UNIVERSIDADE ESTADUAL PAULISTA
“JULIO DE MESQUITA FILHO”
Campus de Araraquara - Instituto de Química
Programa de Biotecnologia

FLÁVIA ALVES DE JESUS SILVA

POLIMORFISMOS DE INSERÇÃO/DELEÇÃO NO CROMOSSOMO X: ANÁLISE DE 32 MARCADORES NA POPULAÇÃO DA REGIÃO SUDESTE DO BRASIL

ARARAQUARA - SP
2020

FLÁVIA ALVES DE JESUS SILVA

POLIMORFISMOS DE INSERÇÃO/DELEÇÃO NO CROMOSSOMO X: ANÁLISE DE 32 MARCADORES NA POPULAÇÃO DA REGIÃO SUDESTE DO BRASIL

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

2020

FICHA CATALOGRÁFICA

Silva, Flávia Alves de Jesus
S586p Polimorfismos de inserção/deleção no cromossomo X : análise de 32 marcadores na população da região sudeste do Brasil / Flávia Alves de Jesus Silva. – Araraquara : [s.n.], 2020
87 f. : il.

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

1. Análise de DNA. 2. Genética da população humana. 3. Polimorfismo (Genética). 4. Cromossomo X. 5. Marcadores genéticos. I. Título.

Bibliotecária Responsável: Ana Carolina Gonçalves Bet - CRB8/8315

unesp

UNIVERSIDADE ESTADUAL PAULISTA

Câmpus de Araraquara

Instituto de Química

CERTIFICADO DE APROVAÇÃO

TÍTULO DA TESE: "Polimorfismos de inserção/deleção no cromossomo X: análise de 32 marcadores na população da região sudeste do Brasil"

AUTORA: FLAVIA ALVES DE JESUS SILVA
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

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

Profª. Drª. RAQUEL MANTUANELI SCAREL CAMINAGA
Departamento de Morfologia / Faculdade de Odontologia - UNESP - Araraquara

Prof.ª Dr.ª LEONOR GUSMÃO
Departamento de Biologia / Universidade do Estado do Rio de Janeiro - UERJ - Rio de Janeiro

Prof.ª Dr.ª GREICIANE GABURRO PANETO
Departamento de Zootecnia / Centro de Ciências Agrárias - UFES - Alegre

Araraquara, 26 de maio de 2020

DEDICATÓRIA

As minhas filhas, Camila e Helena, por suportar e compreender a minha ausência em diversos momentos, para que este trabalho pudesse ser realizado. Isto é por vocês meus amores!

Ao meu amado esposo e companheiro de todas as horas, Elias, que sempre me apoiou e incentivou a prosseguir, por maiores que fossem as dificuldades, fortalecendo-me a cada dia.

A todos vocês, que amo muito, a minha gratidão e o meu reconhecimento. Nos méritos desta conquista há muito de vocês!

AGRADECIMENTOS

*Agradeço primeiramente a **Deus** pelo dom da vida, por iluminar sempre o meu caminho, me proporcionando saúde e sabedoria para buscar sempre mais conhecimento.*

*Agradeço a toda minha família pelo tempo doado para a realização das pesquisas, pela imensa compreensão, em especial as minhas filhas, **Camila e Helena**, ao meu esposo **Elias**, sem o apoio de vocês nada disso seria possível.*

*Aos meus pais, **Conceição e Geraldo**, meu profundo apreço por todo o apoio durante toda a minha vida e, sobretudo, nos momentos mais difíceis dos últimos anos, suprindo minhas ausências junto as minhas filhas e permitindo que eu me dedicasse à realização do meu trabalho.*

*Aos meus irmãos, **Fabrício e Flaviane**, que sempre estiveram ao meu lado, me dando as maiores forças e os melhores conselhos, os quais sempre me mantiveram no caminho certo.*

A vocês a minha eterna gratidão.

*A minha Orientadora, **Profa. Dra Regina Cicarelli**, por dedicar seu tempo a me ensinar um pouco do muito que ela sabe. Serei sempre grata por sua confiança, compreensão e paciência. Tenho um imenso orgulho em ter você como minha orientadora. A você minha professora, meu muito obrigado.*

*A **Profa. Dra Leonor Gusmão**, por estar sempre disponível a nos atender, mesmo tendo inúmeros compromissos. Agradeço a enorme contribuição que fez a este trabalho, por disponibilizar as amostras do Rio de Janeiro e por todo ensinamento e atenção que dedicou a mim neste período, serei sempre eternamente grata.*

*A **Juliana Martinez**, amiga que me incentivou, aconselhou e ensinou desde o desenvolvimento teórico do projeto até a sua conclusão final, agradeço pelas numerosas sugestões e, antes de tudo, pela amizade, carinho e disponibilidade em ajudar.*

*A **Fernanda Polverari** e **Bianca Belon**, pelas conversas serias quando necessárias, pelos momentos de descontração, pelo apoio, ensinamentos, por tirar minhas dúvidas e me apontar soluções quando tudo parecia estar perdido. Agradeço também pela amizade sincera que dedicaram a mim.*

*A **Isabela Brunelli** e **Danilo Braganholi**, companheiros na viagem mais fantástica que já realizei, agradeço pelas conversas e por sempre se disponibilizarem a me ajudarem, mesmo com a correria do dia-a-dia.*

*A **Silvia Loiola**, por auxiliar na seleção das amostras da população do Rio de Janeiro. Muito obrigada por ceder este tempo a este trabalho.*

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior ***(CAPES)*** *pela concessão da bolsa de estudo.*

A todos os colegas da Faculdade de Ciências Farmacêuticas da UNESP de Araraquara, pela gentileza, pela atenção, sempre disposta a ajudar.

A todos as funcionárias da Seção Técnica de Pós-Graduação do Instituto de Química da UNESP de Araraquara, em especial ***Wennia****,* ***Robson, Cíntia*** *e* ***Ana Paula****, pelos serviços prestados e pela atenção.*

A todas as pessoas que fizeram parte na construção deste trabalho sempre me apoiando.

A estes dedico este trabalho, sem a ajuda e compreensão de todos a realização deste sonho não seria possível.

Muito obrigada por tudo!

RESUMO

Na rotina forense, há casos em que a amostra biológica se encontra degradada ou com baixa concentração de DNA, o uso exclusivo dos marcadores STRs pode gerar um resultado estatístico inconclusivo, tornando fundamental a análise de marcadores adicionais para a resolução do caso. Utilizado como método complementar, os polimorfismos de inserção e deleção (InDels) têm mostrado grande potencial para superar as limitações dos marcadores tradicionais. Adicionalmente, estudos de genética forense tem sido cada vez mais explorados no âmbito do cromossomo X, especialmente nos casos em que a análise dos autossômicos não é suficiente. Nessa perspectiva, este trabalho avaliou a utilidade de um conjunto de 32 polimorfismos de inserção/deleção do cromossomo X em amostras da população da região sudeste do Brasil. Afim de avaliar a diversidade genética destas populações, foram analisados os perfis genotípicos de 627 indivíduos sem relação genética pelo cromossomo X, residentes nos estados de Minas Gerais (90 mulheres e 116 homens), Espírito Santo (201 homens) e Rio de Janeiro (220 homens). Os resultados indicam que o painel dos 32 X-InDels é bastante eficiente para a sua finalidade, sendo que praticamente todos os marcadores se mostraram altamente informativos para as populações estudadas. O teste de equilíbrio de Hardy-Weinberg foi realizado nas amostras femininas de Minas Gerais e não foram verificados desvios significativos. Em todas populações analisadas o painel demonstrou alta eficiência forense, confirmado pelo alto poder de discriminação para Minas Gerais (PD_F=0,999999999994 e PD_M=0,99999994), Espírito Santo (PD_F=0,999999999994 e PD_M=0,99999995) e Rio de Janeiro (PD_F=0,999999999997 e PD_M=0,99999997), e pelo elevado poder de exclusão estimado em trios: pai/mãe/filha (Minas Gerais: 0,999996; Espírito Santo: 0,999996; Rio de Janeiro: 0,999997) e duos: pai/filha (Minas Gerais: 0,9994; Espírito Santo: 0,9995; Rio de Janeiro: 0,9996). No estudo comparativo com outras populações, a maior similaridade observada foi das 3 populações estudadas neste trabalho com as outras populações brasileiras, seguido pelas populações dos departamentos colombianos e Europa. Durante a análise foi observada, nas três populações estudadas, a presença de uma variante no MID1361. Na análise de ancestralidade para estas populações, foi identificada proporções ancestrais de europeus (MG= 39%, RJ= 35% e ES= 38%), africanos (MG= 36%, RJ= 38% e ES= 37%) e nativo-americanos (MG= 25%, RJ= 27% e ES= 25%). Os dados para os marcadores de inserção/deleção no cromossomo X disponíveis na literatura vêm sendo cada vez

mais explorados, mas o conhecimento sobre este assunto deve ser ampliado. Os valores de poder de discriminação e exclusão obtidos neste trabalho já demonstram seu potencial como método complementar para a análise de amostras forenses.

Palavras-chave: Identificação humana. Genética populacional. Polimorfismos de inserção/deleção. InDels. Cromossomo X. X-InDels. Minas Gerais. Rio de Janeiro. Espírito Santo. Brasil.

Abstract

In the forensic routine, there are cases in which the biological sample is degraded or with a low concentration of DNA, the exclusive use of STR markers can generate an inconclusive statistical result, making the analysis of additional markers essential for the resolution of the case. Used as a complementary method, the insertion and deletion polymorphisms (InDels) have shown great potential to overcome the limitations of traditional markers. Additionally, the X chromosome has been increasing significant importance in forensic genetics studies, especially in cases where the analysis of autosomal is not enough. In this perspective, this work evaluated the usefulness of a set of 32 polymorphisms of insertion/deletion of the X chromosome in samples from the population of the southeastern region of Brazil. In order to evaluate the genetic diversity of these populations, the genotypic profiles of 627 individuals without genetic relationship were analyzed by the X chromosome, living in the states of Minas Gerais (90 females and 116 males), Espírito Santo (201 males) and Rio de Janeiro (220 males). The results indicate that the 32 X-InDels panel is enough efficient for its purpose, with practically all the markers proving to be highly informative for the studied populations. The Hardy-Weinberg equilibrium test was performed on female samples from Minas Gerais and no significant deviations were found. In all analyzed populations, the panel demonstrated high forensic efficiency, confirmed by the high power of discrimination for Minas Gerais (PDF= 0.9999999999994 and PDM= 0.999999994), Espírito Santo (PDF=0.999999999994 and PDM= 0.99999995) and Rio de Janeiro (PDF= 0.999999999997 and PDM= 0.99999997), and the high exclusion power estimated in trios: father/mother/daughter (Minas Gerais: 0.999996; Espírito Santo: 0.999996; Rio de Janeiro: 0.999997) and duo: father/daughter (Minas Gerais: 0.9994; Espírito Santo: 0.9995; Rio de Janeiro: 0.9996). In the comparative study with other populations, the greatest similarity observed was between the 3 populations studied in this work, followed by the other Brazilian populations, from the Colombian departments and Europe. During the analysis, the presence of a variation in MID1361 was observed in the three populations studied. In the analysis of ancestry for these populations, ancestral proportions of Europeans were identified (MG = 39%, RJ = 35% and ES = 38%), Africans (MG = 36%, RJ = 38% and ES = 37%) and native-americans (MG = 25%, RJ = 27% and ES = 25%). The data for insertion / exclusion markers on the X chromosome are available in the has been increasingly explored literature, but the knowledge on this subject must

be expanded. The values of discriminating power and excluding them in this work already demonstrate its potential as a complementary method for analyzing forensics.

Keywords: Human identification. Population genetics. Insertion/deletion polymorphisms. InDels. X chromosome. X-InDels. Minas Gerais. Rio de Janeiro. Espírito Santo. Brazil

LISTA DE FIGURAS

Figura 1:Esquema representando a simulação do resultado de um sistema de STRs em eletroforese para investigação de paternidade comparando o número de repetições de um determinado loci pesquisado 24

Figura 2:O polimorfismo de nucleotídeo único 25

Figura 3: Eletroferograma de detecção de Inserção e Deleção 27

Figura 4: Ideograma do cromossomo X indicando a localização dos 32 marcadores X-InDel selecionados. 43

Figura 5: Eletroferograma de uma amostra masculina amplificada com o multiplex 32 X-InDels. 49

Figura 6: Eletroferograma de uma amostra feminina amplificada com o multiplex 32 X-InDels. 50

Figura 7: Eletroferograma parcial de três amostras de Minas Gerais obtidas com o multiplex 32 X-InDels. Em destaque a variante alélica do alelo curto do marcador MID1361 51

Figura 8: Eletroferograma parcial de uma amostra do Espírito Santo obtida com o multiplex 32 X-InDels. Em destaque a variante alélica do alelo curto do marcador MID1361. 51

Figura 9: Eletroferograma parcial de uma amostra do Rio de Janeiro obtida com o multiplex 32 X-InDels. Em destaque a variante alélica do alelo curto do marcador MID1361. 52

Figura 10: Parâmetros estatísticos de interesse forense para os marcadores 32 X-InDels, para a população da Somália, Angola e Moçambique, Portugal e Macau. 66

Figura 11: Gráfico MDS bidimensional elaborado a partir da distância genética (F_{ST}) entre África, Europa, Leste Asiático, Nativo-americano (Pasto - Colômbia), diversas populações da Colômbia e Brasil (São Paulo, Mato Grosso, Brasília, Minas Gerais, Rio de Janeiro e Espírito Santo). 69

Figura 12: Representação esquemática para a população de Minas Gerais (MG), Rio de Janeiro (RJ) e Espírito Santo (ES) utilizando os marcadores 32 X-InDels obtidos atraves da anlise do STRUCTURE baseado na contibuição da população africana (AFR), europeira (EUR) e nativo-americano (NAM) (K=3).....70

Figura 13: Proporção de ancestralidade identificada em marcadores X-InDels, mtDNA, Y-SNP e AIM-InDels nas populações de Minas Gerais, Rio de Janeiro e Espírito Santo. .72

LISTA DE TABELAS

Tabela 1: InDels do cromossomo X e sequências iniciadoras de PCR usadas no multiplex 44

Tabela 2: Frequência alélica dos 32 marcadores X-InDels nas populações de Minas Gerais, Espírito Santo e Rio de Janeiro. Número de cromossomos analisados: 296 (Minas Gerais) 201 (Espírito Santo) e 220 (Rio de Janeiro). 53

Tabela 3: Parâmetros estatísticos da diversidade gênica para os 32 marcadores X-InDels na população feminina de Minas Gerais. Tamanho da amostra: 90. 55

Tabela 4: Desequilíbrio de ligação dos pares de marcadores analisados na população masculina de Minas Gerais. Tamanho da amostra: 116 57

Tabela 5: Desequilíbrio de ligação dos pares de marcadores analisados na população do Rio de Janeiro. Tamanho da amostra: 220 58

Tabela 6: Desequilíbrio de ligação dos pares de marcadores analisados na população do Espírito Santo. Tamanho da amostra: 201 59

Tabela 7: Frequência haplotípica observada na população masculina de Minas Gerais para os marcadores em desequilíbrio de ligação. Tamanho da amostra: 116. 60

Tabela 8: Frequência haplotípica observada na população masculina de Espírito Santo para os marcadores em desequilíbrio de ligação. Tamanho da amostra: 201. 60

Tabela 9: Frequência haplotípica observada na população masculina de Rio de Janeiro para os marcadores em desequilíbrio de ligação. Tamanho da amostra: 220. 61

Tabela 10: Parâmetros estatísticos de interesse forense para os marcadores 32 X-InDels na população de Minas Gerais. Número de cromossomos analisados: 296. 62

Tabela 11: Parâmetros estatísticos de interesse forense para os marcadores 32 X-InDels na população do Espírito Santo. Número de cromossomos analisados: 201 64

Tabela 12: Parâmetros estatísticos de interesse forense para os marcadores 32 X-InDels na população do Rio de Janeiro. Número de cromossomos analisados: 220 65

Tabela 13: Amostras usadas na comparação entre as populações, número de cromossomos analisados (N) e respectivas referências. 67

Tabela 14: Distância genética (F_{ST}) entre as populações (abaixo da diagonal) e o valor de p correspondente (acima da diagonal) 68

LISTA DE ABREVIATURAS E SIGLAS

DNA	*Deoxyribonucleic acid*
dbSNP	*Single Nucleotide Polymorphism database*
DG	Diversidade gênica
EUA	Estados Unidos da América
GHEP-ISFG	Grupo de Línguas Espanhola e Portuguesa da Sociedade Internacional de Genética Forense
HID	*Human Identification*
InDel	Inserção/deleção
LD	*Linkage Disequilibrium*
MID	*Marshfield identification*
MCMC	*Markov chain Monte Carlo*
MDS	*Multidimensional scaling*
MEC	Chance média de exclusão
MEC_T	Chance média de exclusão em trios envolvendo filha
MEC_D	Chance média de exclusão em duo pai/filha
MEC_{PGM}	Chance média de exclusão em casos deficientes envolvendo mãe, filha e avó paterna
PCR	*Polymerase Chain Reaction*
PD	Poder de discriminação
PD_F	Poder de discriminação em mulheres
PD_M	Poder de discriminação em homens
RFU	*Relative fluorescence unit*
SNP	*Single Nucleotide Polymorphism*
STR	*Short Tandem Repeat*
UNESP	Universidade Estadual Paulista Júlio Mesquita Filho
X-InDel	Inserção/deleção no cromossomo X
X-STR	*Short Tandem Repeat* no cromossomo X

LISTA DE SÍMBOLOS

~	Aproximadamente
%	Porcento
ng	Nanogramas
nº	Número
ºC	Graus Célsius
®	Marca registrada
ml	Mililitros
µL	Microlitros
cm	Centímetros
mg	Miligramas
D’	Valor normalizado do desequilíbrio de ligação
p-LD	Valor de *p* do desequilíbrio de ligação
p-HWE	Valor de *p* do equilíbrio de Hardy-Weinberg
pb	Par de bases
kb	Kilobase
Mb	Megabase
q.s.p.	Quantidade suficiente para
rpm	Rotações por minuto
min	Minutos
seg	Segundos

Sumário

1. INTRODUÇÃO 19

2. REVISÃO DA LITERATURA 21

2.1. Marcadores biológicos para identificação humana 22

2.2 . Polimorfismos genéticos e análise do DNA 22

***2.2.1. Short Tandem Repeats* (STRs)** 23

2.2.2. *Single Nucleotide Polymorphisms (*SNPs) 25

2.2.3. *Insertion/Deletion Polymorphisms* (InDels) 26

2.3 Cromossomo X 29

2.3.1. Marcadores genéticos do cromossomo X 30

2.3.2. Aplicação dos marcadores genéticos do cromossomo X 32

2.4. Aspectos históricos do povoamento do Brasil 34

2.5. Justificativa e hipótese do estudo 36

3. OBJETIVOS 38

3.1. Objetivo Geral 39

3.2. Objetivos Específicos 39

4. MATERIAIS E MÉTODOS 40

4.1. Aspectos éticos da pesquisa 41

4.2. Sujeitos da pesquisa 41

4.3. Extração do DNA 41

4.4. Amplificação do DNA e Genótipo dos X-InDels 42

4.5. Análise dos resultados e dos parâmetros estatísticos 46

4.6. Análise comparativa 46

5. RESULTADOS E DISCUSSÃO 48

5.1. Avaliação da metodologia 49

5.2. Variante alélica identificada nas populações estudadas 50

5.3. Variabilidade genética populacional 53

5.4. Desequilíbrio de ligação entre os marcadores X-InDels 56

5.5. Parâmetros estatísticos de eficiência forense 61

5.6. Comparação genética entre as populações e análise de ancestralidade 66

6. CONCLUSÕES 74

7. REFERÊNCIAS 76

ANEXO I 85

1. INTRODUÇÃO

Desde os anos 80, os avanços nas técnicas de análise do ácido desoxirribonucleico (DNA) vêm se tornando uma poderosa aliada na identificação humana e para a investigação criminal proporcionando um significativo impacto no campo da ciência forense (ZIĘTKIEWICZ *et al*, 2012). O aprimoramento destas técnicas favoreceu o alto poder de discriminação de um indivíduo/vestígio biológico pelo DNA, o qual permite comparações genéticas entre indivíduos e/ou material biológico (THOMPSON; ZOPPIS; MCCORD,2012)

O genoma humano possui mais 3 bilhões de pares de bases localizados em cada célula nucleada do corpo. O genoma, que se manteve bem conservado ao longo da evolução, é no mínimo 99,5% idêntico entre quaisquer seres humanos do planeta e esta pequena fração variável é o que nos torna único (LEVY *et al*, 2007) e são essas regiões variáveis que possibilitam que o DNA seja utilizado na identificação humana (BUTLER, 2007).

Para a identificação humana são considerados marcadores genéticos os polimorfismos herdáveis, que podem ser identificados em uma ou mais populações (DAVEY *et al*, 2011). Os marcadores nucleares autossômicos estão amplamente padronizados na prática forense, pois apresentam alto grau de diversidade e alta resolubilidade dos casos (ALVAREZ-CUBERO *et al*, 2012).

Entretanto, há circunstâncias em que o material biológico pode ser coletado de casos *post-mortem* oriundos de corpos exumados, putrefeitos e de vestígios. Nessas condições, o DNA disponível pode estar degradado e em baixa quantidade, dificultando a solução do caso com a análise apenas dos marcadores nucleares autossômicos. Nesse contexto, a análise de outros marcadores e em outras regiões polimórficas, como no cromossomo X, podem complementar a análise tradicional de forma eficaz (ALVAREZ-CUBERO *et al*, 2012).

2. REVISÃO DA LITERATURA

2.1. Marcadores biológicos para identificação humana

A identificação humana pelo DNA baseia-se na individualização biológica que cada ser humano representa e, na exclusividade do seu perfil genético, exceto nos casos envolvendo gêmeos monozigóticos (PINHEIRO, 2004).

