado por pesquisadores sejam adicionadas aos bancos de dados para serem publicados.

Um exemplo de banco de dados que foi desenvolvido com rigorosos parâmetros de qualidade é o banco de dados mundial de DNAmt denominado EMPOP. Para manter um padrão de qualidade, é necessário ter uma padronização da geração, análise e transferência de dados, além do uso de *softwares* para a visualização de dados, interpretação e detecção de erros.

20.10 TÉCNICA PASSO A PASSO

Protocolo para a análise de 38 INDELs autossômicos

Extração de DNA

A extração do DNA do interior das células é o primeiro passo para sua análise. Descreveremos a seguir um protocolo[41] para a extração de DNA de sangue armazenado em papel de filtro FTA (Whatman), utilizando a resina Chelex 100 (Biorad) a 5%:

µ

1) Colocar em um tubo de 1,5 mL 2 discos (3 mm cada) de papel de filtro com sangue.
2) Adicionar 50 µL de H_2O, vortexar rapidamente e descartar a H_2O.
3) Adicionar 50 µL de H_2O e incubar a temperatura ambiente por 30 minutos.
4) Descartar a H_2O e adicionar 100 µL da solução homogeneizada da resina chelex 100 (Biorad) a 5%.
5) Incubar por 1 hora a 56 °C.
6) Vortexar rapidamente e incubar por 8 miunutos a 100 °C.
7) Centrifugar a 14.000 RPM por 3 minutos.
8) Armazenar o sobrenadante.

Reação em cadeia da polimerase (PCR)

Segue a descrição do preparo da PCR para amplificação dos 38 INDELs padronizado por Pereira et al. (2009)[15], utilizando o *Multiplex PCR kit* (QIAGEN):

H_2O	3 µL
Qiagen *Multiplex PCR master mix* (2×)	5 µL
Primer mix (10×)	1 µL
DNA 0,3 a 5 ng/µ L	1 µL
Volume total	10 µL

Em seguida, é utilizada a ciclagem em termociclador:

95 °C – 15 min.	
94 °C – 30 seg.	
60 °C – 90 seg.	10 ciclos
72 °C – 60 seg.	
94 °C – 30 seg.	
58 °C – 90 seg.	20 ciclos
72 °C – 60 seg.	
72 °C – 60 min.	
4 °C – até a retirada do termociclador.	

Observação: um passo adicional de purificação do produto da reação de PCR pode ser utilizado se necessário.

Análise do produto da reação de PCR

O produto da PCR deve ser preparado para eletroforese capilar pela adição de 1 µL do produto amplificado em 9,7 µL de formamida HI-DI (Applied Biosystems) e 0,3 µL de GeneScan 600 LIZ (Applied Biosystems). A separação e detectção são realizadas por eletroforese capilar.

20.11 CONCLUSÕES

As ferramentas para identificação humana utilizando o DNA são, atualmente, bastante eficazes na discriminação entre indivíduos e/ou vestígios e na investigação de paternidade, gerando índices de verossimilhança bastante elevados. Entretanto, cabe ressaltar que o exame de DNA na genética

forense constitui-se em uma evidência a mais na elucidação dos casos do Poder Judiciário, favorecendo ou não a hipótese de acusação ou de defesa. O perito deve analisar os resultados obtidos de maneira cuidadosa, isenta e ética, relatando as evidências científicas encontradas através das análises dos perfis genéticos de maneira clara e objetiva, valorando os dados com os cálculos estatísticos adequados, mas deixando ao juiz a tarefa de avaliação final da importância de tais evidências no julgamento do referido caso.

REFERÊNCIAS

1. França GV. Medicina Legal. 6. ed. Rio de Janeiro: Guanabara-Koogan; 2001. 579 p.
2. Jeffreys AJ, Wilson V, Thein SL. Hypervariable minisatellite regions in human DNA. Nature. 1985;314:67-72.
3. Jeffreys AJ, Wilson V, Thein SL. Individual specific "fingerprints" of human DNA. Nature. 1985;316:75-79.
4. Wyman AR, White R. A highly polymorphic locus in human DNA. Proc Natl Acad. Sci USA. 1980;77:6754-8.
5. Jeffreys AJ. Genetic fingerprinting. Nature Medicine. 2005;11(10):XIV-XVIII.
6. Saiki RK, Scharf S, Faloona F, Mullis K, Horn G, Erlich HA, et al. Enzymatic amplification of ß-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science. 1985;230:1350-4.
7. Saiki RK, Gelfand DH, Stoffel S, Scharf S, Higuchi RH, Horn GT, et al. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science. 1988;239:487-91.
8. Mullis K, Faloona F. Specific synthesis of DNA in vitro via a polymerase catalyzed chain reaction. Method Enzymol. 1987;155: 335-50.
9. Watson JD. DNA: o segredo da vida. São Paulo: Companhia das Letras; 2005. 470 p.
10. The FBI Federal Bureau of Investigation. Combined DNA Index System (Codis) [Internet]. Available from: http://www.fbi.gov/about-us/lab/biometric-analysis/codis.
11. Consultor Jurídico. Técnica usada na tipagem genética não está livre de erros [Internet]. 2006 Jun 13. Available from: http://www.conjur.com.br/2006-jun-13/tecnica_usada_tipagem_genetica_nao_livre_erros.
12. Paradela ER, Figueiredo AL dos S; Smarra ALS. A identificação humana por DNA: aplicações e limites [Internet] [Cited Nov 2013]. In: Âmbito Jurídico, Rio Grande, IX, n. 30, jun 2006. Available from: http://www.ambito-juridico.com.br/site/index.php?artigo_id=1175&n_link=revista_artigos_leitura.
13. Houck MM. A realidade do CSI. Scientific American Brasil. Available from: http://www.uol.com.br/sciam/reportagens/a_realidade_do_csi.html.
14. Bassette F. País padroniza técnicas para perícia. Folha de S. Paulo. 27 set 2010:Cotidiano.
15. Pereira R, Phillips C, Alves C, Amorim A, Carracedo A, Gusmao L. A new multiplex for human identification using insertion/deletion polymorphisms. Electrophoresis. 2009;30:3682-90.
16. Weber JL, David D, Heil J, Fan Y, Zhao C, Marth G. Human diallelic insertion/deletion polymorphisms. Am J Hum Genet. 2002;71(4):854-862.