Os estudos modernos na identificação humana se iniciaram por volta do princípio do século XX, com a descoberta dos grupos sanguíneos ABO (JORDE; WATKINS; BAMSHAD, 2001), sendo tal sistema aplicado por várias décadas nos exames forenses de amostras biológicas em evidências relacionadas a crimes ou à identificação de pessoas.

Marcando uma segunda fase na evolução desta ciência, em 1954, foi demonstrada a ocorrência de um sistema de histocompatibilidade mediado por antígenos na superfície dos leucócitos, conhecido por complexo HLA (*histocompatibility leucocyte antigen*), os quais são determinados por genes alélicos muito próximos localizados no braço curto do cromossomo 6, com acentuado poder de discriminação individual ou determinação da individualidade genética (CALABREZ, 1999).

A terceira fase do desenvolvimento das ciências forenses voltadas à identificação humana veio com a publicação de um artigo na Revista *Nature* pelo geneticista inglês Alec Jeffreys e seu grupo de trabalho, que revolucionou a Genética Forense. Este grupo de trabalho descobriu que em um dos *íntrons* do gene codificador da proteína mioglobina, havia certas regiões do DNA que variavam de um indivíduo para o outro, produzindo perfis genéticos conhecidos como “impressão digital do DNA” ou “*DNA fingerprinting*” (JEFFREYS; WILSON; THEIN, 1985a).

A tipagem molecular de material genético foi utilizada oficialmente pela primeira vez, em 1985, por Jeffreys, na Inglaterra para a resolução de um problema de imigração (JEFFREYS *et al*, 1985b). Um ano após, o mesmo autor empregou esta técnica para identificar o verdadeiro estuprador e assassino de duas vítimas. A partir deste caso, que ficou conhecido como Enderby (Queen x Pitchfork), a tipagem molecular do DNA passou a ser potencialmente utilizada na rotina forense (TANDE, 1989).

2.2. Polimorfismos genéticos e análise do DNA

Desde a descoberta dos grupos sanguíneos os polimorfismos em humanos vêm auxiliando a resolução de questões criminais. Passando pelos polimorfismos de grupos

sanguíneos, polimorfismos proteicos/enzimáticos e pelos primeiros polimorfismos de comprimento de fragmentos de DNA (RFLP), obtidos através do tratamento do DNA com enzima de restrição. O grande salto deu-se com a identificação dos polimorfismos de sequência repetitiva do tipo microssatélites.

A palavra polimorfismo tem sua origem no grego antigo, onde “poli’ significa “muitos” e “*morphos*” significa “formas”, portanto, polimorfismos exprime a ideia de “muitas formas”. Uma variação na sequência de DNA que ocorre em uma população com ao menos dois alelos, e ambos com frequência mínima de 1% é denominada como polimorfismo genético (NUSSBAUM, 2008; BEIGUELMAN, 1996).

Os polimorfismos resultam de uma extensa variação genética entre os indivíduos, sendo originária de mutações e forças evolucionárias, como seleção, deriva genética e recombinação, que resultam em genomas únicos com características particulares (HARTI; CLARK, 2006). Essas variações são constituídas por polimorfismos de sequência ou de comprimento, a depender das alterações ocorridas na sequência do DNA.

A grande maioria dos polimorfismos de DNA pode ser dividida em dois grupos: os baseados em substituições nucleotidicas (SNPs – *Single nucleotide polymorphism*), que são os polimorfismos de sequência, e os polimorfismos de comprimento, baseados em inserções e deleções de um ou mais nucleotídeos, os mais conhecidos são os InDels e os *Short Tandem Repeats* (STR) (BUTLER, 2005).

2.2.1. Short Tandem Repeats (STRs)

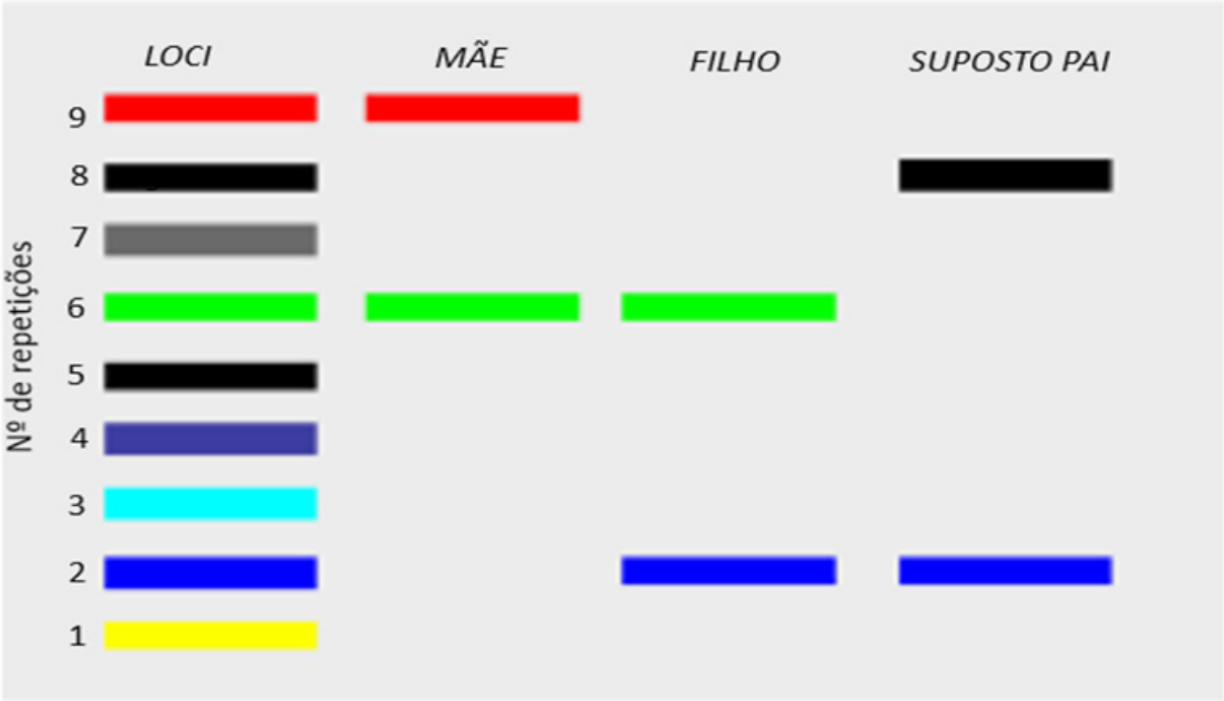
Os microssatélites, ou STRs (*Short Tandem Repeats*), estão entre os *loci* mais polimórficos no genoma humano e possuem características que os tornam marcadores de escolha em diversas temáticas forenses, tais como uma ampla distribuição no genoma humano e elevado poder de discriminação (CHARLESWORTH *et al*, 1994). Os STRs são polimorfismos de comprimento, com unidades de 2 a 6 nucleotídeos que se repetem em tandem por dezenas de vezes. Estes loci são altamente propensos a mutações, apresentando uma taxa de mutação de aproximadamente $2x10^{-3}$/gerações, isso se deve à sua susceptibilidade aos eventos de deslizamento da DNA polimerase durante a replicação do DNA (BUTLER, 2007).

Os STRs substituíram rapidamente os marcadores convencionais de grupos sanguíneos e de proteínas séricas, bem como as sondas de número variável de repetições em tandem (VNTR), que foi a primeira geração de sistemas de tipagem de

DNA baseados em hibridação que eram bastante poderosos, mas difíceis de usar pois, requer DNA íntegro e em grande quantidade (100 - 500ng), tornando praticamente inviável a tipagem de amostra biológica antiga, degradada ou para pouca quantidade de DNA (HALLENBERG; MORLING, 2001). Estes marcadores são facilmente amplificados pela reação em cadeia da polimerase (PCR), o número de repetições pode ser altamente variável entre os indivíduos (BUTLER, 2011) e, 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. Além disso, a detecção pode ser realizada em sistemas fluorescentes que permitem a automação dos sistemas de eletroforese e a interpretação dos perfis de DNA (BUTLER, 2005).

Os STRs são amplamente utilizados na prática forense. Na maioria dos casos, a análise de 20 STRs autossômicos é suficiente para a identificação humana ou para casos de trio (suposto pai/mãe/filho[a]) e duo (pai/filho[a]) na investigação de paternidade (YUAN *et al*, 2014). Para os testes de paternidade, a utilização dos marcadores STRs torna a análise e a visualização dos resultados mais simples. O filho herda um cromossomo do pai e outro da mãe, portanto, o filho terá marcadores STRs conforme o número de repetições atribuído a ele (figura 1) (FRAIGE, 2007).

Figura 1:Esquema representando a simulação do resultado de um sistema de STRs em eletroforese para investigação de paternidade comparando o número de repetições de um determinado loci pesquisado



Fonte: adaptado de FRAIGE, 2007

Por outro lado, como abordado por Schneider (2012), nos casos de relações de parentesco complexas, por exemplo, há certas desvantagens em sua utilização, uma vez que os marcadores STRs podem sofrer mutações que dificultam a identificação. Esta desvantagem ocorre principalmente em casos de reconstrução genética, em que o suposto pai está ausente, principalmente quando, são utilizadas amostras de parentes não tão próximos como primos e meios-irmãos. Pois, quanto mais distantes forem os parentescos genéticos dos indivíduos, maior é a probabilidade de observar mutações nos STRs analisados, gerando incompatibilidade. Essas mutações requerem testes adicionais envolvendo outros marcadores para compensação (PINTO *et al*, 2013a).

Essas propriedades dos marcadores STRs podem ser compensadas por sistemas de marcadores genéticos que estão distribuídos mais amplamente no genoma e são menos propensos a eventos de mutação. Nesse contexto, há autores que sugerem os marcadores bialélicos como complemento, tais como os polimorfismos de base única (SNP) ou de inserção/deleção (InDel), uma vez que esses são muito menos susceptíveis a mutações quando comparados aos STRs (PINTO *et al*, 2013a), o que os tornam melhores para analisar amostras de DNA degradado.

2.2.2. *Single Nucleotide Polymorphisms (*SNPs)

Polimorfismos de nucleotídeo único (SNPs) são variações na sequência de DNA (figura 2) que ocorrem quando um único nucleotídeo na sequência do genoma é alterado, sendo o tipo mais comum de variação genética entre as pessoas (BUDOWLE; van DAAL, 2008).

Figura 2:O polimorfismo de nucleotídeo único

Fonte: adaptado de NUSSBAUM et. al, 2016.

As vantagens de analisar SNPs como marcadores complementares aos STRs em casos forenses têm sido objeto de muito debate nos últimos anos. Entre as vantagens destes polimorfismos está a capacidade de gerar *amplicons* mais curtos do que é

possível com análise com STR, tornando assim os SNPs mais vantajosos na análise de DNA degradado (JOBLING; GILL, 2004; PIMENTA; PENA, 2010). Outras vantagens destes marcadores, incluem uma taxa de mutação menor do que aos dos STR, aproximadamente 10^{-8}/geração, tornando-as adequadas como ferramenta complementar, para análise de paternidade e testes complexos de relacionamento, e a ausência de *stutter* (produtos de amplificação que apresentam uma unidade de repetição a mais ou a menos do que um alelo verdadeiro, devido a erro da DNA polimerase durante a PCR), o que facilita a correta interpretação (SOBRINO; BRION; CARRACEDO, 2005; BUTLER, 2003).

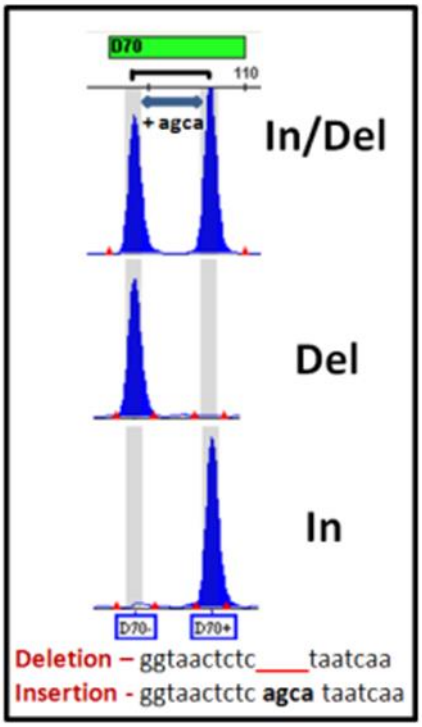
Apesar destes marcadores apresentarem uma diversidade genética menor que os STRs, estes polimorfismos constituem uma excelente ferramenta que pode contribuir nas resoluções de análises mais complexas (NACHMAN; CROWELL, 2000; PIMENTA; PENA, 2010).

2.2.3. *Insertion/Deletion Polymorphisms* (InDels)

Polimorfismos de inserção/deleção (InDels) constituem na inserção ou deleção de um ou mais nucleotídeos no genoma. Os InDels são abundantes, representando aproximadamente de 16% a 25% de todos os polimorfismos no genoma humano. Semelhante aos SNPs, também possuem uma baixa taxa de mutação, na ordem de $2{,}3 \times 10^{-9}$ e uma baixa recorrência, o que se refere a uma mutação que não ocorre mais de uma vez, ou seja, eventos mutacionais de origem única. A baixa taxa de mutação, torna esses marcadores mais adequados para investigar questões de ancestralidade. A simplicidade da genotipagem de InDels também é uma vantagem no estudo de grandes conjuntos amostrais populacionais (NACHMAN; CROWELL, 2000; IBARRA *et al*, 2014; WEBER *et al*, 2002; PIMENTA; PENA, 2010).

Os dois alelos dos marcadores InDels são compostos por um alelo classificado como alelo curto e outro como longo (BUTLER, 2005). O tamanho reduzido e a necessidade de verificar apenas a presença ou ausência de uma inserção/deleção tornaram esses marcadores mais práticos e fáceis de analisar (figura 3).

Figura 3: Eletroferograma de detecção de Inserção e Deleção



Fonte: Fondevila M. *et al.* (2011)

Apenas nos últimos anos esses polimorfismos têm recebido atenção dos estudiosos da área forense (PEREIRA *et al*, 2009). Em 2002, Weber e colaboradores identificaram e caracterizaram 2.000 InDels com diferentes variações de comprimento no genoma humano e calcularam as frequências desses alelos em Europeus, Africanos, Japoneses e Nativo Americanos. Em 2006, Mills e colaboradores iniciaram um mapa da variação de InDels no genoma, o qual continha mais de 415.000 polimorfismos descritos.

Utilizado como complementação à análise dos STR, os InDels podem superar as limitações da análise tradicional em casos de relações de parentesco complexas, pois apresentam baixa taxa de mutação em comparação com os STR; são amplamente distribuídos no genoma, sendo encontrados em média a cada 7,2kpb (MILLS, 2006); podem ser detectados por diferentes técnicas (MULLANEY *et al*, 2010), o que favorece verificar o vínculo biológico entre dois indivíduos. Além disso, as técnicas utilizadas na identificação humana geram *amplicons* de tamanho reduzido (50-150pb), o que favorece a análise de material biológico comprometido e, assim como para os STRs, estes podem ser analisados em multiplex, ou seja, em uma única reação (TORRES *et al*, 2014).

Dixon e colaboradores (2005), sugeriram que, durante a morte celular por apoptose ou necrose, o nucleossoma pode oferecer proteção contra as enzimas que

degradam o DNA a aproximadamente 146 pares de bases. Portanto, para a análise de material biológico degradado, técnicas que geram *amplicons* de tamanho reduzido podem apresentar maior sucesso na amplificação do DNA.

Apesar de ser uma importante ferramenta de complementação da análise tradicional em casos inconclusivos, para a análise de InDels em cromossomos autossômicos existe atualmente apenas dois kits comerciais, o DIPplex Kit® - Qiagen sendo sua eficácia na identificação humana comprovada em diversas populações, inclusive no Brasil (WANG *et al*, 2014; TORRES *et al*, 2014; PALHA *et al*, 2015). O outro kit, InDelPlex Kit® - Genomica, é o resultado do trabalho realizado por Pereira e colaboradores (2009), que descreveram esse novo multiplex para a identificação humana na qual foi possível analisar 38 InDels autossômicos em uma única reação. Os *amplicons* gerados na PCR possuem no máximo 160 pb e podem ser analisados diretamente em analisadores genéticos de eletroforese capilar. Os autores ressaltam que bons resultados foram identificados em concentrações de 0,3 a 5ng de DNA, possibilitando sua utilização também em casos de amostras degradadas. A análise do desequilíbrio de ligação e da probabilidade de coincidência de perfis genéticos em Africanos, Europeus e Asiáticos mostrou a eficiência desse multiplex na diferenciação entre os grupos, indicando a possibilidade de uso na rotina forense.

Partindo desse multiplex de 38 InDels (PEREIRA *et al*, 2009), Manta e colaboradores (2012a) avaliaram sua aplicabilidade em casos forenses no Brasil. A estatística mostrou diferenças entre as frequências alélicas da população do Rio de Janeiro e de Europeus, Africanos e Asiáticos, apresentadas por Pereira e colaboradores (2009). Diante dos casos forenses, os autores analisaram restos humanos em testes de paternidade e identificaram uma eficiente amplificação de DNA degradado, o que contribuiu com a resolução de processos em que apenas o uso de STRs foi insuficiente. Em 2018, um esforço colaborativo foi realizado pelo Grupo de Trabalho de Língua Espanhola e Portuguesa da Sociedade Internacional de Genética Forense (GHEP-ISFG), para promover o intercâmbio de conhecimento entre laboratórios associados, interessados na implementação de metodologias baseadas em InDel, e construir bancos de dados de frequência de alelo de 38 InDels para aplicações forenses (PEREIRA *et al*, 2018).

2.3 Cromossomo X

O cromossomo X é um dos cromossomos sexuais em humanos, sendo o outro o cromossomo Y. Cada pessoa possui normalmente um par de cromossomos sexuais em cada célula, sendo que as mulheres herdam um cromossomo X de cada um dos genitores e os homens herdam o cromossomo X materno e o Y paterno (NLM, 2012). Com esse padrão de herança parental, o cromossomo X se desloca entre ambos os sexos a cada geração, o que possibilita investigar o padrão migratório de homens e mulheres ao longo da história. Além disso, sua recombinação ocorre somente em mulheres, o que leva a uma menor diversidade genética quando comparada aos autossomos e, por consequência, um maior desequilíbrio de ligação; já em homens, a falta de recombinação permite a identificação direta de haplótipos (PEREIRA et al, 2007; ZARRABEITIA *et al*., 2009; MARTINS et al, 2014).

Este cromossomo tem aproximadamente 155 Mb e foram identificados 1.098 genes, representando 4% do total de genes humanos. Logo, o cromossomo X é um dos maiores cromossomos humano, mas possui uma densidade gênica um pouco menor que os cromossomos autossômicos. Cerca de 5% do genoma das células femininas e 2,5% do das células masculinas é representado pelo cromossoma X. A quantidade de regiões repetitivas é elevada, correspondendo a 56% do DNA deste cromossomo, sendo que a média do genoma é de 45%. Dentro dessas regiões repetitivas foram identificados os elementos intercalares curtos (SINEs – *Short Interspersed Nuclear Elements)* em quantidades abaixo da média dos cromossomos e os elementos intercalares longos (LINEs – *Long Interspersed Nuclear Elements*) em quantidades acima da média (ROSS *et al*, 2005).

Indivíduos do sexo feminino têm o dobro de cópias dos genes ligados ao cromossomo X em relação ao sexo masculino. Com a finalidade de tornar proporcional os produtos gênicos do cromossomo X entre homens e mulheres, um dos cromossomos X da mulher é inativado, aleatoriamente, no início do desenvolvimento e este estado inativo é transmitido para todas as células (CHOW *et al*., 2005). Na linhagem germinativa, o segundo cromossomo X é reativado durante a ovogênese para recombinar-se com o outro cromossomo X feminino (GRIFFITHS *et al*., 2009). No homem, o cromossomo X não se recombina ao longo de todo o seu comprimento, sendo tal processo restrito a algumas regiões pseudoautossômicas (PAR, *pseudoautosomal region*) 1 e 2, localizadas nas extremidades dos cromossomos X e Y. Tais regiões são necessárias para a normal

segregação dos cromossomos sexuais durante a meiose (ELLIS & GOODFELLOW, 1989).

O cromossomo X exibe características genéticas distintas em relação aos autossomos, especialmente: (1) menor tamanho populacional efetivo e (2) taxa de mutação menor. Todos os cromossomos masculinos possuem maior taxa de mutação que os femininos devido ao elevado número de mitoses nas células masculinas. Como dois cromossomos X estão presentes nas mulheres e apenas um nos homens a taxa de mutação global deste cromossomo é inferior à dos autossomos e (3) maior desequilíbrio de ligação, uma vez que o cromossomo X recombina apenas na meiose feminina (ZARRABEITIA *et al*., 2009).

2.3.1. Marcadores genéticos do cromossomo X

Devido ao seu alto poder de individualização e praticidade, o cromossomo X vem ganhando significativa importância em estudos de genética forense. Em testes de paternidade em que a análise estatística se apresenta inconclusiva, a análise de marcadores presentes no cromossomo X pode apresentar parâmetros mais eficientes quando comparados aos autossomos, especialmente quando as relações pai e filha ou mãe e filho estão sendo questionadas. (PEREIRA *et al*, 2012).

O cromossomo X apresenta diversos tipos de polimorfismos, que não se diferenciam dos descritos nos demais. Nele também são encontrados inserções, deleções, substituições, minissatélites e microssatélites. Na prática forense, as variações genéticas (STRs, SNPs e InDels) e as metodologias de genotipagem utilizadas são as mesmas também descritas para os marcadores autossômicos (TILLMAR *et al*., 2017).

Em termos estatísticos, os parâmetros de conteúdo de informação polimórfica (PIC) (BOTSTEIN *et al*, 1980) e a heterozigosidade esperada (NEI; ROYCHOUDHURY, 1974) são calculados da mesma forma para os marcadores autossômicos e cromossomo X. Já a chance média de exclusão (MEC) calculada para os autossômicos (KRÜGER *et al*., 1968) não pode ser a mesma para X, exceto nos casos complexos de parentesco em que se analisa amostras de avós. Assim, Desmarais *et al.* (1998) introduziu uma possibilidade de cálculo de MEC para trios e duos, bem como do poder de discriminação em indivíduos do sexo masculino e do sexo feminino para esses marcadores.

Os dois primeiros microssatélites descritos no cromossomo X (X-STRs) que apresentaram um importante papel foram HPRTB (HEARNE; TODD, 1991; EDWARDS *et al*., 1992) e ARA (EDWARDS *et al*., 1992; DESMARAIS *et al*., 1998). Após esses, mais

de 50 outros X-STRs foram identificados e Szibor e colaboradores (2007) descreveram os 30 marcadores de uso forense mais conhecidos até aquela época.

É importante destacar que os marcadores X-STRs utilizados na identificação humana e testes de paternidade estão localizados em regiões não recombinantes entre os cromossomos sexuais X e Y. Entretanto, ocorre recombinação entre os cromossomos X na mulher, garantido a variabilidade desses marcadores. Já nos homens isso é inviável devido a hemizigose do cromossomo X nestas regiões (SZIBOR *et al*., 2007).

Para a análise desses marcadores, alguns sistemas multiplex comerciais foram desenvolvidos. O primeiro deles, Mentype® Argus X-UL (Biotype Diagnostic GmbH, Dresden, Alemanha), amplificava 4 marcadores X-STRs e permaneceu no mercado entre 2003 e 2008, mas não foi disponibilizado no Brasil. Desenvolvido pelo Grupo Espanhol-Português da *Internacional Society for Forensic Genetics* (ISFG), X-Decaplex (GUSMÃO *et al*, 2009), consiste em um sistema com amplificação simultânea de 10 marcadores X-STRs Em 2005, a empresa Biotype (Dresden, Alemanha) lança um novo kit, Mentype® Argus X-8, com a amplificação de 8 marcadores mais amelogenina, porém, também não foi disponibilizado no Brasil. O mais recente kit, dessa vez disponível no Brasil, é o *Investigator Argus X-12* (Qiagen, Hilden, Alemanha), que amplifica simultaneamente 12 X-STRs mais amelogenina (http://www.chrx-str.org/, acesso em 21 de março de 2019).

Para a análise de marcadores X-STRs, estudos populacionais vêm sendo produzidos em diversas partes do mundo (DONG *et al*, 2014) e um banco de dados de marcadores X-STRs está disponível para análise dos perfis genéticos das populações (www.bgbx.com.br) (MARTINS *et al*, 2014). O Banco Genético Brasileiro do cromossomo X foi criado com o objetivo de reunir dados genéticos da população brasileira para polimorfismos X-STRs, caracterizando-se como uma ferramenta útil para a Polícia Técnico-Cientifica, os laboratórios de DNA e a comunidade acadêmica (MARTINS *et al*, 2014). Nele há dados de frequências alélicas e haplotipicas, parâmetros estatísticos de interesse forense e informações sobre a sequência de *primers* de trabalhos publicados e indexados na base de dados PubMed (MARTINS *et al*, 2014).

Entretanto, quando tratamos de marcadores InDels para esse cromossomo, pouco se tem publicado. Um grupo de pesquisadores chineses elaboraram um multiplex com 18 marcadores X-InDels que, em conclusão, alegam atender os requisitos para ser utilizado como ferramenta complementar na análise de material genético, fornecendo dados complementares eficazes para a resolução de casos difíceis (SUN *et al*, 2014).

Buscando a padronização de marcadores X-InDels para uso forense, um grupo brasileiro (FREITAS *et al*, 2010) desenvolveu uma metodologia utilizando 33 X-InDels em

uma única reação de PCR em multiplex. Os autores demonstraram a possibilidade de aplicação da metodologia em casos de identificação humana e testes de paternidade em uma população miscigenada da Amazônia e concluíram que se trata de marcadores robustos e eficientes, que podem facilmente ser adotados na rotina forense.

Em 2017, Caputo e colaboradores analisaram a eficiência forense do painel 33 X-InDels em amostras da população da Argentina. Os autores identificaram que 5 marcadores e 2 blocos haplotípicos apresentaram heterozigosidade esperada abaixo de 20%, indicando que este painel seria pouco informativo para tal população. O poder de discriminação e a chance média de exclusão também foram analisados e estes apresentaram resultados levemente abaixo de outros painéis já estudados. Esses resultados possivelmente se devem ao fato de que muitos dos marcadores analisados estão associados com a ancestralidade africana, o qual não consiste no principal componente ancestral dos argentinos.

Nessa mesma perspectiva, Pereira e colaboradores (2012) desenvolveram uma PCR em multiplex com 32 marcadores X-InDel e identificaram seu potencial uso para a identificação humana e em testes de paternidade em Africanos, Europeus e Asiáticos. Quando comparado ao estudo descrito anteriormente, esses 32 marcadores possuem a vantagem de gerarem *amplicons* curtos, o que é mais adequado para amostras provenientes de casos forense, em especial aqueles em que o DNA se encontra degradado.

Explorando ainda mais a potencialidade desses marcadores, Ibarra *et al.* (2014) utilizaram o multiplex 32 X-InDels (PEREIRA *et al*, 2012) para ampliar o conhecimento atual sobre a base genética das populações da Colômbia. Com o uso desses marcadores foi possível realizar uma caracterização ancestral da população colombiana, apontando diferenças significativas entre as diferentes regiões geográficas do país.

Também utilizando o multiplex 32 X-InDels (PEREIRA *et al*, 2012) em amostras da população de São Paulo, Martinez e colaboradores (2019) relataram que este painel se mostrou polimórfico para esta população sendo bastante informativo. A eficiência forense também foi analisada e apresentou um alto poder discriminativo.

2.3.2. Aplicação dos marcadores genéticos do cromossomo X

As propriedades de transmissão do cromossomo X o torna altamente interessante para estudos de identificação de indivíduos e de diversidade genética nas

populações humanas. Dentro dos estudos da genética forense, o cromossomo X vem ganhando interesse nos últimos anos especialmente na análise das seguintes situações:

a) Análise de relações biológicas

Investigação de paternidade em casos envolvendo trios comuns com suposto pai, mãe e filha podem facilmente ser solucionados com o uso exclusivo de marcadores autossômicos, não necessitando de marcadores adicionais (SZIBOR *et al*., 2003). Se o teste de paternidade envolve um duo, com suposto pai e filho, os X-STRs também não são utilizados, já que não há herança paterna do cromossomo X.

Por outro lado, para casos de paternidade que envolvem um duo (suposto pai e filha) em que se observa uma ou duas incompatibilidades, sendo necessária a confirmação de exclusão ou ocorrência de possíveis mutações, os marcadores no cromossomo X são capazes de complementar a análise de forma eficiente. Como o pai transfere 100% de seu material genético do cromossomo X para a filha, os marcadores nesse cromossomo apresentam elevada chance média de exclusão (MEC, do inglês Mean Exclusion Chance), o que significa que possuem alta capacidade de exclusão nas investigações de parentesco (SZIBOR *et al*., 2003; SZIBOR, 2007).

Nas investigações de maternidade entre suposta mãe e filha esses marcadores não apresentam grande contribuição, mas na investigação entre suposta mãe e filho esses marcadores também apresentam maior MEC e podem colaborar com a resolução do caso (SZIBOR *et al*., 2003).

Estudos de irmandade ou meia-irmandade, em indivíduos do sexo feminino, apresentam ótimos resultados com X-STRs. Como há apenas uma possibilidade de cromossomo X a ser transmitido do pai para todas as suas descendentes do sexo feminino, todas as irmãs e meias-irmãs por via paterna devem compartilhar do mesmo haplótipo paterno. A análise exclusiva dos cromossomos autossômicos nesses casos não é capaz de fornecer essa informação, pois eles não são capazes de excluir esses casos (SZIBOR *et al*., 2003; SZIBOR, 2007).

Ainda na ausência do suposto pai, a análise do perfil genético de sua mãe (suposta avó) é uma importante chave no processo de reconstrução. Todos os alelos de X do suposto pai podem ser determinados pela análise genotípica de sua mãe e, portanto, esses devem ser compatíveis com os alelos da filha questionada. Os cromossomos autossômicos também não têm poder de exclusão nessas situações (SZIBOR *et al*., 2003).

b) Análise de relações biológicas através de restos mortais e de vestígios biológicos

Em testes de investigação de paternidade ou identificação humana com dificuldade para se analisar o material biológico, tais como em exumação, embora o material a ser genotipado é do suposto pai, o número de STRs amplificados com sucesso pode ser menor do que geralmente se obtém com amostras de sangue ou saliva e o laudo pode permanecer inconclusivo. Como os X- STRs possuem maiores valores de MEC que os autossômicos, é possível excluir uma relação de parentesco ou um aumento estatístico ao se complementar aos dados autossômicos, colaborando com resolução de casos em que só é possível obter um pequeno número de marcadores analisados (SZIBOR *et al*., 2003; SZIBOR, 2007).

Por outro lado, nas situações em que se pretende identificar um perfil feminino em meio a uma mistura contendo material genético masculino, os marcadores de X tendem a ser mais eficientes, pois os alelos femininos somente poderão coincidir completamente com os masculinos se a mulher for homozigotica para todos os loci (SZIBOR *et al*., 2003).

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, primeiros estrangeiros a entrar no Brasil, chegaram ao litoral brasileiro em 22 de abril de 1500 e encontraram a nova terra povoada por cerca de 5 milhões de nativos. A mistura, predominantemente entre homens europeus e mulheres nativas americanas, começou imediatamente, no entanto, conflitos e doenças contribuíram para uma redução drástica da população nativa americana (SALZANO; BOTOLINI, 2002). Segundo dados do IBGE (1998) estima-se que esta população fora reduzida a aproximadamente 302 mil indivíduos. Porém, o que muitas vezes se ignora no Brasil é que grande parte da população indígena não pereceu, mas foi assimilada pela sociedade brasileira. Muitos índios foram viver ao lado de português e africanos, com eles se miscigenando e dando origem a grande parcela da população brasileira, sendo

que seus descendentes não mais se identificam como “índios” (NARLOCH & TEIXEIRA, 2011).

Quando os descobridores portugueses se fixaram como colonizadores no território, que seria mais tarde o país Brasil, já trouxeram séculos de interação genética e cultural com outros povos de origens geográficas distintas. Os portugueses foram os principais colonizadores europeus a introduzirem suas heranças genéticas à nossa população, inicialmente com as mulheres indígenas, pois estudos apontam que os primeiros brasileiros foram resultados do cruzamento entre homens portugueses e mulheres ameríndias (CARVALHO-SILVA *et al*., 2001). Mas a posterior entrada maciça de escravos africanos permitiu a miscigenação a partir do cruzamento destes primeiros colonizadores com as mulheres africanas.

O Brasil foi, no continente americano, a região que mais importou escravos africanos durante os 300 anos de duração do tráfico transatlântico. Não se sabe ao certo o ano da chegada dos primeiros negros ao Brasil. Goulart (1975) sugere que este evento tenha ocorrido entre 1516 e 1526, com o início da exploração de cana-de-açúcar, em substituição a mão-de-obra indígena, procediam de várias áreas culturais da África. Estimam-se que cerca de 3 a 18 milhões de africanos teriam chegado ao país entre os séculos XVI e XIX, destes, 32 milhões teriam passado pelo Rio de Janeiro, principal porta de entrada de navios negreiros no país (MONTEIRO, 1994). Essa

Os escravos africanos foram utilizados não apenas na produção de produtos de exportação (açúcar, café, algodão, minérios e outros), mas também na agricultura de abastecimento interno, na criação do gado, trabalhos domésticos e toda ordem de ocupações urbanas. 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 (Brasil: 500 anos de povoamento, IBGE, 2007).

O desbravamento na região que hoje compreende o estado de Minas Gerais se iniciou no século XVI, por meio do trabalho dos bandeirantes, em busca de ouro e pedras preciosas. No início do século XVIII, a região tornou-se um importante centro econômico da colônia, com rápido povoamento (BUENO, 2009).

No período colonial (sec. XVI e XIX) o povoamento do Espírito Santo ficou limitado ao litoral. No século XIX, o interior do Espírito Santo ainda era coberto de matas virgens. Ainda no começo do século XIX, com a chegada dos plantadores de café vindos de Minas

e Rio de Janeiro e mais tarde dos imigrantes europeus estabelecidos na região central do Brasil, provocou o povoamento dessas áreas (NEVES *et al*., 1997).

A descoberta e a exploração de ouro em Minas Gerais, aumentou a importância do Rio de Janeiro, principalmente a atividade portuária. 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).

Movidos por interesses principalmente econômicos, importantes movimentos migratórios estavam ocorrendo dentro do território brasileiro. Com isso, a história de povoamento do Brasil se mostrou heterogênea nas diferentes regiões geopolíticas do país (Norte, Nordeste, Centro-Oeste, Sudeste e Sul), e a forma de distribuição desses grupos populacional, 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).

2.5. Justificativa do estudo

Na rotina forense, tanto no contexto da identificação humana quanto nas investigações de relações biológicas, há situações em que somente a análise dos marcadores STRs autossômicos não é suficiente para a resolução do caso, devendo ser complementada com a análise de outros marcadores e de outras regiões polimórficas. Marcadores InDels podem superar as limitações dos marcadores tradicionais e solucionar casos tratados como inconclusivos.

Diante desses aspectos, Pereira e colaboradores (2012) desenvolveram um multiplex com 32 marcadores X-InDels e identificaram seu potencial uso na prática forense em Africanos, Europeus e Asiáticos. Utilizando este mesmo multiplex, Martinez e colaboradores (2019), analisaram a aplicabilidade deste multiplex em amostras da população do estado de São Paulo, onde evidenciaram a eficiência forense destes marcadores para populações miscigenadas

Tendo em vista que as populações de Minas Gerais, Espírito Santo e Rio de Janeiro não foram analisadas por Pereira e colaboradores (2012) e Martinez e colaboradores (2019), assim como a eficiência forense dos marcadores descrita pelos

autores, estudos relacionados as variações no cromossomo X nas populações de Minas Gerais, Espírito Santo e Rio de Janeiro permitem compreender a diversidade desses marcadores, expandindo o conhecimento para toda região sudeste do Brasil, permitindo auxiliar na resolução de casos forenses em nosso país.

3. OBJETIVOS

3.1. Objetivo Geral

Descrever a variabilidade genética de um multiplex de 32 polimorfismos de inserção/deleção do cromossomo X nas populações dos estados de Minas Gerais, Espírito Santo e Rio de Janeiro-Brasil e comparar os valores obtidos com dados de outras populações disponíveis na literatura.

3.2. Objetivos Específicos

- Determinar as frequências alélicas, desequilíbrio de ligação entre os marcadores de 32 polimorfismos de inserção/deleção do cromossomo X em amostras da população dos estados de Minas Gerais, Espírito Santo e Rio de Janeiro;
- Analisar a eficiência forense dos 32 marcadores X-InDels nessas populações, calculando o poder de discriminação feminino (PDF), poder de discriminação masculino (PDM), chance média de exclusão em trios envolvendo filha (MEC_T); chance média de exclusão em duo pai/filha (MEC_D), e chance média de exclusão em casos deficientes envolvendo mãe, filha e avó paterna (MEC_{PGM});
- Verificar se há diferença genética significativa entre as populações estudadas com dados relatados na literatura;
- Analisar a ancestralidade das populações estudadas obtidas com dados baseados nestes marcadores e comparar com os dados publicados de Y-SNP, DNA autossômico (específico para ancestralidade) e DNA mitocondrial.

4. MATERIAIS E MÉTODOS

4.1. Aspectos éticos da pesquisa

Este trabalho se enquadra dentro de um projeto mais amplo intitulado “Estudo de polimorfismo do cromossomo X na população da região sudeste do Brasil”, coordenado pelo Profa. Dra. Regina Maria Barretto Cicarelli e pela Dra. Joyce Aparecida Martins Lopes Ferraz. Aprovado pelo Comitê de Ética em Pesquisa da Faculdade de Ciências Farmacêuticas de Araraquara – UNESP. Protocolos: Minas Gerais - Hemominas – Ofício N° 89/2008 e Espírito Santo - CEP/FCF/CAr n° 02/2009.

4.2. Sujeitos da pesquisa

Para a caracterização genética dos 32 polimorfismos de inserção/deleção, foi estudado um total de 627 amostras sanguíneas de indivíduos aparentemente saudáveis, não aparentados e residentes em Minas Gerais (90 mulheres e 116 homens), Espírito Santo (201 homens) e Rio de Janeiro (220 homens). As amostras foram coletadas apenas após o esclarecimento ao participante sobre os objetivos e aspectos éticos da pesquisa e a assinatura do Termo de Consentimento Livre e Esclarecido.

Para a análise do perfil genético foi feita uma punção capilar no dedo anelar dos participantes da pesquisa para a coleta de amostras de sangue em papel FTA® (*Whatman*®) e estas foram armazenadas a 8ºC até o momento da extração do DNA.

Critérios de Inclusão

- Indivíduos sem relação biológica pelo cromossomo X;
- Idade mínima de 18 anos;
- Indivíduos que aceitaram participar do estudo.

4.3. Extração do DNA

A extração de DNA das amostras de sangue foi realizada com o auxílio da resina *Chelex* 100 (Bio-Rad), adaptado da metodologia proposta por Singer-Sam *et al.* (1989) e Walsh, Metzger e Higuchi (1991).

Com o auxílio da caneta Harris Micro Punch® (*Whatman*® GE), foram coletados 2 discos de 2mm do centro da gota de sangue do cartão FTA® (*Whatman*®) e depositados em tubo de 0,2ml. Os discos foram lavados com 50µL de água ultrapura (Milli-Q® -

Millipore Corporation) e agitados. O sobrenadante foi descartado e os discos incubado em 50µL de água ultrapura (Milli-Q® - Millipore Corporation) por 30 minutos em temperatura ambiente. Em seguida, o sobrenadante foi descartado e os discos incubado em 100µL de *Chelex* 100 (Bio-Rad) 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 por 3 minutos a 13000 rpm e o sobrenadante então transferido para outro tubo de forma que a resina não seja transferida.

Para obter melhores resultados na extração do DNA, algumas amostras foram extraídas utilizando o Kit DNA IQ (Promega), de acordo com as normas do fabricante.

Assim, promoveu-se a lise celular por meio de uma solução composta por 250µL de *Lysis Buffer* e 2,5µL de 1M Dithiothreitol (DTT). O DNA foi "capturado" por 7µL de uma resina composta por partículas magnéticas, lavado por três vezes com 100 L de solução *Wash Buffer* por lavagem e por fim eluído em 50 L de *Elution Buffer*.

4.4. Amplificação do DNA e Genótipo dos X-InDels

Para a amplificação do DNA foi solicitada a síntese dos 32 pares de *primers* marcados com as fluorescências 6-FAM™, VIC®, PET™, NED™ da Applied Biosystems® (Thermo Fischer Scientific), conforme sequência descrita por Pereira e colaboradores (2012). A distribuição dos 32 marcadores ao longo do cromossomo X assim como a sequências dos *primers* de PCR usadas no multiplex estão dispostos nas figuras 1 e 2.

Figura 4: Ideograma do cromossomo X indicando a localização dos 32 marcadores X-InDel selecionados.