17. Mills RE, Luttig CT, Larkins CE, Beauchamp A, Tsui C, Pittard WS, Devine SE. An initial map of insertion and deletion (INDEL) variation in the human genome. Genome Res. 2006;16(9):1182-90.
18. Butler J. Advanced Topics in Forensic DNA Typing: Methodology. 1st ed. Waltham/San Diego/London:Academic Press; 2011. 704 p.
19. Hochmeister MN, Budowle B, Borer UV, Eggmann U, Comey CT, Dirnhofer R. Typing of deoxyribonucleic acid (DNA) extracted from compact bone from human remains. J Forensic Sci. 1991;36:1649-61.
20. Loreille OM, Diegoli TM, Irwin JA, Coble MD, Parsons TJ. High efficiency DNA extraction from bone by total demineralization. Forensic Sc International. 2007;1:191-5.
21. Zidkova A, Horinek A, Kebrdlova V, Korabecna M. Application of the new insertion–deletion polymorphism kit for forensic identification and parentage testing on the Czech population. Int J Legal Méd. 2011;127(1):7-10.
22. Phillips C, Fondevila M, Garcia-Margarinos M, Rodriguez A, Salas A, Carracedo A, Lareu MV. Resolving relationship tests that show ambiguous STR results using autosomal SNPs as supplementary markers. Forensic Science International Genetics. 2008;2:198-204.
23. Mullaney JM, Mills RE, Pittard WS, Devine SE. Small insertions and deletions (INDELs) in human genomes. Hum Mol Genet. 2010;19(2):131-6.
24. Yang N, Li H, Criswell LA, Gregersen PK, Alarcon-Riquelme ME, Kittles R, et al. Examination of ancestry and ethnic affiliation using highly informative diallelic DNA markers: application to diverse and admixed populations and implications for clinical epidemiology and forensic medicine. Hum Genet. 2005;118(3-4):382-92.
25. Larue BL, Ge J, King JL, Budowle B. A validation study of the Qiagen Investigator DIPplex® kit; an INDEL-based assay for human identification. Int J Legal Med. 2012;126(4):533-40.
26. Wallace DC, Ye J, Neckelmann SN, et al. Sequence Analysis of cDNAs for the Human and Bovine ATP Synthase Beta Subunit: Mitochondrial Genes Sustain Seventeen Times more Mutations. Current Genetics. 1987;12:81-90.
27. Kashyap VK, Sitalaximil T, Chattopadhyay P, Trivedi R. DNA Profiling Technologies in Forensic Analysis. International Journal of Human Genetics. 2004;3: 11-30.
28. Lutz S, Wittig H, Weisser HJ, Heizmann J, Jungle A, Simonin ND, et al. Is it possible to differentiate mtDNA by means of HVIII in samples that cannot be distinguished by sequencing the HVI and HVII regions? Forensic Science International. 2000;113:97-101.
29. Paneto GG, et al. Heteroplasmy in Hair: Study of Mitochondrial DNA Third Hypervariable Region in Hair and Blood Samples. J Forensic Sci. 2010;55:715-7.
30. Anderson S, Bankier AT, Barrel BG, et al. Sequence and organization of the human mitochondrial genome. Nature. 1981;290:457-65.

31. Andrews R, et al. Reanalysis and revision of Cambridge reference sequence for human mitochondrial DNA. Nat Genet. 1999;23:147.
32. Huang D, Gi C, Yi S, et al. Typing of 24 mtDNA SNPs in Chinese Population Using SNaPshot Minisequencing. Med Sci. 2010;30:291-98.
33. Bini C, et al. Different informativeness of the three hypervariable mitochondrial DNA regions in the population of Bologna (Italy). Forensic Sci Int. 2003;135:48-52.
34. Vanecek T, Vorel F, Sip M. Mitochondrial DNA D-loop hypervariable regions: Czech population data. Int J Legal Med. 2004;118:14-8.
35. Zhang YJ, et al. Haplotype diversity in mitochondrial DNA hypervariable region I, II and III in northeast China Han. Forensic Sci Int. 2005;149:267-9.
36. Holland MM, Fisher DL, Mitchell LG. Mitochondrial DNA sequence analysis of human skeletal remains: identification of remains from the Vietnam war. J Forensic Sci. 1993;38:542-53.
37. Farah SB. DNA segredos e mistérios. 2 ed. São Paulo: Sarvier; 2007.
38. Wilson AC, et al. Mitochondrial DNA and two perspectives on evolutionary genetics. Biol J Linn. 1985;6:375-400.
39. Scientific Working Group on DNA Analysis Methods (Swgdam). Guidelines for Mitochondrial DNA (mtDNA) Nucleotide Sequence Interpretation. Forensic Science Communications. 2003. 5 p.
40. Bandelt HJ, Macaulay V, Richards M. Human mitochondrial DNA and the evolution of Homo sapiens. 1st ed. Heidelberg: Springer; 2006. p.153.
41. Singer-Sam J, Tanguay R, Riggs AD. Use of chelex to improve the PCR signal from a small number of cells. Amplifications: A Forum for PCR Users. 1989;3:11.