Fonte: Modificado de Pereira *et al*., 2012

Tabela 1: InDels do cromossomo X e sequências iniciadoras de PCR usadas no multiplex

MID	número de rs	Posição (pb)	Alelos	*Primers Forward*	*Primers reverse*
MID2612	rs3048996	10234839	-/ATC	*g*ACCCACGGTGTTGAATTCAG	NED-CACAGCACCAGGAAAATAGC
MID3712	rs55877732	12572196	-/GAA	*gttt*AGTCTTGCTGCAATGTACCC	VIC-TTCAAGGGCAATGATGTTTG
MID357	rs25581	12912861	-/TGAGA	*g*TTTTATAGACTGTGGCCCCC	PET-GTTAGTGGTTGGATTGCTCG
MID356	rs25580	12918048	-/CTT	*gttt*CCAACTCCACGTGAGAAATG	PET-AGTCTGATGCAGTGGCAAAC
MID3703	rs59400186	13711300	-/GTTA	VIC-AGCTTCCAAGTAGTTCTGCC	*g*TTTGGCTTACTTCCTCCTCC
MID3774	rs5901519	13809000	-/ACC	*g*AAGACGGGAATTGAGTCACC	NED-TTTTTGTGCACAGGCACTCC
MID3692	rs67929163	19516252	-/CATAT	6FAM-ACATAAAAGCAAGCTTTGGC	*gtttctt*CCCGGTGTGTGAACTTTTTC
MID3716	rs63344461	24235114	-/GAG	6FAM-AAAGGGAGCATCTACTCCAG	*gtttct*AGGGCAATCCAGAATTGGAC
MID3690	rs60283667	28984076	-/TCAC	GGGCACCATATTAGGCATGT	VIC-CCCACCATCTAACCCATTTC
MID3719	rs3078486	29040938	-/TTAACT	*g*TTCTTTCTCATCTGGCACCC	VIC-CTATGAAGCCTATAGATTGG
MID2089	rs2308280	29157973	-/TTA	VIC-AATCCATTCTGGAATAAGATGTCA	*gtttctt*TCCACTCTCAGGGATTCCTTT
MID2692	rs3047852	38262701	-/ATT	*gtttctt*CAAGTTCATATGGTGTCTTGG	PET-TGCATTACACAGAGCAACTC
MID3701	rs4030406	45539201	-/ATTA	*gtttct*AGTTGGAGATGCAATGAAGC	NED-AGAGACAGGTGAATTGAGGC
MID198	rs16637	47680386	-/CAACCAAT	6FAM-CAGGCACAGGAGAGGAAGAG	*g*TCCACCCCTAGTTAAACAGC
MID1736	rs2307932	68733480	-/ATA	VIC-GTGAAAGGTGAGCTTGTCTG	*gtttct*AGGCCTTTTTGGTTAACTGG
MID3730	rs3215490	88009689	-/GACA	6FAM-AGGATCCTGACTAAGATAGC	GAATCTCTGGAAACACTTGG
MID1511	rs2307707	93392006	-/GTCT	*g*CTGCCTGGGATTTTTCCTTT	PET-CAGGGGAGAACACCCACTAA
MID3740	-	97906546	-/GT	*gtttct*ACTTGCTTTGCTTTTCCCTC	6FAM-GTACAACTGCAAGGAACRAGc
MID3732	rs11277082	98331815	-/ACCTCACTCA	VIC-CAGAGTCATCTATTCCCCAG	*gtttctt*CACCCATGTGGTTTCATTTC
MID3727	rs3050111	99165489	-/TT	*gtttct*GGTGGAATTCTTTTGCATGTG	PET-TTTTGGGAAGCACACTCACC
MID3754	rs57608175	116901987	-/GGTCATCACGAG	NED-TTTCACCAAGGACTTGAAAGG	*gtt*CAGCTCACACTAGGGCCTTC
MID3722	–	118156157	-/AAAGTGTACACAT	VIC-TGGCCCTTCTGAGTTCAAAC	*g*CAGTGTAATAAGGTGGGAGC
MID1361	rs2307557	118748515	-/ACA	6FAM-TCAGTCCTTAACAAGGGAGC	*gttt*GTCATTGTGAAGGCTACCTG
MID243	rs16680	122370414	-/TGT	PET-TACAGTTGGCTGCTTTTCCC	*gtttctt*ATACGAAGATCTGTGGGAAC
MID2637	rs3053615	124135529	-/CT	*gtttct*TATGTGTCAAAATGGGAGGC	6FAM-AATCCTCAAATCACAGTGGC
MID111	rs16397	127958384	-/GTG	GAGGCAGGGAAATCAGTTAG	NED-TTGATTCCAGCTTTCCCTTT
MID3736	rs56162621	130975547	-/CT	6FAM-GGGTTAGGAGCCCCTGCT	*gtttctt*GGATGTATGACACACAACGC
MID3753	rs72417152	131760172	-/GTATAT	GCTACACCAATGGACAGATG	PET-TGTGGTGTGCATGATTTGC
MID1839	rs2308035	135695920	-/CA	*gtt*GATATCCCATAACGCCCATTT	NED-TCCTTTTGCTACGCAGACCT
MID3760	rs66676381	137369795	-/TTAAA	*g*CAAGGTTCGTACTCATTTAG	NED-AACCTAGTTCACAACCCCTG
MID329	rs25553	147393784	-/TACTCT	*g*TCTCAAAACCTTCCCATGGC	6FAM-AGAAGTTAGAGGGTGTCTGG
MID2652	rs3080039	154561961	-/TAA	PET-GCTGCTCTTTGCTTTAATTTC	*g*TATGGTAGGCACTGTGCTAA

Fonte: Adaptado de Pereira *et al*., 2012; Letras minúsculas em itálico representam caudas de nucleotídeos adicionadas aos *primers*.

Para a Reação em Cadeia da Polimerase (PCR) em multiplex foi utilizado o protocolo descrito por Martinez (2017) que fora adaptado do originalmente proposto por Pereira e colaboradores (2012). Um controle positivo (amostra comercial 9947A, Promega, Wisconsin, EUA) e um negativo foram incluídos em todas as reações. Os *primers* foram aliquotados em um único mix e a PCR seguiu conforme descrito a seguir:

2,5µL de PCR Master Mix (Qiagen)
0,5µL do mix de *primers*
1,0µL de DNA extraído (~ 5ng de DNA)
Água ultrapura (Milli-Q® - Millipore Corporation) autoclavada q.s.p. 5µL

Obs: a mistura de primers contém todos os primers na concentração de 1µM exceto para os marcadores MID2089, MID3727 (2µM) e MID3719 (3µM)

Os fragmentos foram amplificados em termociclador Veriti® Thermal Cycler (Thermo Fisher Scientific), conforme ciclagem abaixo:

90ºC – 15 min
94ºC – 30 seg
60ºC – 90 seg } 29 ciclos
72ºC – 45 seg
72ºC – 60 min

Obs: Esta ciclagem apresenta uma modificação em comparação com a publicada por Pereira e colaboradores (2012). O passo de ciclagens foi reduzido de 30 para 29 ciclos para diminuir o sinal de fluorescência excessivo obtido no eletroferograma (PEREIRA *et al*. 2012).

Os produtos amplificados foram preparados para análise de *Human Identification* (HID) com 8,7µL de Hi-Di™ *Formamide* (Thermo Fisher Scientific) e 0,3µl de GeneScan™600 LIZ® (*Applied Biosystems*® by Life Technologies) e depois foram submetidos à eletroforese capilar pelo Analisador Genético ABI 3500 *Applied Biosystems*® (Thermo Fisher Scientific) com POP-4 Polymer da

Applied Biosystems® (Thermo Fischer Scientific) e capilar de 36cm da Applied Biosystems® (Thermo Fischer Scientific).

4.5. Análise dos resultados e dos parâmetros estatísticos

A análise dos perfis alélicos gerados foi realizada com o auxílio do *software* GeneMapper® ID-X v1.2.1 da Applied Biosystems® (Thermo Fischer Scientific).

Com o auxílio do programa Arlequin v. 3.5 (EXCOFFIER; LISCHER, 2010), os perfis genéticos foram utilizados para calcular a frequência alélica e desequilíbrio de ligação. O cálculo dos valores de heterozigosidade esperada e observada e o teste ao equilíbrio de Hardy-Weinberg foram realizados somente para a amostra dos genótipos femininos (Minas Gerais) com o auxílio do mesmo *software.*

A análise do desequilíbrio de ligação foi feita utilizando apenas os haplótipos masculinos. Para estas amostras, o número de passos de cadeia de Markov utilizados foi 100.000 e o número de passos de desmemorização foi 10.000.

Para essas análises, a correção para múltiplos testes proposta por Bonferroni foi utilizada (BONFERRONI, 1936). A correção de Bonferroni é usada 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 comparações que foram executadas.

Para identificar a eficiência forense desses marcadores foram calculados o poder de discriminação (PD) e de exclusão (PE), chance média de exclusão em trios envolvendo filha (MEC_T); chance média de exclusão em duo pai/filha (MEC_D), todos segundo Desmarais *et al*. (1998); e chance média de exclusão em casos deficientes envolvendo mãe, filha e avó paterna (MEC_{PGM}), segundo Krüger *et al*. (1968). Os resultados foram obtidos através da ferramenta de cálculo do banco de dados internacional de X-STRs (http://www.chrx-str.org/).

4.6. Análise comparativa

A fim de se obter uma visão mais completa da diversidade genética das populações de Minas Gerais, Espírito Santo e Rio de Janeiro, os resultados foram

comparados com populações (tabela 13) da Europa, África, América Latina, Nativo-Americano, Ásia e América do Norte.

A distância genética (F_{ST}) entre as diferentes populações foi calculada utilizando o software Arlequin v3.5 (EXCOFFIER; LISCHER, 2010). Para melhor visualização dos resultados, os dados de F_{ST} foram dispostos em um gráfico de Escalonamento Multidimensional (MDS) utilizando o *software* STATISTICA 13.2 Trial (StatSoft).

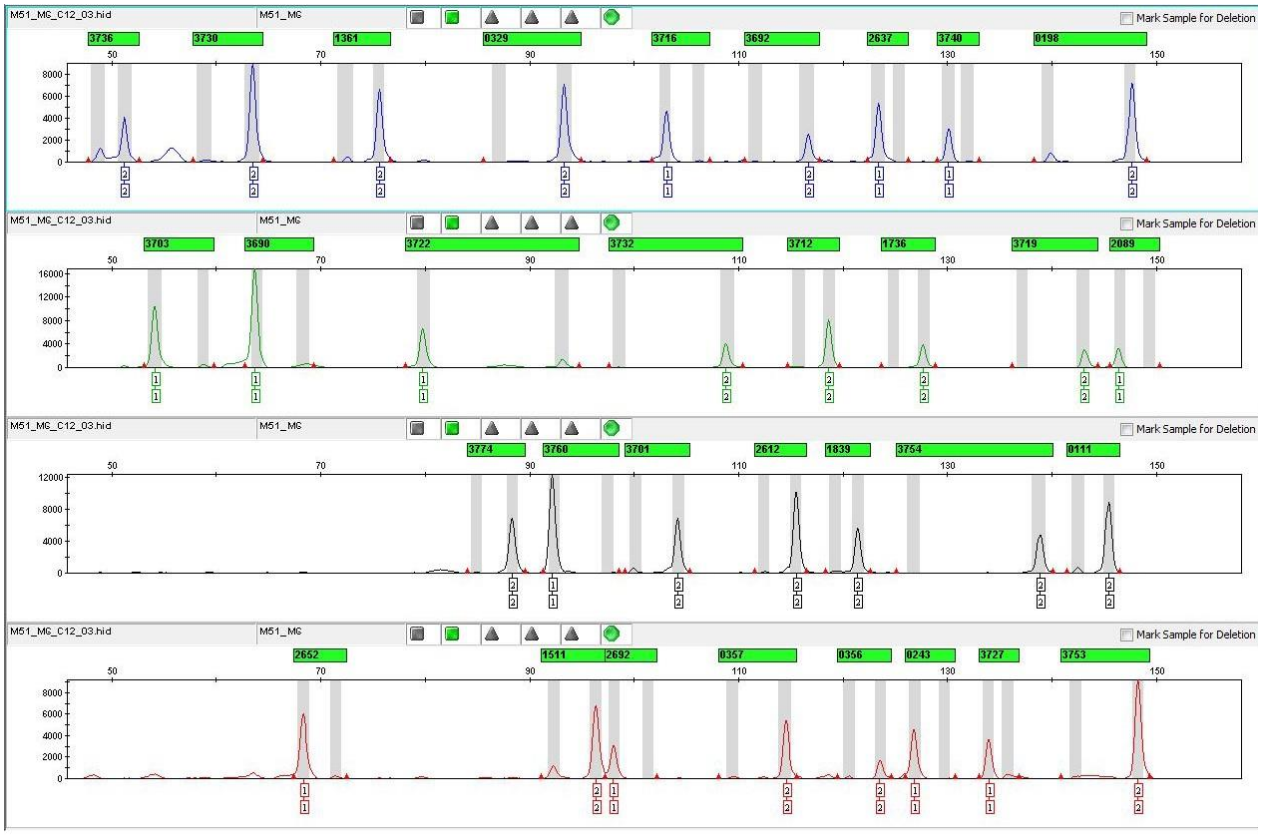
A proporção da contribuição ancestral genética foi estimada utilizando o *software* STRUCTURE v.2.3.4 (PRITCHARD; STEPHENS; DONNELLY, 2000). Considerando a formação histórica do Brasil, assumiu-se uma contribuição tri-híbrida de nativo-americanos, europeus e africanos (K = 3). Para a análise, três interações independentes foram realizadas com 100.000 passos de *burn-in* seguido por 100.000 passos de Cadeia de Markov Monte Carlo (MCMC), os valores estimados entre as três interações foram consistentes, sendo usada a interação com menor valor de ln Pr(X|K); a opção "*use population information to test for migrants*" foi utilizada no modelo *admixture*.

5. RESULTADOS E DISCUSSÃO

5.1. Avaliação da metodologia

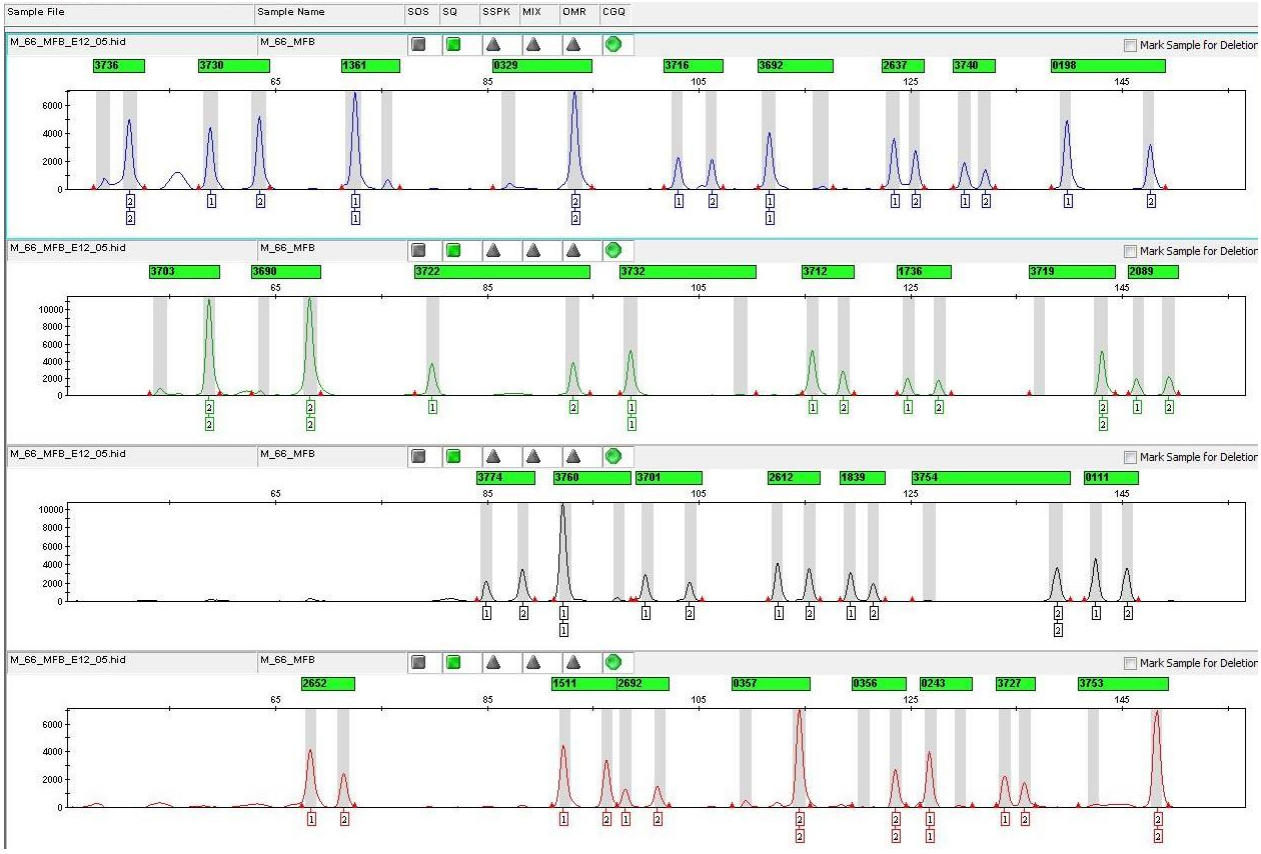
A reação de amplificação descrita por PEREIRA e colaboradores (2012) foi eficiente para análise dos 32 X-InDels. Os eletroferogramas dos perfis alélicos obtidos em analisador genético ABI3500 de amostras masculina e feminina estão apresentados nas figuras 3 e 4. Pode-se notar que a metodologia foi eficiente, pois todos os marcadores foram bem amplificados, obtendo-se uma adequada determinação dos alelos.

Figura 5: Eletroferograma de uma amostra masculina amplificada com o multiplex 32 X-InDels.



Fonte: Dados da pesquisa

Figura 6: Eletroferograma de uma amostra feminina amplificada com o multiplex 32 X-InDels.

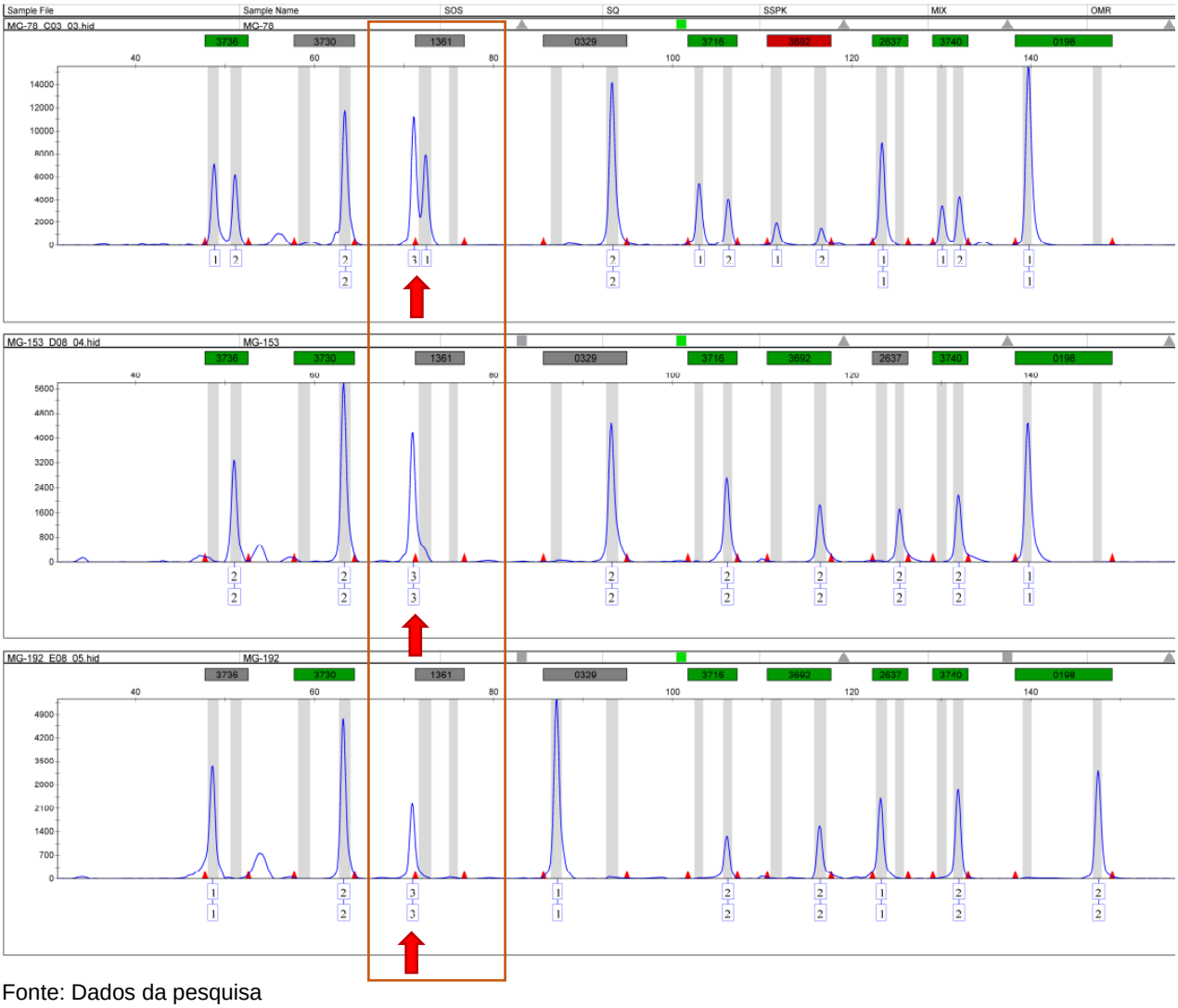


Fonte: Dados da pesquisa

5.2. Variante alélica identificada nas populações estudadas

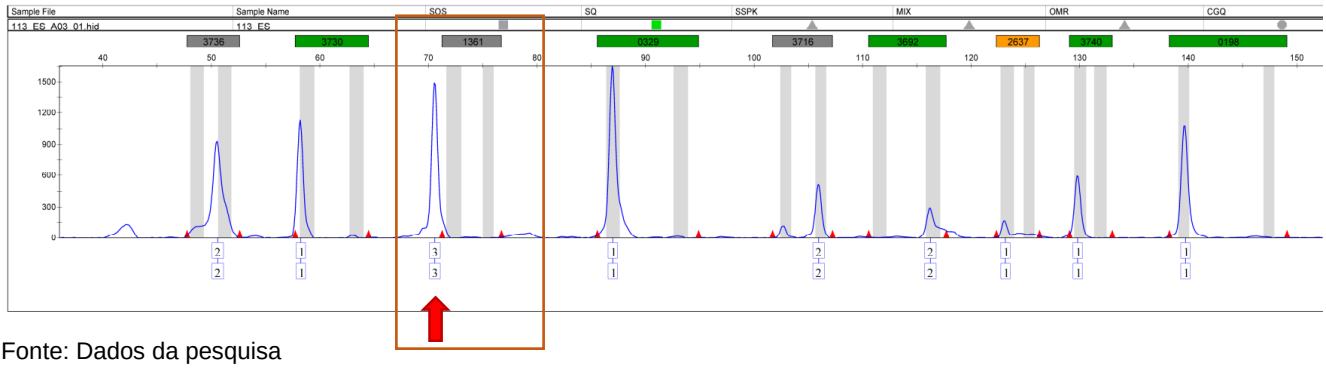
Em cada população estudada, foram obtidos mais de 200 perfis genotípicos de indivíduos sem relação biológica pelo cromossomo X, sendo 116 homens e 90 mulheres da população de Minas Gerais, 201 homens na do Espírito Santo e 220 homens da população do Rio de Janeiro. Na análise dos genótipos das amostras, foi observado em cinco amostras um alelo variante no marcador MID1361, sendo uma amostra feminina e duas masculinas na população de Minas Gerais (figura 8), uma na população do Espírito Santo (figura 9) e uma na população do Rio de Janeiro (figura10).

Figura 7: Eletroferograma parcial de três amostras de Minas Gerais obtidas com o multiplex 32 X-InDels. Em destaque a variante alélica do alelo curto do marcador MID1361



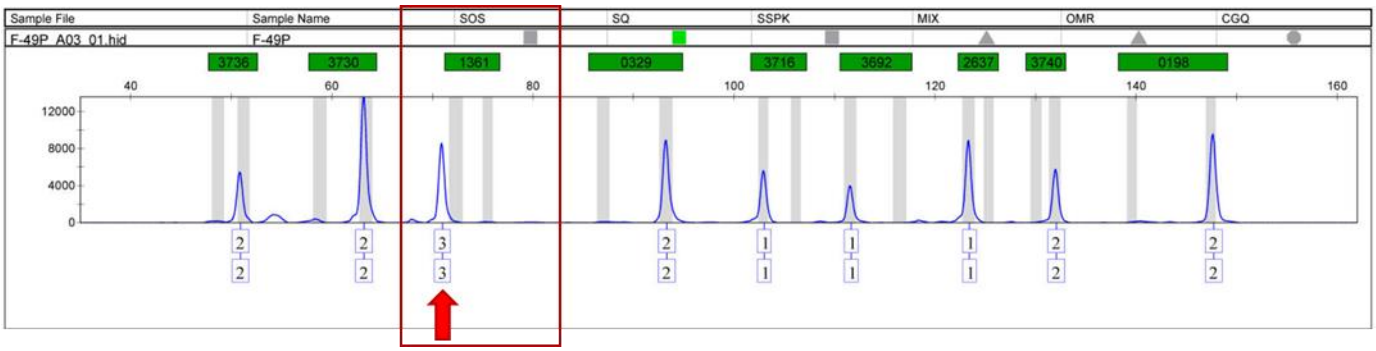
Fonte: Dados da pesquisa

Figura 8: Eletroferograma parcial de uma amostra do Espírito Santo obtida com o multiplex 32 X-InDels. Em destaque a variante alélica do alelo curto do marcador MID1361.



Fonte: Dados da pesquisa

Figura 9: Eletroferograma parcial de uma amostra do Rio de Janeiro obtida com o multiplex 32 X-InDels. Em destaque a variante alélica do alelo curto do marcador MID1361.



Fonte: Dados da pesquisa

A variante alélica (aqui chamado de alelo 3), foi observada pela primeira vez na população do estado de São Paulo (MARTINEZ *et al*, 2019), sendo relatada a presença desta mesma variante em casos de trio, onde é possível observar a transferência do alelo 3 do genitor para seu descendente, descartando a possibilidade de ser um artefato.

Neste trabalho não foi possível observar tal transferência, pois as amostras aqui estudas são de indivíduos sem relação genética pelo cromossomo X.

O MID1361 geralmente apresenta os InDels "-/ ACA", mas no banco de dados de SNPs (bdSNP) já consta registrado um padrão trialélico com a combinação de InDels "−/AA/ACA." O padrão encontrado em nossas amostras tem 1pb a menos do que o alelo curto esperado, representando, portanto, um padrão tetralélico para este marcador. O projeto 1000 genomas (The 1000 Genomes Project Consortium, 2015) descreve uma deleção de um nucleotídeo T (rs202070714) na posição 119614551, perto o locus MID1361 (posição 119614552). Na análise do marcador MID1361, esta deleção da base “T” associada com o alelo curto do marcador gera um *amplicon* com o mesmo tamanho em pares de base que foi observado em nossas análises, indicando sua presença. Martinez e colaboradores (2019) corroboram com esta associação na identificação da variante alélica observada em suas análises. Para a confirmação desta hipótese, foi realizado um sequenciamento do *amplicon* nas amostras que apresentaram tal variante. Entretanto, não foi possível obter uma amplificação eficiente desta região, que fosse favorável ao alinhamento. Isso provavelmente se deva ao tamanho pequeno dos primers escolhidos, o que limitou à amplificação das amostras. Em futuros trabalhos desta equipe de pesquisa, um novo sequenciamento poderá ser realizado para esclarecer aa causa deste *amplicon* nessas amostras.

Nos estudos realizados por Pereira e colaboradores (2012), foram analisadas amostras de europeus, africanos e asiáticos, não sendo observada a presença desta variante. A população brasileira é composta principalmente por contribuição populacional de europeus, africanos e ameríndios. Portanto, há indícios de que está variante tenha como herança populacional os descendentes nativos americanos, representando assim uma importante ferramenta em análise de ancestralidade.

5.3. Variabilidade genética populacional

Para a população de Minas de Gerais, as frequências alélicas para os 32 X-InDels foram calculadas separadamente para homens e mulheres, e não se observou distância genética significativa entre os sexos (F_{ST} ≤0,0000; p = 0,6497 ± 0,005). Por esse motivo, as frequências alélicas da população de Minas Gerais foram calculadas considerando a amostra total (tabela 2).

Tabela 2: Frequência alélica dos 32 marcadores X-InDels nas populações de Minas Gerais, Espírito Santo e Rio de Janeiro. Número de cromossomos analisados: 296 (Minas Gerais) 201 (Espírito Santo) e 220 (Rio de Janeiro).

	Minas Gerais			**Espírito Santo**			**Rio de Janeiro**		
Locus	**Alelo Curto**	**Alelo Longo**	**DG**	**Alelo Curto**	**Alelo Longo**	**DG**	**Alelo Curto**	**Alelo Longo**	**DG**
MID3736	0,6284	0,3716	0,4686	0,6070	0,3930	0,4795	0,5500	0,4500	0,4973
MID3730	0,2601	0,7399	0,4686	0,2687	0,7313	0,3949	0,2364	0,7636	0,3626
MID1361*	0,2331	0,7534	0,3792	0,2637	0,7313	0,3976	0,2500	0,7455	0,3835
MID329	0,2027	0,7973	0,3243	0,3184	0,6816	0,4362	0,1864	0,8136	0,3047
MID3716	0,5169	0,4831	0,5011	0,5124	0,4876	**0,5000**	0,5682	0,4318	0,4929
MID3692	0,2635	0,7365	0,3895	0,2139	0,7861	0,3380	0,3136	0,6864	0,4325
MID2637	0,8581	0,1419	0,2443	0,8707	0,1294	0,2264	0,8955	0,1046	0,1881
MID3740	0,3176	0,6824	0,4349	0,3731	0,6269	0,4702	0,3364	0,6636	0,4485
MID198	0,5608	0,4392	0,4943	0,5124	0,4876	**0,5000**	0,5500	0,4500	0,4973
MID3703	0,3345	0,6655	0,4467	0,2537	0,7463	0,3806	0,3182	0,6818	0,4359
MID3690	0,3514	0,6487	0,4574	0,3731	0,6269	0,4702	0,3091	0,6909	0,4291
MID3722	0,3480	0,6520	0,4553	0,3333	0,6667	0,4467	0,3409	0,6591	0,4514
MID3732	0,3041	0,6960	0,4246	0,3731	0,6269	0,4702	0,3273	0,6727	0,4423
MID3712	0,2939	0,7061	0,4165	0,3184	0,6816	0,4362	0,3455	0,6546	0,4543
MID1736	0,3074	0,6926	0,4273	0,3582	0,6418	0,4621	0,2682	0,7318	0,3943
MID3719	0,1960	0,8041	0,3162	0,2189	0,7811	0,3437	0,2727	0,7273	0,3985
MID2089	0,3176	0,6824	0,4349	0,3085	0,6915	0,4288	0,3364	0,6636	0,4485
MID3774	0,3311	0,6689	0,4444	0,3632	0,6368	0,4649	0,4273	0,5727	0,4917
MID3760	0,8514	0,1487	0,2540	0,7413	0,2587	0,3855	0,8182	0,1818	0,2989
MID3701	0,5439	0,4561	0,4978	0,4527	0,5473	0,4980	0,5046	0,4955	**0,5000**

MID2612	0,4966	0,5034	**0,5000**	0,3980	0,6020	0,4816	0,3955	0,6046	0,4803
MID1839	0,5439	0,4561	0,4978	0,5572	0,4428	0,4959	0,5864	0,4136	0,4873
MID3754	0,3074	0,6926	0,4273	0,3333	0,6667	0,4467	0,3727	0,6273	0,4697
MID111	0,5811	0,4189	0,4885	0,5522	0,4478	0,4970	0,5136	0,4864	0,5019
MID2652	0,6520	0,3480	0,4553	0,6766	0,3234	0,4398	0,5682	0,4318	0,4929
MID1511	0,5743	0,4257	0,4906	0,5821	0,4179	0,4890	0,4455	0,5546	0,4963
MID2692	0,4831	0,5169	0,5011	0,5174	0,4826	0,5019	0,4227	0,5773	0,4903
MID357	0,2939	0,7061	0,4165	0,3930	0,6070	0,4795	0,3091	0,6909	0,4291
MID356	0,3345	0,6655	0,4467	0,3582	0,6418	0,4621	0,3273	0,6727	0,4423
MID243	0,7804	0,2196	0,3439	0,7512	0,2488	0,3756	0,8046	0,1955	0,3159
MID3727	0,2601	0,7399	0,3862	0,4030	0,5970	0,4836	0,3000	0,7000	0,4219
MID3753	0,1182	0,8818	0,2092	0,2090	0,7910	0,3322	0,0909	0,9091	0,1660
MÉDIA			0,4176			0,4381			0,4234

Fonte: Dados da pesquisa; DG: diversidade gênica; * marcador possui frequência do alelo variante nas 3 populações estudadas: MG: 0,0135; ES: 0,0050; RJ:0,0046; em negrito os marcadores que apresentaram maior diversidade gênica.

Como esperado, todos os marcadores se mostraram informativos para as populações estudadas (tabela 2), sendo que os marcadores a apresentarem o maior valor de diversidade gênica aceitável para marcadores bialélicos (0,5000) foram os MID2612 e MID3701 respectivamente para as populações de Minas Gerais e Rio de Janeiro, e os MID3716 e MID198 para a população do Espírito Santo.

A população do Espírito Santo foi a que apresentou a maior média de diversidade gênica sendo de 0,438 (±0.064), seguida pela população do Rio de Janeiro 0,423 (±0.086) e a menor média observada foi a da população de Minas Gerais 0,418 (± 0.078). Comparada com outras populações estudadas por Pereira e colaboradores (2012), estes valores apresentados por estas três populações aqui estudadas foram maiores que os valores apresentados nas populações da Europa (0,388), Ásia (0,347) e de Angola e Moçambique (0,408), sendo inclusive maior que os valores encontrados por Martinez e colaboradores (2019) para a população de São Paulo (0,414). Tendo em conta a contribuição ancestral da população brasileira, esses valores já eram esperados pois, conforme demostrado por Martinez e colaboradores (2019), esses marcadores são altamente informativos em análise de populações miscigenadas.

Todos os marcadores analisados nas populações de Minas Gerais e Espírito Santo apresentaram-se com mais de 20% de diversidade gênica, sendo que os marcadores MID3753 (0,2092) em Minas Gerais e MID2637 (0,2264) no Espírito Santo foram os que apresentaram o menor valor de DG. Estes mesmos marcadores também apresentaram menor DG na população do Rio de Janeiro (MID3753 (0,1660) - MID2637 (0,1881)). Comparando estes dados com outras populações estudadas, observou-se que

o MID2637 apresentou menor DG nas populações de Portugal (0,174), Macau (0,187) e São Paulo (0,1880) (PEREIRA et. al, 2012; MARTINEZ et. al, 2019). Já o MID3753 apresentou resultado similar em amostras da população de Angola e Moçambique (0,187) (PEREIRA et. al, 2012).

Conforme proposto por Szibor e colaboradores (2003), é recomendável que os marcadores usados em ciência forense se apresentem em equilíbrio de Hardy-Weinberg, pois só assim poderá ser calculada a frequência do genótipo. Este teste, se caracteriza por ser uma ferramenta importante para detectar erros de genotipagem em estudos genéticos de larga escala. Pressupõe-se que uma população em equilíbrio de Hardy-Weinberg é infinita, não está sujeita a eventos de mutação, migração e seleção e que os cruzamentos ocorreram de forma aleatória. Assim, com base nas frequências de seus alelos, as proporções genotípicas permanecem inalteradas ao longo das gerações e atingem um equilíbrio estável (ANDREWS, 2010).

Os dados gerados das amostras femininas de Minas Gerais foram utilizados no teste exato para o equilíbrio de Hardy-Weinberg. Utilizando um nível de significância (*p*) de 0,00156, obtido através da correção de Bonferroni (0,05/32) para testes múltiplos, não foram detectados desvios significativos para os genótipos femininos da população de Minas Gerais. O menor valor observado foi o do MID1361 (0,0091). No estudo realizado na população de São Paulo por Martinez e colaboradores (2019), este marcador apresentou um desvio significativo (0,0014). Não há na literatura alelo nulo descrito que justifique o desequilíbrio relatado por Martinez e colaboradores (2019).

A maioria dos estudos realizados demonstra uma distribuição genotípica de acordo com o equilíbrio de Hardy-Weinberg, o que também foi observado neste estudo (tabela 3) onde os valores não foram significativos. Diante disso, as frequências genotípicas encontradas na amostragem podem ser consideradas representativas para as populações estudadas neste trabalho.

Tabela 3: Parâmetros estatísticos da diversidade gênica para os 32 marcadores X-InDels na população feminina de Minas Gerais. Tamanho da amostra: 90.

Locus	**Ho**	**He**	**P-HWE**
MID3736	0,5111	0,4728	0,5040
MID3730	0,3778	0,3933	0,7868
MID1361	0,4111	0,4145	0,0091
MID329	0,3667	0,3284	0,3443
MID3716	0,5000	0,5012	1,0000
MID3692	0,3222	0,3880	0,1677

MID2637	0,2000	0,2158	0,6128
MID3740	0,4444	0,4541	1,0000
MID198	0,4111	0,4888	0,1364
MID3703	0,4222	0,4469	0,6382
MID3690	0,3778	0,4728	0,0722
MID3722	0,4889	0,4541	0,4938
MID3732	0,3333	0,4310	0,0480
MID3712	0,3667	0,3880	0,5922
MID1736	0,4000	0,4132	0,7991
MID3719	0,3111	0,3081	1,0000
MID2089	0,4778	0,4640	0,8217
MID3774	0,4556	0,4506	1,0000
MID3760	0,2000	0,2324	0,1796
MID3701	0,4000	0,4966	0,0875
MID2612	0,5222	0,4998	0,6764
MID1839	0,5778	0,4988	0,1434
MID3754	0,4333	0,4084	0,6125
MID111	0,5889	0,4849	0,0493
MID2652	0,3889	0,4506	0,2397
MID1511	0,4778	0,4978	0,8322
MID2692	0,4889	0,5018	0,8358
MID357	0,3667	0,4268	0,2150
MID356	0,4000	0,4541	0,3501
MID243	0,4000	0,3826	0,7842
MID3727	0,3889	0,3880	1,0000
MID3753	0,2222	0,2158	1,0000

Fonte: Dados da pesquisa; Ho, heterozigose observada; He, heterozigose esperada; p-HWE, valor de p para o equilíbrio de Hardy-Weinberg, com nível de significância de 0,00156 (após correção de Bonferroni para múltiplos testes).

5.4. Desequilíbrio de ligação entre os marcadores X-InDels

Quando estudamos marcadores que estão localizados próximos no mesmo cromossomo, existe uma maior probabilidade de existir desequilíbrio de ligação (LD) entre eles. Sendo assim, para marcadores X-InDels é necessário avaliar a presença de LD para sua posterior utilização na prática forense.

Ligação ou "*linkage*" está associada à dependência física entre loci, ou seja, refere-se à co-segregação de loci que estão localizados próximos. É a tendência de dois ou mais loci serem herdados juntos. Já LD consiste na dependência de alelos em diferentes loci, não necessariamente no mesmo cromossomo, e é definido como uma associação não aleatória de alelos na população. Quando os marcadores em análise estão em LD, não se podem inferir frequências haplotipicas com base nas frequências

alélicas, sendo que as frequências haplotipicas observadas devem ser calculadas diretamente na população (TILLMAR *et al*., 2011).

Utilizando apenas os dados das amostras masculinas, o teste exato para avaliação de LD foi realizado para todos os marcadores nas três populações estudadas. A análise resultou em 496 pareamentos/população, que após a correção de Bonferroni para múltiplas análises ($p \leq 0,05/496=0,00010$), observou-se associações significativas nas três populações. Nas populações de Minas Gerais e Rio de Janeiro, foram observados associação em 3 pares de marcadores (tabela 4 e 5) sendo eles o par MID356-MID357, MID3719-MID2089, MID3690-MID3719.

Tabela 4: Desequilíbrio de ligação dos pares de marcadores analisados na população masculina de Minas Gerais. Tamanho da amostra: 116

Par de InDels	Distância em cM	p-LD
MID3719-MID2089	**0,08**	**0,0000**
MID357-MID356	**0,01**	**0,0000**
MID3690-MID3719	**0,06**	**0,0001**
MID3732-MID1736	17,70	0,0010
MID2612-MID243	115,17	0,0017
MID1511-MID3727	6,48	0,0052
MID3703-MID3774	0,27	0,0062
MID3716-MID2089	2,73	0,0087
MID1736-MID3753	53,36	0,0091
MID3736-MID3730	44,69	0,0147
MID2637-MID3760	10,01	0,0201
MID3760-MID243	13,13	0,0233
MID3690-MID1839	0,40	0,0292
MID3703-MID1839	122,35	0,0324
MID2637-MID356	115,84	0,0332
MID2652-MID2692	128,86	0,0351
MID111-MID243	6,22	0,0402
MID2637-MID3727	29,10	0,0458
MID3760-MID3753	4,74	0,0473
MID3692-MID1511	70,40	0,0483

Fonte: Dados da pesquisa; p-LD, valor de p para o teste de desequilíbrio de ligação; em negrito amostras que apresentaram nível de significância abaixo de 0,00010 (após correção de Bonferroni para múltiplos testes); distância em centimorgans (cM) entre os marcadores.

Valores significativos de *p* entre estes 3 pares de marcadores foram relatados em outras populações estudadas. A associação do par MID3690-MID3719 já foi observada nas populações da Angola, Moçambique (PEREIRA et. Al, 2012), Boiacá-

Cundinamarca, Huila, Santander, Norte de Santander (IBARRA et. al, 2014), Alemanha e países Bálticos (EDELMANN *et al*., 2016).

O desequilíbrio do par MID3719-MID2089 foi observado nas populações de Portugal, Somália (PEREIRA et. Al, 2012), Maiorca (FERRAGUT et. al, 2017), Boiacá-Cundinamarca, Huila, Santander, Norte de Santander, Arauca, Casanare, Meta, Chocó (IBARRA et. al, 2014), Iraque (PEREIRA *et al*., 2011), Alemanha, países Bálticos (EDELMANN *et al*., 2016), São Paulo (MARTINEZ *et al*., 2019). Neste trabalho, estes pares de marcadores também aparecem em desequilíbrio na população do Rio de Janeiro (tabela 5).

Tabela 5: Desequilíbrio de ligação dos pares de marcadores analisados na população do Rio de Janeiro. Tamanho da amostra: 220

Par de InDels	**Distância em cM**	***p*-LD**
MID3690-MID3719	**0,06**	**0,0000**
MID3719-MID2089	**0,08**	**0,0000**
MID357-MID356	**0,01**	**0,0000**
MID1736-MID2652	100,86	0,0004
MID1839-MID111	6,91	0,0007
MID3690-MID2692	16,93	0,0020
MID3692-MID3712	10,94	0,0048
MID3754-MID111	15,93	0,0052
MID1511-MID243	32,07	0,0062
MID1511-MID3727	6,48	0,0090
MID3716-MID3754	82,37	0,0099
MID2637-MID1511	35,58	0,0150
MID2692-MID3753	81,35	0,0152
MID3716-MID2692	19,52	0,0195
MID3692-MID1736	57,90	0,0203
MID3730-MID329	78,23	0,0225
MID2612-MID1839	128,30	0,0232
MID1839-MID1511	45,20	0,0266
MID1736-MID3753	53,36	0,0320
MID3712-MID243	113,41	0,0346

Fonte: Dados da pesquisa; p-LD, valor de p para o teste de desequilíbrio de ligação; em negrito amostras que apresentaram nível de significância abaixo de 0,0001 (após correção de Bonferroni para múltiplos testes); distância em centimorgans (cM) entre os marcadores.

A associação dos marcadores MID357-MID356 foi observada também na população do Espírito Santo (tabela 6). Outra associação observada nesta população ocorreu entre os marcadores MID3719-MID2089. No estudo realizado por Pereira et. al, 2012, os autores destacam a distância entre os marcadores 3719-2089-3690

(~0,14cM), ressaltando que estes marcadores formam um *cluster*. Por esse motivo, Pereira e colaboradores (2012) sugerem que esses três marcadores devem ser analisados como um único bloco haplotípico.

Tabela 6: Desequilíbrio de ligação dos pares de marcadores analisados na população do Espírito Santo. Tamanho da amostra: 201

Par de InDels	Distância em cM	*p*-LD
MID3719-MID2089	**0,08**	**0,0000**
MID357-MID356	**0,01**	**0,0000**
MID3760-MID3753	4,74	0,0004
MID3732-MID1839	39,99	0,0005
MID198-MID3732	28,38	0,0006
MID3732-MID1736	17,70	0,0008
MID3703-MID3774	0,27	0,0008
MID198-MID1839	68,37	0,0009
MID329-MID3760	28,36	0,0009
MID329-MID3732	68,75	0,0010
MID3703-MID3722	100,16	0,0011
MID329-MID1736	86,45	0,0016
MID198-MID3760	68,76	0,0020
MID3732-MID3760	40,39	0,0023
MID3760-MID357	125,87	0,0023
MID329-MID3753	33,09	0,0059
MID1736-MID3753	53,36	0,0066
MID3727-MID3753	34,38	0,0072
MID3732-MID3753	35,65	0,0073
MID3732-MID111	33,08	0,0077

Fonte: Dados da pesquisa; p-LD, valor de p para o teste de desequilíbrio de ligação; em negrito amostras que apresentaram nível de significância abaixo de 0,0001 (após correção de Bonferroni para múltiplos testes); distância em centimorgans (cM) entre os marcadores.

Como já era esperado, os marcadores MID356-MID357 se mostraram em desequilíbrio em ambas as populações estudadas. Estes dados estão de acordo com os resultados já observado em outras populações estudadas, sendo em descendentes de judeus (FERRAGUT *et al*., 2017), Somália e Iraque (PEREIRA *et al*., 2011), Argentina (CAPUTO *et al*., 2017), Colômbia (IBARRA *et al*., 2014), Alemanha e países Bálticos (EDELMANN *et al*., 2016), Portugal e Miranda do Douro, Espanha (Zamora e Maiorca), Angola, Moçambique, Macau (PEREIRA *et al*., 2012; PINTO *et al*., 2015; FERRAGUT *et al*., 2017) e no Brasil, nas populações de Belém (FREITAS *et al*., 2010) e São Paulo (MARTINEZ *et al*., 2019).

Como é possível observar na tabela 4, 5 e 6, estes marcadores (MID357-MID356) apresentam-se muito próximos no cromossomo X (~0,01cM de distância) e, assim como o *cluster* 3719-2089-3690, devem ser tratados como haplótipos quando forem utilizados juntos.

Tabela 7: Frequência haplotípica observada na população masculina de Minas Gerais para os marcadores em desequilíbrio de ligação. Tamanho da amostra: 116.

BLOCO I		
357	**356**	**Frequência**
1	1	0,2759
1	2	0,0000
2	1	0,0431
2	2	0,6810

BLOCO II			
3690	**3719**	**2089**	**Frequência**
1	1	1	0,0000
1	1	2	0,0000
1	2	1	0,0517
1	2	2	0,2586
2	1	1	0,1552
2	1	2	0,0431
2	2	1	0,4397
2	2	2	0,0517

Fonte: Dados da pesquisa

Tabela 8: Frequência haplotípica observada na população masculina de Espírito Santo para os marcadores em desequilíbrio de ligação. Tamanho da amostra: 201.

BLOCO I		
357	**356**	**Frequência**
1	1	0,2985
1	2	0,0945
2	1	0,0597
2	2	0,5473

BLOCO II			
3719	**2089**	**3690**	**Frequência**
1	1	1	0,0199
1	1	2	0,1095
1	2	1	0,0299
1	2	2	0,0597

2	1	1	0,1244
2	1	2	0,0547
2	2	1	0,1990
2	2	2	0,4030

Fonte: Dados da pesquisa

Tabela 9: Frequência haplotípica observada na população masculina de Rio de Janeiro para os marcadores em desequilíbrio de ligação. Tamanho da amostra: 220.

BLOCO I		
357	**356**	**Frequência**
1	1	0,1818
1	2	0,0045
2	1	0,0227
2	2	0,6682

BLOCO II			
3690	**3719**	**2089**	**Frequência**
1	1	1	0,0000
1	1	2	0,0227
1	2	1	0,1091
1	2	2	0,1773
2	1	1	0,1818
2	1	2	0,0682
2	2	1	0,0455
2	2	2	0,3955

Fonte: Dados da pesquisa

5.5. Parâmetros estatísticos de eficiência forense

Em uma resolução de caso forense, antes de identificar a representatividade de uma evidência de DNA, deve-se considerar a eficiência dos marcadores selecionados para determinada população (TILLMAR, 2010). Para testar a eficiência dos marcadores utilizados neste trabalho para as populações de Minas Gerais, Espírito Santo e Rio de Janeiro, uma série de parâmetros estatísticos foram calculados utilizando uma ferramenta *online* disponível em http://www.chrx-str.org.

Testes de paternidade consistem nas análises mais comuns na rotina da genética forense. A eficiência de um marcador nesses testes é mensurada por meio de sua habilidade de excluir um homem aleatório na população como pai da criança, sabendo que este não é o pai da criança. Essa probabilidade é conhecida como “poder de

exclusão” (PE, ou chance média de exclusão, MEC) e pode ser calculada quando o genótipo da mãe está ou não disponível (PINTO *et al*., 2013b).

A chance média de exclusão (MEC) foi introduzida por Krüger *et al*. (1968), para tipos de marcadores autossômicos em trios envolvendo mãe, filho e suposto pai. Este parâmetro não é adequado para marcadores do cromossomo X, exceto para casos de deficiência em que a avó paterna é investigada em vez do suposto pai (MEC_{PGM}) (SZIBOR *et al*., 2003).

Em 1998, Desmarais e colaboradores introduziram uma possibilidade de cálculo de MEC para trios (MEC_T) e duos (MEC_D), bem como do poder de discriminação em indivíduos do sexo masculino e do sexo feminino para esses marcadores.

O indicativo de parâmetros estatísticos de eficiência forense para os 32 X-InDel foi calculado usando as informações de *locus* único (tabela 2), exceto aqueles marcadores apresentando a associação, que foram tratados como haplótipos (usando as frequências das tabelas 7, 8 e 9)

Diante de uma situação de paternidade de trio (mãe, filha e suposto pai), o valor combinado de MEC_T foi de 0,999996, para as populações de Minas Gerais e Espírito Santo e de 0,999997 para população do Rio de Janeiro. Para situações envolvendo apenas suposto pai e filha (MEC_D) o valor combinado foi 0,9994 para a população de Minas Gerais, no Espírito Santo 0,9998 e 0,9996 na população do Rio de Janeiro. Nos casos em que o pai está ausente e os genótipos disponíveis são da mãe, filha e suposta avó paterna, o valor combinado de MEC_{PGM} foi 0,997 nas três populações estudadas (tabela 10, 11, 12).

Sabendo que o MEC_D considera a ausência da mãe, não é possível determinar quais alelos essa doou à filha, portanto, em situações em que a filha é heterozigota com perfil alélico AB, seu pai poderá ser qualquer homem da população, uma vez que os marcadores InDels são bialélicos, o que por consequência aumentam-se os pais possíveis e diminui-se o MEC. Nessas circunstâncias, os InDels são marcadores informativos somente nos casos em que a filha é homozigota.

Tabela 10: Parâmetros estatísticos de interesse forense para os marcadores 32 X-InDels na população de Minas Gerais. Número de cromossomos analisados: 296.

Locus	**PD_F**	**PD_M**	**MEC_T**	**MEC_D**	**MEC_{PGM}**
bloco1	0,6339	0,4583	0,3857	0,2534	0,2116
bloco2	0,6906	0,4766	0,4412	0,2985	0,2712
MID3727	0,5476	0,3849	0,3108	0,1925	0,1554
MID3753	0,3518	0,2085	0,1868	0,1043	0,0934

MID3736	0,6069	0,4670	0,3580	0,2335	0,1790
MID3730	0,5476	0,3849	0,3108	0,1925	0,1554
MID1361	0,5511	0,3779	0,3160	0,1961	0,1638
MID329	0,4897	0,3232	0,2710	0,1616	0,1355
MID3716	0,6247	0,4994	0,3747	0,2497	0,1874
MID3692	0,5503	0,3881	0,3128	0,1941	0,1564
MID2637	0,3981	0,2435	0,2139	0,1218	0,1069
MID3740	0,5851	0,4334	0,3395	0,2167	0,1698
MID198	0,6212	0,4926	0,3713	0,2463	0,1856
MID3703	0,5931	0,4452	0,3461	0,2226	0,1730
MID3690	0,6000	0,4558	0,3519	0,2279	0,1760
MID3722	0,5987	0,4538	0,3508	0,2269	0,1754
MID3732	0,5778	0,4232	0,3337	0,2116	0,1668
MID3712	0,5717	0,4151	0,3289	0,2075	0,1645
MID1736	0,5797	0,4258	0,3352	0,2129	0,1676
MID3774	0,5916	0,4429	0,3448	0,2215	0,1724
MID3760	0,4101	0,2531	0,2211	0,1266	0,1105
MID3701	0,6230	0,4961	0,3731	0,2481	0,1865
MID2612	0,6250	0,5000	0,3750	0,2500	0,1875
MID1839	0,6230	0,4961	0,3731	0,2481	0,1865
MID3754	0,5797	0,4258	0,3352	0,2129	0,1676
MID111	0,6182	0,4869	0,3683	0,2434	0,1842
MID2652	0,5987	0,4538	0,3508	0,2269	0,1754
MID1511	0,6193	0,4890	0,3694	0,2445	0,1847
MID2692	0,6247	0,4994	0,3747	0,2497	0,1874
MID243	0,5093	0,3427	0,2840	0,1714	0,1420
TOTAL	**0,999999999994**	**0,99999994**	**0,999996**	**0,9994**	**0,997**

Fonte: Dados da pesquisa; PDF, poder de discriminação feminino; PDM, poder de discriminação masculino; MECT, chance média de exclusão em trios envolvendo filha, mãe e suposto pai; MECD, chance média de exclusão em duo pai/filha; MEC_{PGM}, chance média de exclusão em casos deficientes envolvendo mãe, filha e suposta avó paterna; Bloco I: MID357-MID356; Bloco II: MID3690-MID3719-MID2089.

O Poder de Discriminação (PD) consiste na probabilidade de dois indivíduos escolhidos ao acaso na população apresentarem genótipos diferentes. Para as populações aqui estudadas, o valor combinado para mulheres (PD_F) foi de 0,999999999994 nas populações de Minas Gerais e Espírito Santo, e de 0,999999999997 na população do Rio de Janeiro. O valor combinado para os homens (PD_M) foi de 0,99999994 na população de Minas Gerais, na população do Espírito Santo 0,99999995 e 0,99999997 na população do Rio de Janeiro (tabelas 10, 11 e 12)

Tabela 11: Parâmetros estatísticos de interesse forense para os marcadores 32 X-InDels na população do Espírito Santo. Número de cromossomos analisados: 201

Locus	PD_F	PD_M	MEC_T	MEC_D	MEC_{PGM}
bloco1	0,7759	0,5989	0,5357	0,3893	0,3391
bloco2	0,9157	0,7627	0,7347	0,6024	0,5683
MID3727	0,6151	0,4812	0,3654	0,2406	0,1827
MID3753	0,4972	0,3306	0,2759	0,1653	0,1380
MID3736	0,6128	0,4771	0,3633	0,2386	0,1816
MID3730	0,5543	0,3930	0,3158	0,1965	0,1579
MID1361	0,5603	0,3956	0,3212	0,2007	0,1629
MID329	0,5855	0,4341	0,3399	0,2170	0,1699
MID3716	0,6248	0,4997	0,3748	0,2498	0,1874
MID3692	0,5030	0,3363	0,2798	0,1682	0,1399
MID2637	0,3744	0,2252	0,1999	0,1126	0,0999
MID3740	0,6074	0,4678	0,3584	0,2339	0,1792
MID198	0,6248	0,4997	0,3748	0,2498	0,1874
MID3703	0,5423	0,3787	0,3070	0,1894	0,1535
MID3722	0,5926	0,4444	0,3457	0,2222	0,1728
MID3732	0,6074	0,4678	0,3584	0,2339	0,1792
MID3712	0,5855	0,4341	0,3399	0,2170	0,1699
MID1736	0,6025	0,4598	0,3541	0,2299	0,1770
MID3774	0,6042	0,4626	0,3556	0,2313	0,1778
MID3760	0,5464	0,3836	0,3100	0,1918	0,1550
MID3701	0,6227	0,4955	0,3728	0,2478	0,1864
MID2612	0,6139	0,4792	0,3644	0,2396	0,1822
MID1839	0,6217	0,4935	0,3717	0,2467	0,1859
MID3754	0,5926	0,4444	0,3457	0,2222	0,1728
MID111	0,6222	0,4945	0,3723	0,2473	0,1861
MID2652	0,5880	0,4376	0,3419	0,2188	0,1709
MID1511	0,6180	0,4865	0,3682	0,2433	0,1841
MID2692	0,6247	0,4994	0,3747	0,2497	0,1873
MID243	0,5380	0,3738	0,3039	0,1869	0,1520
Total	**0,999999999994**	**0,99999995**	**0,999996**	**0,9995**	**0,997**

Fonte : Dados da pesquisa; PDF, poder de discriminação feminino; PDM, poder de discriminação masculino; MECT, chance média de exclusão em trios envolvendo filha, mãe e suposto pai; MECD, chance média de exclusão em duo pai/filha; MEC_{PGM}, chance média de exclusão em casos deficientes envolvendo mãe, filha e suposta avó paterna; Bloco I: MID357-MID356; Bloco II: MID3690-MID3719-MID2089.

Tabela 12: Parâmetros estatísticos de interesse forense para os marcadores 32 X-InDels na população do Rio de Janeiro. Número de cromossomos analisados: 220

Locus	PD_F	PD_M	MEC_T	MEC_D	MEC_{PGM}
bloco1	0,7395	0,5199	0,4899	0,3442	0,1907
bloco2	0,9115	0,7600	0,7291	0,5951	0,5572
MID3727	0,5754	0,4200	0,3318	0,2100	0,1659
MID3753	0,2896	0,1653	0,1516	0,0826	0,0758
MID3736	0,6225	0,4950	0,3725	0,2475	0,1862
MID3730	0,5265	0,3610	0,2958	0,1805	0,1479
MID1361	0,5483	0,3818	0,3123	0,1934	0,1582
MID329	0,4686	0,3033	0,2573	0,1516	0,1286
MID3716	0,6202	0,4907	0,3703	0,2454	0,1852
MID3692	0,5830	0,4305	0,3379	0,2153	0,1689
MID2637	0,3219	0,1872	0,1697	0,0936	0,0849
MID3740	0,5939	0,4464	0,3468	0,2232	0,1734
MID198	0,6225	0,4950	0,3725	0,2475	0,1862
MID3703	0,5854	0,4339	0,3398	0,2169	0,1699
MID3722	0,5958	0,4494	0,3484	0,2247	0,1742
MID3732	0,5898	0,4403	0,3434	0,2202	0,1717
MID3712	0,5977	0,4522	0,3500	0,2261	0,1750
MID1736	0,5539	0,3925	0,3155	0,1963	0,1577
MID3774	0,6195	0,4894	0,3697	0,2447	0,1848
MID3760	0,4623	0,2975	0,2533	0,1488	0,1266
MID3701	0,6250	0,5000	0,3750	0,2500	0,1875
MID2612	0,6134	0,4781	0,3638	0,2391	0,1819
MID1839	0,6172	0,4851	0,3674	0,2425	0,1837
MID3754	0,6072	0,4676	0,3583	0,2338	0,1791
MID111	0,6248	0,4996	0,3748	0,2498	0,1874
MID2652	0,6202	0,4907	0,3703	0,2454	0,1852
MID1511	0,6220	0,4940	0,3720	0,2470	0,1860
MID2692	0,6188	0,4881	0,3690	0,2440	0,1845
MID243	0,4806	0,3145	0,2650	0,1572	0,1325
Total	**0,999999999997**	**0,99999997**	**0,999997**	**0,9996**	**0,997**

Fonte: Dados da pesquisa; PDF, poder de discriminação feminino; PDM, poder de discriminação masculino; MECT, chance média de exclusão em trios envolvendo filha, mãe e suposto pai; MECD, chance média de exclusão em duo pai/filha; MEC_{PGM}, chance média de exclusão em casos deficientes envolvendo mãe, filha e suposta avó paterna; Bloco I: MID357-MID356; Bloco II: MID3690-MID3719-MID2089.

Figura 10: Parâmetros estatísticos de interesse forense para os marcadores 32 X-InDels, para a população da Somália, Angola e Moçambique, Portugal e Macau.

	Somalia	Ang+Moz	Portugal	Macau
PD_{Males}	0.99999998	0.99999995	0.9999997	0.9999990
$PD_{Females}$	0.999999999998	0.999999999993	0.99999999997	0.9999999995
MEC_{Duos}	0.9996	0.9995	0.9990	0.998
MEC_{Trios}	0.999998	0.999996	0.999990	0.99997

Fonte: Pereira, et. al., 2012

Comparando com dados de outras populações (Figura 9), os 32 X-InDels se mostraram tão eficientes em Minas Gerais e Espírito Santo quanto foram para São Paulo (MARTINEZ, et. al., 2019) e Somália (PEREIRA *et al*., 2012), sendo mais eficientes do que para Angola e Moçambique (PD_F), Portugal e Macau (PEREIRA *et al*., 2012). Para a população do Rio de Janeiro, esses marcadores se mostraram mais eficientes do que para São Paulo, Somália, Angola e Moçambique, Portugal, Macau (MARTINEZ, et. al., 2019; PEREIRA *et al*., 2012), Minas Gerais e Espírito Santo. Em síntese, não há relatos de grupos de marcadores bialélicos localizados no cromossomo X com poder de discriminação maior que os encontrados nesse trabalho, mesmo para aqueles que analisam mais do que 32 polimorfismos (FREITAS *et al*., 2010; CAPUTO *et al*., 2017; EDELMANN, *et al*., 2016). Esses valores já eram esperados, dada a maior diversidade relatada para essas populações em comparação com as anteriormente citadas. Além disso, já fora demostrado por Martinez e colaboradores (2019), que estes marcadores são mais discriminativos para populações miscigenadas, tal como a população brasileira.

5.6. Comparação genética entre as populações e análise de ancestralidade

A população do Brasil é constituída por uma vasta variabilidade genética, que foi moldada ao longo dos mais de 500 anos de cruzamentos interétnicos entre nativos americanos, colonizadores europeus e escravos africanos. Com os dados obtidos neste estudo, foi realizada a análise da distância genética (F_{ST}) entre as populações dos estados de Minas Gerais, Espírito Santo e Rio de Janeiro. Além disso, elas foram comparadas com as populações do estado de São Paulo, Mato Grasso, Brasília e

populações estrangeiras com dados disponíveis na literatura que analisaram o mesmo conjunto de marcadores X-InDels (tabela 13).

Tabela 13: Amostras usadas na comparação entre as populações, número de cromossomos analisados (N) e respectivas referências.

População	N	Referência
EUROPA		
Portugal	324	Pereira *et al*. (2012)
Espanha (Maiorca)	112	Ferragut *et al*. (2017)
ÁFRICA		
Angola e Moçambique	116	Pereira *et al*. (2012)
ÁMÉRICA LATINA		
Brasil		
Minas Gerais	206	
Espírito Santo	201	Este trabalho
Rio de Janeiro	220	
Mato Grosso	406	Polverari (2018)
Brasília	346	
São Paulo	782	Martinez *et al*. (2019)
Colômbia		
Antioquia	145	
Boiacá-Cundinamarca	144	
Huila	137	
Santander	183	
Norte de Santander	36	
Arauca	154	Ibarra *et al*. (2014)
Casanare	22	
Meta	55	
Cartagena	194	
Chocó	104	
Narinho	84	
NATIVOAMERICANO		
Pasto (Colômbia)	83	Ibarra *et al*. (2014)
LESTE ASIÁTICO		
Macau	100	Pereira *et al*. (2012)
Taiwan	20	Pereira *et al*. (2015)

A análise de distância genética (F_{ST}) (tabela 14) revelou valores significativos na maioria das comparações entre as populações brasileiras com as populações estrangeiras. A fim de melhorar a visualização dos resultados, com os valores de distância genética (F_{ST}) foi realizada uma análise em escalonamento multidimensional (MDS) em 2 dimensões (Figura 10). A adição de uma terceira dimensão ao gráfico MDS não contribui para uma melhoria significativa na visualização dos resultados.

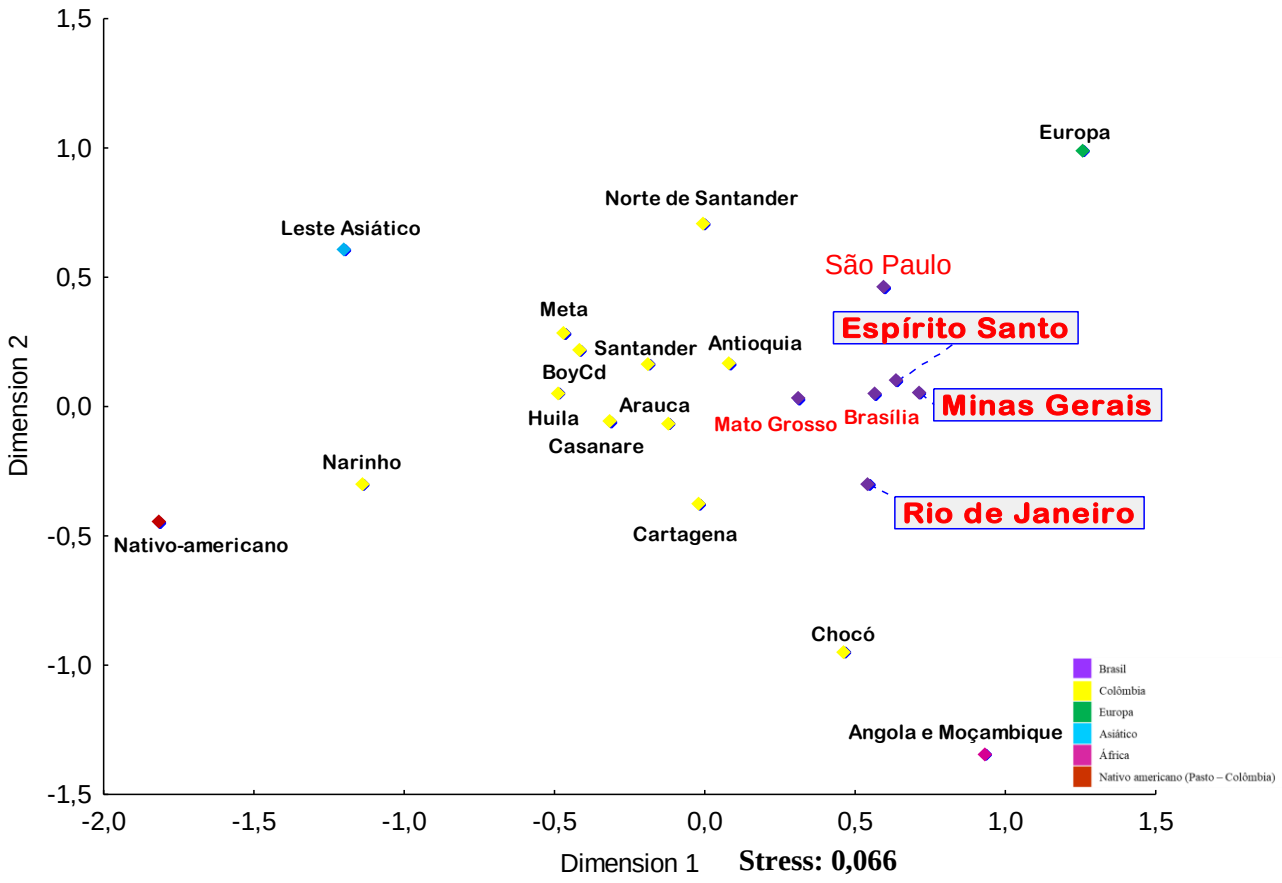
Tabela 14: Distância genética (F_{ST}) entre as populações (abaixo da diagonal) e o valor de p correspondente (acima da diagonal)

	Antioquia	CunBoy	Huila	Santander	N. de Santander	Arauca	Casanare	Meta	Cartagena	Choco	Nariño	Nativo	Ang e Moç	Europa	Leste Asiático	São Paulo	Brasília	Mato Grosso	Minas Gerais	Espírito Santo	Rio de Janeiro
Antioquia	*	0.03237	0.00723	0.43649	0.02584	0.02930	0.32442	0.04010	0.00040	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00059	0.00713	0.09088	0.00020	0.00000	0.00000
CunBoy	0.00500	*	0.16256	0.43075	0.03911	0.01693	0.52589	0.57588	0.00010	0.00000	0.00099	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00010	0.00000	0.00000	0.00000
Huila	0.00813	0.00214	*	0.14949	0.03564	0.00515	0.32036	0.29086	0.00119	0.00000	0.01723	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00010	0.00000	0.00000	0.00000
Santander	-0.00007	-0.00001	0.00203	*	0.04505	0.05732	0.50450	0.28799	0.00129	0.00000	0.00020	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00366	0.00000	0.00000	0.00000
N. de Santander	0.01364	0.01240	0.01254	0.01075	*	0.01990	0.14850	0.03851	0.00723	0.00000	0.00050	0.00000	0.00000	0.00000	0.00000	0.04247	0.02000	0.01544	0.00495	0.00000	0.00000
Arauca	0.00507	0.00659	0.00889	0.00357	0.01593	*	0.77586	0.03683	0.00158	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00198	0.00000	0.00000	0.00000
Casanare	0.00242	-0.00187	0.00248	-0.00128	0.01168	-0.0067	*	0.14622	0.11019	0.00149	0.19553	0.00198	0.00000	0.00000	0.00000	0.00782	0.04059	0.12702	0.01594	0.00238	0.00198
Meta	0.00840	-0.00138	0.00159	0.00138	0.01602	0.00857	0.01025	*	0.01208	0.00000	0.00505	0.00000	0.00000	0.00000	0.00000	0.00010	0.00079	0.00386	0.00000	0.00000	0.00000
Cartagena	0.01382	0.01315	0.01085	0.01008	0.02080	0.01135	0.01050	0.01216	*	0.00564	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00020	0.00099	0.00000	0.00000	0.00149
Choco	0.04039	0.04879	0.04421	0.04287	0.05390	0.03742	0.04260	0.04511	0.00938	*	0.00000	0.00000	0.01079	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00079
Nariño	0.03462	0.01768	0.00924	0.02022	0.03962	0.02676	0.00696	0.01923	0.02446	0.06046	*	0.09118	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
Nativo	0.07367	0.05032	0.03932	0.05677	0.08534	0.05929	0.04783	0.04810	0.06136	0.10701	0.00582	*	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
Ang e Moç	0.06401	0.08384	0.07662	0.07141	0.08453	0.06193	0.06523	0.08134	0.03066	0.00708	0.09699	0.14329	*	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
Europa	0.03810	0.06017	0.06726	0.04848	0.03562	0.06084	0.07306	0.06292	0.06273	0.08426	0.12450	0.18085	0.10351	*	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
Leste Asiático	0.09289	0.06734	0.05944	0.07977	0.06784	0.08733	0.09191	0.06756	0.07336	0.11430	0.07090	0.09369	0.16089	0.13489	*	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
São Paulo	0.00924	0.02115	0.02373	0.01278	0.00957	0.01970	0.02642	0.02404	0.02000	0.04112	0.06094	0.10548	0.06200	0.01315	0.08578	*	0.00050	0.00000	0.00079	0.00020	0.00000
Brasília	0.00643	0.02272	0.02249	0.01246	0.01604	0.01567	0.01797	0.02271	0.01181	0.02475	0.05444	0.09789	0.03919	0.02448	0.10186	0.00455	*	0.03138	0.23909	0.00653	0.03614
Mato Grosso	0.00197	0.01197	0.01316	0.00520	0.01232	0.00666	0.00800	0.01216	0.00708	0.02515	0.03923	0.07850	0.04477	0.03369	0.08719	0.00517	0.00199	*	0.00238	0.00000	0.00000
Minas Gerais	0.01043	0.02746	0.02897	0.01737	0.0183	0.01987	0.02072	0.03029	0.01582	0.02636	0.06392	0.11168	0.04187	0.02112	0.10536	0.00425	0.00056	0.00372	*	0.00455	0.01515
Espírito Santo	0.01404	0.02761	0.02738	0.01804	0.02306	0.01911	0.02488	0.02805	0.01571	0.02916	0.06201	0.10622	0.04345	0.02596	0.09755	0.00698	0.00454	0.0061	0.0041	*	0.00020
Rio de Janeiro	0.01485	0.0268	0.02632	0.01843	0.01963	0.01818	0.02263	0.02531	0.00574	0.00876	0.05473	0.09853	0.022	0.03648	0.09527	0.01016	0.00238	0.00613	0.00276	0.00628	*

Fonte: Dados da pesquisa; *valores de p após correção de Bonferroni (0,0002)

Como já era esperado, a maior similaridade dos perfis de X-InDels foi observada entre as populações brasileira. Comparando com as populações estrangeiras, as populações de Minas Gerais, Espírito Santo e Rio de Janeiro foram mais similares as populações da Europa e da Colômbia (caracterizadas por possuir maior componente europeu). Os valores de distância obtidos para estas populações (tabela 14 e figura 11), com relação à Europa, se devem ao fato de estas populações tem pouco europeu por via materna. Assim como, a população do Rio de Janeiro, tem uma maior contribuição por via materna em função dos escravos africanos que permaneceram no estado após chegar ao Brasil (figura 12 e 13).

Figura 11: Gráfico MDS bidimensional elaborado a partir da distância genética (F_{ST}) entre África, Europa, Leste Asiático, Nativo-americano (Pasto - Colômbia), diversas populações da Colômbia e



Fonte: Dados da pesquisa

Brasil (São Paulo, Mato Grosso, Brasília, Minas Gerais, Rio de Janeiro e Espírito Santo).

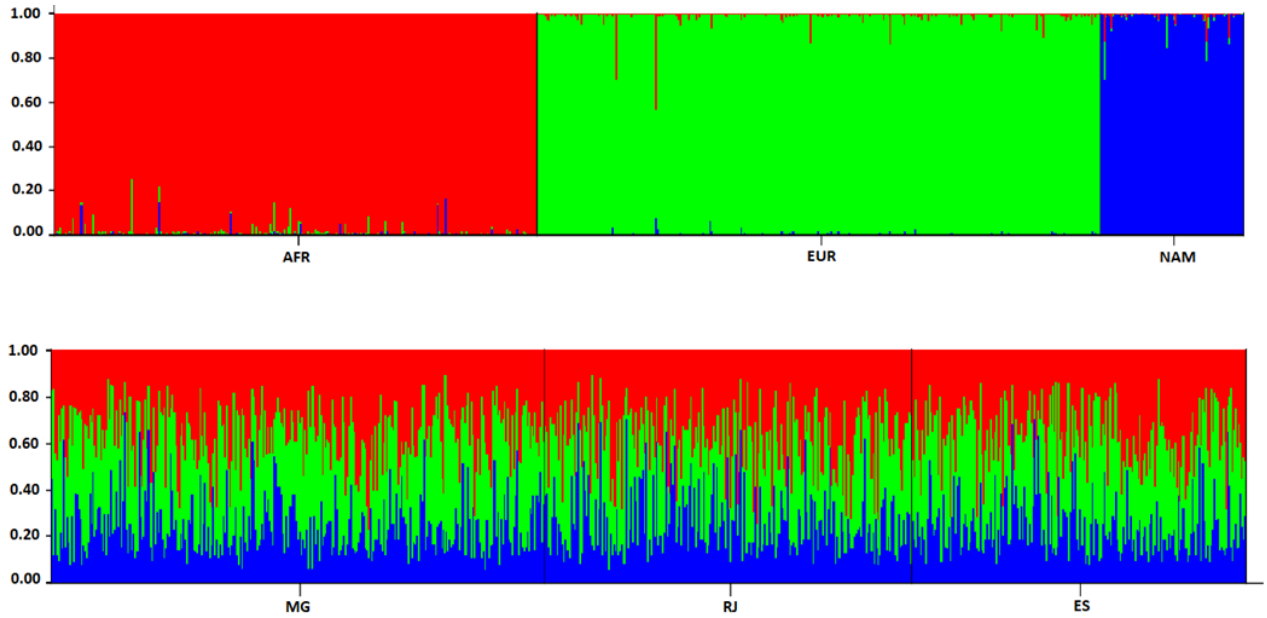
Esses resultados estão de acordo com a história da formação da população brasileira, onde o maior contingente migratório para o Brasil foi de europeus, composto principalmente por portugueses, italianos e espanhóis. O segundo maior grupo de

imigrantes são os africanos que foram trazidos pelos europeus entre os séculos XVI e XIX. A contribuição asiática para a formação populacional do país foi menor, visto que a imigração desta população para o Brasil ocorreu em menor escala.

Em um estudo realizado por Moura et. al (2015), foi relatado que em média a contribuição ancestral no sudeste do Brasil quando analisados marcadores autossômicos é de 67% de europeus, seguido por 23% de africanos e 10% de nativos-americanos. Wang e colaboradores (2008) também relataram que em toda a América Latina a contribuição de europeus para marcadores autossômicos é maior que 50%, entretanto a média cai 35% quando o cromossomo X é analisado.

Baseado no estudo de Ibarra e colaboradores (2014) em que os 32 X-InDels foram utilizados para identificar as proporções de miscigenação da população colombiana, foi realizada uma análise da contribuição ancestral genética desses marcadores nas populações de Minas Gerais, Espírito Santo e Rio de Janeiro (figura 13).

Figura 12: Representação esquemática para a população de Minas Gerais (MG), Rio de Janeiro (RJ) e Espírito Santo (ES) utilizando os marcadores 32 X-InDels obtidos através da análise do STRUCTURE baseado na contribuição da população africana (AFR), europeia (EUR) e nativo-americano (NAM) (K=3)



Fonte: Dados da pesquisa

Resque e colaboradores (2010) relatam que a contribuição de Europeus para o cromossomo X é maior na região sudeste do Brasil, mas com valores muito próximos aos da contribuição africana, sendo estruturada por 38% de europeus, 35% de africanos e 27% de nativo-americano. Os dados gerados na análise de ancestralidade pelo software STRUCUTURE, relatam uma proporção europeia ligeiramente maior em Minas Gerais

(39%) e no Espírito Santo (38%) (figura 12 e 13). Estes valores vão de encontro com a história da construção populacional destes estados, que assim como relatado anteriormente, Minas Gerais e Espirito Santo receberam uma grande quantidade de imigrantes europeus, que vieram atraídos por plantações de café, e também pelos bandeirantes (maioria descentes de europeus), que permaneceram nestas regiões após o seu desbravamento. Na população do Rio de Janeiro, a maior proporção ancestral identificada para estes marcadores foi a de contribuição africana (38%) (figura 12 e 13). Nos dados de distância genética, esta população também se apresentou mais próxima a populações africanas que de populações europeias. De acordo com a história do povoamento do Brasil (Brasil: 500 anos de povoamento, IBGE, 2007), o Rio Janeiro, assim como alguns estados do nordeste foram os que mais receberam escravos africanos, acarretando uma grande contribuição para o povoamento desta região.

Entretanto, os 32 X-InDels não se tratam de marcadores de ancestralidade, pois são marcadores que foram selecionados para identificação e não para ancestralidade, o que pode ter gerado uma subestimativa da contribuição europeia e uma sobre estimativa da contribuição africana e nativa, uma vez que os dados de distância genética confirmam a maior proximidade de São Paulo com a Europa.

Os valores de ancestralidade obtidos com os marcadores 32 X-InDels, foram comparados com dados populacionais disponíveis na literatura, que utilizaram marcadores do cromossomo Y, DNA mitocondrial e autossômicos.

Figura 13: Proporção de ancestralidade identificada em marcadores X-InDels, mtDNA, Y-SNP e AIM-InDels nas populações de Minas Gerais, Rio de Janeiro e Espírito Santo.

	Pop. Ref.	MG	RJ	ES	SP	SUDESTE
X-InDels	EUR	39%	35%	38%	41,80%	38,45%
	AFR	36%	38%	37%	31,60%	35,65%
	NAM	25%	27%	25%	26,60%	25,90%
mtDNA	EUR	31%	31%	30%	27,30%	29,83%
	AFR	34%	42%	43%	46,60%	41,40%
	NAM	33%	26%	26%	26,10%	27,78%
Y-SNP	EUR	96%	81,95%	85,88%	95,40%	89,81%
	AFR	4%	15,73%	11,37%	3,10%	8,56%
	NAM	0	2,32%	2,75%	1,50%	1,64%
AIM-InDels	EUR	59,20%	55,20%	74,10%	62,90%	62,85%
	AFR	28,90%	31,10%	13,40%	25,50%	24,73%
	NAM	11,90%	13,70%	12,50%	11,60%	12,43%

Fonte: número de amostras (n) por marcador, X-InDels (MG, ES e RJ dados da pesquisa; Martinez et al, 2019, SP n=500,), Y-SNP (MG n=50, RJ n=605- Oliveira *et al*, 2014; ES n= 253 - Figueiredo et al, 2016; SP n=65- Carvalho et al, 2010), mtDNA (ES n=100 - Ambrosio *et al*, 2015; MG n=99 - Alves-Silva et al, 2000; RJ n=205 - Simão et al, 2018; SP n=160 - Paneto et al, 2011) AIM-InDels (MG n=88 , ES n=92, RJ n=280, SP n=49 - Manta et al, 2012b).

As proporções de ancestralidade calculadas são relativas, ou seja, os 32 X-InDels permitem observar a miscigenação destas populações. Observando os dados gerados por meio da média da região sudeste, nota-se uma proporção equilibrada entre europeus, africanos e nativo-americano. Entretanto, pelo tipo de seleção dos marcadores, estes não são valores absolutos.

Na identificação humana e no estudo de genética de populações, a análise do cromossomo Y e do DNA mitocondrial são ferramentas muito importantes uma vez que, por não se recombinarem, permitem a identificação das linhagens paterna e materna, respectivamente. Já marcadores biparentais como o DNA autossômico se recombinam, e quando utilizados para análise de origem ancestral de uma amostra ou população, tende a gerar um resultado que se assemelha a uma média dos marcadores uniparentais (BRAGANHOLI, 2016). Entretanto, o cromossomo X se comporta de uma maneira mista, onde, na transmissão paterna, comporta-se como uniparental enquanto na transmissão materna, como biparental. Para os marcadores X-InDels, se esperava que os dados assemelhassem a uma média obtida através da soma de duas vezes o valor materno

(DNA mitocondrial) mais uma vez o valor paterno (cromossomo Y), sendo estes divididos por três. Os valores obtidos para proporção ancestral, na sua maioria não estão de próximo ao esperado, isso provavelmente se deva a baixa sensibilidade dos marcadores do multiplex 32 X-InDels para análise de ancestralidade.

Comparando estes valores com os dados disponíveis na literatura para o cromossomo Y, DNA mitocondrial e autossômicos se observa que os dados obtidos para os 32 X-InDels ficaram entre os dados disponíveis para DNA mitocondrial e autossômicos. Já para os valores relatados para autossômico, é possível observar que estes estão próximos a uma média entre os dados do cromossomo Y e DNA mitocondrial.

6. CONCLUSÕES

Os dados para os marcadores de inserção/deleção no cromossomo X disponíveis na literatura ainda são pouco explorados e este conhecimento sobre o assunto deve ser ampliado. Os valores de poder de discriminação e exclusão obtidos neste trabalho já demonstram seu potencial como método complementar para a análise de amostras forenses. Com os resultados obtidos neste trabalho pode-se concluir que:

- Os 32 X-InDels mostraram-se polimórficos nas populações de Minas Gerais, Rio de Janeiro e Espírito Santo, demostrando que estes marcadores genéticos são bastantes informativos para estas populações;

- Através da identificação de uma variante no MID1361, concluiu-se existir um alelo exclusivo para a população brasileira não sendo identificado em populações estrangeiras;

- Para todos os parâmetros de eficiência forenses, esse painel se mostrou altamente discriminativo nas três populações estudadas;

- Nas comparações de distância genética, se obteve valores significativos entre as populações estudadas e as estrangeiras com dados publicados na literatura. Concluindo que estas informações estão de acordo com a história do povoamento do Brasil;

- Com os dados de proporção ancestral, conclui-se que a ancestralidade identificada confirma a miscigenação destas populações.

7. REFERÊNCIAS

ALVAREZ-CUBERO, M.J. *et al*. Genetic identification of missing persons: DNA analysis of human remains and compromised samples. **Pathobiology**. v. 79. n. 5 p. 228–38. 2012.

ALVES-SILVA et. Al. 2000. The ancestry of Brazil-ian mtDNA lineages. **Am J Hum Genet** 67:444–461.

AMBROSIO *et al*. DNA 30-SNP data confirm high prevalence of African lineages in the population of Espirito Santo, Brazil Forensic Sci. Int.: Gene. Suppl. Ser., 5 (2015), pp. e346-e347 ANDREWS, C. The Hardy-Weinberg Principle. **Nature Education Knowledge.** v. 3. n. 10. p. 65. 2010.

BONFERRONI, C.E. **Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze.** v. 8. p. 3-62. 1936.

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

BUDOWLE, B.; van DAAL, A. Forensically relevant SNP classes. **Biotechniques**. V.44. p. 603-8. 2008.

BUENO, B. P. S. Dilatação dos confins: caminhos, vilas e cidades na formação da capitania de São Paulo (1532-1822), **Na. Mus. Paul**. V.17, São Paulo, 2009.

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

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

BUTLER, J.M. Advanced Topics in Forensic DNA Typing: Methodology. 1 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.

CAPUTO, M. *et al*. Potential forensic use of a 33 X-Indel panel in the Argentinean population. **Int J Legal Med.** v. 131. n. 1. p. 107-12. 2017.

CARVALHO *et al*. Analysis of paternal lineages in Brazilian and African populations. **Genet. Mol. Biol**. 2010; 33: 422-427.

CARVALHO-SILVA D, Santos FR, Rocha J and Pena SDJ (2001) The phylogeography of Brazilian y-chromosome lineages. **Am J Hum Genet** 68:281-286.

CHARLESWORTH, B.; SNIEGOWSKI, P.; STEPHAN, W. The evolutionary dynamics of repetitive DNA in eukaryotes. **Nature**. v. 371, P. 215-20, 1994.

CHOW, J.C *et al*. Silencing of the mammalian X chromosome. **Annual Review of Genomics and Human Genetics.** v. 6, p. 69–92, 2005.

DAVEY, J.W. *et al*. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. **Nat Rev Genet**. v. 12. n. 7. p. 499-510. 2011.

DESMARAIS, D. *et al*. Development of a highly polymorphic STR marker for identity testing purposes at the human androgen receptor gene (HUMARA). **J Forensic Sci.** v. 43. n.5. p. 1046–9. 1998.

DIXON, L.A. *et al*. Validation of a 21-locus autosomal SNP multiplex for forensic identification purposes. **Forensic Sci Int.** v. 154. n. 1. p. 62-77. 2005.

DONG, C. *et al*. Development of three X-linked tetrameric microsatellite markers for forensic purposes. **Mol Biol Rep**. 2014. In Press.

EDELMANN, J. *et al*. X-chromosomal 21-indel marker panel in German and Baltic populations. **Int J Legal Med.** v. 130. n. 2. p. 357-60. 2016.

EDWARDS, A. *et al*. Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. **Genomics**. v. 12. n. 2. p. 241-53. 1992.

ELLIS, N.; GOODFELLOW, P.N. The mammalian pseudoautosomal region. **Trends in Genetics**, v.5, p.406-410, 1989.

EXCOFFIER, L.; LISCHER, H.E.L. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. **Molecular Ecology Resources**. v. 10. p. 564-7. 2010.

FERRAGUT, J.F. *et al*. Genetic portrait of Jewish populations based on three sets of Xchromosome markers: Indels, Alu insertions and STRs. **Forensic Sci Int Genet**. pii: S1872-4973. n. 17. p. 30196-5. 2017.

FIGUEIREDO, R.F. *et al*. Male-specific contributions to the Brazilian population of Espirito Santo. **Int. J. Legal Med,** v. 130, p. 679-681, 2016

FONDEVILA et. al. Forensic Performance of Insertion-Deletion Marker System**s. ISFG (2011)** p-24.

FRAIGE, K. **Avaliação de Novos Sistemas Eletroforéticos Miniaturizados para Testes de Paternidades**. Dissertação de Mestrado. São Carlos. 2007

FREITAS, N.S. *et al*. X-linked insertion/deletion polymorphisms: forensic applications of a 33-markers panel. **Int J Legal Med**. v. 124. n. 6. p. 589-93. 2010.

GOULART, M. A escravidão Africana no Brasil: das origens à extinção do tráfico. Editora Alfa-Omega. São Paulo-SP, 1975, 316 p.

GRIFFITHS, A.J.F.; WESSLER, S.R.; LEWONTIN, R.C.; CARROLL, S.B. Introdução à Genética. 9. ed. Rio de Janeiro, **Guanabara Koogan**, 2009.

GUSMÃO *et al*. A GEP-ISFG collaborative study on the optimization of an X-STR Decaplex: data on 15 Iberian and Latin American populations. **Int. J. Legal Med**. v. 123. p. 227-234. 2009.

HALLENBERG C, MORLING N. A report of the 1997, 1998 and 1999 Paternity Testing Workshops of the English Speaking Working Group of the International Society for Forensic Genetics. **Forensic Sci Int**. 2001; 116:23–33.

HEARNE, C.M.; TODD, J.A. Tetranucleotide repeat polymorphism at the HPRT locus. **Nucleic Acids Res.** v. 19. n. 19. p. 5450. 1991.

IBARRA, A. *et al*. Evaluating the X chromosome-specific diversity of Colombian populations using insertion/deletion polymorphisms. **PLoS One.** v. 9. n. 1. p. e87202. 2014.

Brasil: **500 anos de povoamento**. Centro de Documentação e Disseminação de Informações. - Rio de Janeiro: **IBGE**, **2007**. 232 p.

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

JEFFREYS, A. J.; BROOKFIELD, J. F. Y; SEMEONOFF, R. Positive identification of an immigration test-case using human DNA fingerprints. **Nature,** v.317, p.818-819, 1985b.

JOBLING, M. A., GILL, P. Encoded evidence: DNA in forensic analysis **Nat. Rev. Genet**., 5 (10) (2004), pp. 739-751

JORDE, L.B.; WATKINS, W.S.; BAMSHAD, M.J. Population genomics: a bridge from evolutionary history to genetic medicine. **Human Mol. Genetics**, v.10, n.20, p.2199-2207, 2001.

KRÜGER, J. *et al*. Zur Verwendung der sauren Erythrocytenphosphatase bei der Vaterschaftsbegutachtung. **Dtsch Z Gerichtl Med**. v. 64. n. 2. p. 127–46. 1968.

LEVY et. al. The diploid genome sequence of an individual human. **PLoS Biol**. 2007;5: e254.

MANTA, F. *et al*. InDel markers: genetic diversity of 38 polymorphisms in Brazilian populations and application in a paternity investigation with post mortem material. **Forensic Sci Int Genet.** v. 6. n. 5. p. 658-61. 2012a.

MANTA et. al. Revisiting the genetic ancestry of brazilians using autosomal AIM-Indels **PLoS One,** 8 (2012), p. 2012b.

MARTINEZ, J. **Polimorfismos de inserção/deleção no cromossomo X:** Análise de 32 marcadores na população do estado de São Paulo (BRASIL). Tese de doutorado. Faculdade de Ciências Farmacêuticas, Universidade Júlio de Mesquita Filho. 2017.

MARTINEZ, J; POLVERARI F. S; SILVA, F. A. J; BRAGANHOLI, D. F; FERRAZ, J. A. M. F; GUSMÃO, L; CICARELLI, R. M. B. Genetic characterization of 32 X-InDels in a population sample from São Paulo State (Brazil). **Int J Legal Med** (2019). https://doi.org/10.1007/s00414-018 01988-w

MARTINS, J.A. **Estudo de polimorfismos do cromossomo X na população da região sudeste do Brasil**. Tese (Doutorado em Biociências e Biotecnologia Aplicadas à Farmácia). Faculdade de Ciências Farmacêuticas, Universidade Estadual Paulista Júlio de Mesquita Filho. 2011.

MARTINS, J.A. *et al*. Brazilian genetic database of chromosome X. **Mol Biol Rep**. v. 41. n. 6. p. 4077-80. 2014.

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

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

MULLANEY, J. M.; MILLS, R. E.; PITTARD, W. S.; DEVINE, S. E. Small insertions and deletions (INDELs) in human genomes. **Hum. Mol. Genet**. 19(2). P.131-136. 2010.

NARLOCH, L & TEIXEIRA, D. Guia politicamente incorreto da América Latina. São Paulo: Editora Leya, 2011. 336 p.

NLM. U.S. National Library of Medicine**.** X Chromosome. In: **Genetic home reference: Your guide to understanding genetic conditions** [Internet]. 2012.

MILLS, R.E. *et al*. An initial map of insertion and deletion (InDel) variation in the human genome. **Genome Res**. v. 16. n. 9. p. 1182-90. 2006.

NACHMAN, M. W.; CROWELL, S. L. Estimate of the mutation rate per nucleotide in humans. **Genetics**, v. 156, p. 297-304, 2000.

NEI, M.; ROYCHOUDHURY, A.K. Sampling variances of heterozygosity and genetic distance. **Genetics**. v. 76. p. 379–90. 1974.

NEVES, G.S. *et al*. Espírito Santo: nossa história nossa gente. Vitória: Grafer, 1997.

NUSSBAUM, R. L. Thompson & Thompson. **Genética Médica**. 7ªed. Rio de Janeiro: Elsevier. 2008.

NUSSBAUM, R. L; MCINNES, R. R; WILLARD, H. F. Thompson & Thompson. **Genética Médica**. 8ªed. Rio de Janeiro: Elsevier. 2016.

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.J.B.F. et al. Population genetic analysis of insertion–deletion polymorphisms in a Brazilian population using the Investigator DIPplex kit. **Forensic Sci Int: Genet.** v. 19, p. 10-14, 2015.

PANETO et. al. A single multiplex PCR and SNaPshot minisequencing reaction of 42 SNPs to classify admixture populations into mitochondrial DNA haplogroups. Mitochondrion 11(2):296–302. 2011.

PEREIRA, R. *et al*. Genetic diversity of 10 X chromosome STRs in northern Portugal. **Int J Legal Med.** V. 121, p. 192-197. 2007.

PEREIRA, V. *et al*. Genetic characterization of Somali and Iraqi populations using a set of 33 X-chromosome Indels. **Forensic Sci. Int. Genet. Suppl. Ser**. v. 3. n. 1. p. e137–8. 2011.

PEREIRA, R. *et al*. A method for the analysis of 32 X chromosome insertion deletion polymorphisms in a single PCR. **Int J Legal Med**. v. 126. n. 1. p. 97-105. 2012.

PEREIRA, R. *et al*. A new multiplex for human identification using insertion/deletion polymorphisms. **Electrophoresis.** v. 30. n. 21. p. 3682-90. 2009.

PEREIRA, R. *et al*. A GHEP-ISFG collaborative study on the genetic variation of 38 autosomal indels for human identification in different continental populations. **Forensic Sci Int: Genet.** v.32, P. 18-25, 2018.

PIMENTA, J. R; PENA S. D. J. Efficient human paternity testing with a panel of 40 short insertion–deletion polymorphisms. Genet Mol Res 9(1):601–607. 2010.

PINHEIRO, Maria de Fátima. Genética e biologia forense, e criminalística. In: Faculdade de Medicina da Universidade do Porto. In: NOÇÕES GERAIS SOBRE OUTRAS CIÊNCIAS FORENSES MEDICINA LEGAL, 2004. 57p.

PINTO, N. *et al*. Assessing paternities with inconclusive STR results: The suitability of bi-allelic markers. **Forensic Sci Int Genet**. v. 7. n. 1. p. 16-21. 2013a.

PINTO, N. *et al*. Paternity exclusion power: comparative behaviour of autosomal and Xchromosomal markers in standard and deficient cases with inbreeding. **Forensic Sci Int Genet.** v. 7. n. 2. p. 290-5. 2013b.

PINTO, J.C. *et al*. Mirandese language and genetic differentiation in Iberia: a study using X chromosome markers. **Ann Hum Biol.** v. 42. n. 1. p. 20-5. 2015.

POLVERARI, F. S. **Caracterização genética da população do Estado do Mato Grosso e do Distrito Federal (Brasília) pela análise de 32 polimorfismos de inserção/deleção (InDels) no cromossomo X.** Tese (Doutorado em Biotecnologia). Instituto de Química, Universidade Estadual Paulista Júlio de Mesquita Filho, 2018.

PRITCHARD, JK; STEPHENS, M.; DONNELLY, P. Inference of population structure using multilocus genotype data. **Genetics**. v. 155. p. 945–59. 2000.

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

ROSS, M.T. *et al*. The DNA sequence of the human X chromosome. **Nature**. v. 434. n. 7031. p. 325- 37. 2005.

SALZANO, F. M.; BORTOLINI M. C. The evolution and genetics of Latin American population. **Cambridge University Press**; Cambridge: 2002. p. 512.

SCHAFFNER, S.F. The X chromosome in population genetics. **Nat Rev Genet**. v. 5. n. 1. p. 43-51. 2004.

SCHNEIDER P. M. Beyond STRs: The Role of Diallelic Markers in Forensic Genetics. **Transfus Med Hemother**, v.39, p.176–180, 2012

SIMÃO *et al*. (2018). Defining mtDNA origins and population stratification in Rio de Janeiro. **Forensic Science International: Genetics**, 34, 97–104. doi:10.1016/j.fsigen.2018.02.003

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.

SOBRINO, B.; BRION, M.; CARRACEDO, A. SNPs in forensic genetics: a review on SNP typing methodologies **Forensic Sci. Int.,** 154 (2–3) (2005), pp. 181-194

SUN, K. *et al*. Development of an 18 X-InDel multiplex PCR system. **Fa Yi Xue Za Zhi.** v. 30. n. 2. p. 101-9. 2014.

SZIBOR, R. *et al*. Use of X-linked markers for forensic purposes. **Int J Legal Med**. v. 117. n. 2. p. 67-74. 2003.

SZIBOR, R. X-chromosomal markers: past, present and future. **Forensic Sci Int Genet**. v. 1. n. 2. p. 93-9. 2007.

TANDE, C.M. DNA Typing: A New Investigatory Tool. **Duke Law Journal**, v.1989, n. 2, p.474-494, 1989.

THE 1000 GENOMES PROJECT CONSORTIUM. A global reference for human genetic variation. **Nature,** v**.** 526, p.68–74. 2015. https://doi.org/10.1038/nature15393

THOMPSON, R.; ZOPPIS, S.; MCCORD, B. An overview of DNA typing methods for human identification: past, present, and future. **Methods Mol Biol.** v. 830. p. 3-16. 2012.

TILLMAR, A. **Populations and Statistics in Forensic Genetics**. Tese (Doutorado). Faculty of Health Sciences da Linköping University. 2010.

TILLMAR, A. *et al*. Using X-chromosomal markers in relationship testing: calculation of likelihood ratios taking both linkage and linkage disequilibrium into account. **Forensic Sci Int Genet.** v. 5. n. 5. p. 506-11. 2011.

TILLMAR, A. *et al.* DNA Commission of the International Society for Forensic Genetics (ISFG): Guidelines on the use of X-STRs in kinship analysis. **Forensic Sci Int Genet.** v. 29. p. 269-75. 2017.

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.** 2014.

WALSH, PS. METZGER DA, HIGUCHI R. Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. **Biotechniques**. 1991.

WANG, Z. *et al*. Population genetics of 30 insertion-deletion polymorphisms in two Chinese populations using Qiagen Investigator® DIPplex kit. **Forensic Sci Int Genet.** v. 11. 2014.

WEBER, J.L. *et al*. Human diallelic insertion/deletion polymorphisms. **Am J Hum Genet**. v. 71. n. 4. p. 854-62. 2002.

YUAN, L. *et al*. Population genetics analysis of 38 STR loci in the she population from Fujian province of China. **Legal Medicine**. 2014. In Press.

ZARRABEITIA MT, *et al*. Analysis of 10 X-linked tetranucleotide markers in mixed and isolated populations. **Forensic Sci Int Genet.** v.3. p. 63-66. 2009.

ZIĘTKIEWICZ, E. *et al*. Current genetic methodologies in the identification of disaster victims and in forensic analysis. **J Appl Genet**. v. 53. n. 1. p. 41-60. 2012.

ANEXO I

International Journal of Legal Medicine
https://doi.org/10.1007/s00414-018-01988-w

POPULATION DATA

Genetic characterization of 32 X-InDels in a population sample from São Paulo State (Brazil)

Juliana Martinez[1] · Fernanda Silva Polverari[1] · Flávia Alves de Jesus Silva[1] · Danilo Faustino Braganholi[1] · Joyce Aparecida Martins Lopes Ferraz[2] · Leonor Gusmão[3] · Regina Maria Barretto Cicarelli[1]

Received: 23 July 2018 / Accepted: 19 December 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract
X-chromosomal markers can be useful in some forensic cases, where the analysis of the autosomal markers is not conclusive. In this study, a population sample of 500 unrelated individuals born in São Paulo State was characterized for 32 X-InDel markers. No deviations from the Hardy–Weinberg equilibrium were detected, except for MID1361. The 32 X-InDels showed an accumulated power of discrimination of 0.9999999999993 in females and 0.99999993 in males and an exclusion chance of 0.999996 in trios and 0.99995 in duos. São Paulo showed lower genetic distances to the Colombian admixed and European populations than to Native American, Asian, or African populations. Ancestry analysis revealed 41.8% European, 31.6% African, and 26.6% Native American contributions. Segregation analysis was performed in 101 trios, and the mutation rate was estimated to be low.

Keywords Forensic genetics · Human identification · Kinship testing · Insertion/deletion polymorphism · X chromosome

The analysis of autosomal short tandem repeats (STRs) is usually performed to solve human identification or kinship cases, although complementary tools can be necessary for complex cases. Due to low mutation rates, SNPs and insertion/deletion polymorphisms (InDels) can complement STRs in kinship investigation, when few incompatibilities are observed [1, 2]. Because men have a single copy of the X chromosome that is transmitted to all his daughters, X-chromosomal markers can be particularly useful in sibling analysis, when the alleged father is not available. In these cases, autosomal markers are not informative to exclude the kinship [3–5]. Markers on the X chromosome are also able to differentiate some pedigrees not distinguishable using autosomal markers (e.g., half-siblings, grandparent–grandchild, and avuncular) [6]. Freitas et al. [7] described a multiplex including 33 X-InDels, with amplicon sizes between 100 bp and 300 bp, presenting high diversities in the population from Belém, Brazil. Another PCR multiplex with 32 X-InDels was described by Pereira et al. [8], to be used in human identification and paternity testing in African, European, and Asian populations. When compared to the previous, the 32 X-InDels are amplified in shorter amplicons (< 160 pb), being suitable for the analysis of degraded samples.

The aim of this study was to characterize the 32 X-InDels from Pereira et al. [8] in a population sample from São Paulo State (in the South West region), to evaluate their utility in forensic analysis. A total of 500 samples from unrelated subjects (282 females and 218 males) born in São Paulo State were selected for this study. Samples from 101 father/mother/daughter trios were also included for segregation analysis. Blood samples were collected on Whatman FTA™ card (GE Healthcare, Little Chalfont, UK) and DNA was extracted using Chelex 100 resin (Bio-Rad Laboratories, California, USA) at 5% concentration.

This study was approved by the ethical committee of the São Paulo State University (UNESP), School of Pharmaceutical Sciences, São Paulo, Brazil, and all participants signed an informed consent.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00414-018-01988-w) contains supplementary material, which is available to authorized users.

✉ Juliana Martinez
jumrtz@hotmail.com

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

[2] Research Centre for Biochemistry and Molecular Biology, Medical School of São José do Rio Preto (FAMERP), São Paulo, Brazil

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

Published online: 05 January 2019

The 32 X-InDels were amplified in a single PCR multiplex, using the conditions described in Pereira et al. [8], and adjusting the final reaction volume to 5 μL. The PCR products were run in an ABI 3500 Genetic Analyzer (AB, Thermo Fischer Scientific). Genotypes were assigned with GeneMapper ID-X™ software (AB, Thermo Fischer Scientific).

The full list of genotypes obtained is in Supplementary Table S1. A variant allele at MID1361 was observed in three females. In two cases, the genotypes of the daughters were also analyzed. The first case consisted in a trio (father, mother, and daughter) in which the daughter received the variant allele from the mother (Supplementary Figure S1-A). The second case included two daughters (full-sibling), where one received the short allele from the mother, and the other one received the variant allele (Supplementary Figure S1-B). The MID1361 usually presents the "-/ACA" alleles, but three alleles are described in dbSNP as "–/AA/ACA." The variant found in our samples has 1 bp less than the short allele, therefore representing a new allele. The 1000 Genomes Project [9] describes a deletion of a T nucleotide (rs202070714) near the MID1361 locus, at position 119,614,551. This deletion, associated with the MID1361 short allele, generates an amplicon with the same size of the variant that we observed (see Supplementary Figure S1).

Allele and haplotype frequencies were calculated using ARLEQUIN software v3.5 [10]. Since no significant differences were observed between allele frequency distributions in male and female samples ($F_{ST} \leq 0.0007$; $p = 0.194 \pm 0.004$), the allele frequencies were estimated for the total sample from São Paulo (Supplementary Table S2). All markers were polymorphic, with a mean locus diversity of 0.414 (± 0.084), which was higher than reported for European (Portugal, 0.388; Miranda do Douro, 0.376; Spain/Zamora, 0.386), Asian (Macau, 0.347), and African (Angola and Mozambique, 0.408) populations [8, 11]. The MID3701 was the most diverse and MID2637 was the less informative.

An exact test of the Hardy–Weinberg equilibrium was performed in female samples using the ARLEQUIN software, revealing a significant deviation in MID1361 (Supplementary Table S2). The same result was observed in a sample from Germany [12]. Although no null alleles were described for this marker, 10 random samples (from homozygotes for the MID1361 long allele) were amplified at a lower annealing temperature of 54 °C. The genotypes remained the same, revealing no signs of null alleles. A search in dbSNP (https://www.ncbi.nlm.nih.gov/snp/) showed two SNPs in the primers' annealing region, the rs914334630 (located in the annealing region of the forward primer) and the rs749840100 (in the reverse primer region). However, these are rare events, not validated by the 1000 Genomes Project [9].

For male samples, an exact test for LD was performed using the ARLEQUIN software, resulting in 496 pairwise analysis. After Bonferroni correction for multiple tests ($p \leq 0.05/496$), a significant association was observed for 4 pairs of markers, namely MID356–MID357 ($p < 0.00005$), MID3703-MID3774 ($p < 0.00005$), MID3719-MID2089 ($p < 0.00005$), MID3692-MID2652 ($p < 0.00005$).

The MID356-MID357 markers are very close on the X chromosome (~ 5.2 Kb; 0.012 cM), and therefore, they should be treated as haplotypes. Significant associations were also found for these markers in different populations from Africa, Europe, America, and Asia [7, 8, 11–16]. For MID3703-MID3774 markers, LD was also observed in samples from Colombia [14], Africa, Europe, and Brazil [17]. Between MID3719-MID2089, LD was previously detected in populations from Spain (Majorca) [13], Portugal [8], Somalia [18], Germany, Baltic [12], and Colombia [14]. The MID3690-MID3719 showed LD in Angola, Mozambique [8], Colombia [14], Germany, and Baltic [12] populations. In fact, MID3690-MID3719-MID2089 are separated from each other by only ~ 170 Kb (0.137 cM), forming a cluster on the X chromosome. For this reason, Pereira et al. [8] suggest analyzing these markers as a single haplotypic block. Finally, for MID3692-MID2652 markers, LD was only detected in the São Paulo population, and not in other populations. Since LD can result from other factors than linkage, such as admixture and population substructure [19], it is possible that population substructure is favoring the occurrence of LD in the sample from São Paulo. For markers in LD, we calculated the haplotype frequency by simple counting and the data are available in supplementary table S3.

Statistical parameters of forensic interest were calculated for the 32 X-Indels (Supplementary Table S4) in the São Paulo population, using the tool available at the ChrX-STR.org 2.0 database (http://www.chrx-str.org/). In kinship trio investigations (mother, daughter, and alleged father), the combined value of mean exclusion chance (MEC_T) is 0.99996. For situations with duos (alleged father and daughter), the combined value of MEC_D is 0.9999. If the father is absent, and the alleged paternal grandmother is tested instead, the combined MEC_{PGM} value is 0.997. The combined power of discrimination is 0.9999999999993 for females and 0.999999993 for males.

Population comparisons by mean of the pairwise genetic distances (F_{ST}), and the corresponding non-differentiation p-values (for 10,000 permutations), were estimated between São Paulo and other populations with data available for the same 32 X-InDels multiplex, using the ARLEQUIN software (Supplementary Table S5). For an easier visualization, the pairwise F_{ST} matrix was represented in a multidimensional scaling (MDS) using the software STATISTICA 13.2 Trial (StatSoft Inc., http://www.statsoft.com.br) (Supplementary Figure S2).

In most comparisons, significant differences ($p < 0.0002$) were obtained between São Paulo and other populations

(Supplementary Table S6), demonstrating the need for specific allele frequency databases for these populations. Low genetic distances, with non-significant p values, were obtained between São Paulo and two Colombian departments (Antioquia [$F_{ST} = 0.0092$] and Norte de Santander [$F_{ST} = 0.0096$]), characterized by a high European ancestry [20]. Additionally, São Paulo showed lower genetic distances with European populations ($0.0120 \leq F_{ST} \leq 0.0169$) than with Native Americans or Sub-Saharan African populations. Although the genetic distance to Portugal ($F_{ST} = 0.0120$; $p < 0.00005$) was lower than to Spain or Denmark, it was statistically significant.

Ancestry contributions were estimated using the software STRUCTURE v.2.3.4. [21]. Considering the historical formation of São Paulo State, we assumed a tri-hybrid contribution from Native Americans, Europeans, and Africans ($K = 3$). Three independent interactions were performed with 100,000 burn-in steps followed by 100,000 Markov Chain Monte Carlo (MCMC). The "use population information to test for migrants" option was used with the admixture model.

In accordance with the historical data, Moura et al. [22] described that for autosomal markers, the ancestry in Brazilian Southeast region is 67% European, followed by 23% African and 10% Native American. Wang et al. [23] also reported that in Latin America, the European contribution of autosomal markers is higher than 50%, but the average falls to 35% when the X chromosome is analyzed. Resque et al. [17] also found that the European contribution of the X chromosome is higher in Brazil southeastern region, but with values closely to the African contribution, with 38% of Europeans, 35% of African, and 27% of Native Americans admixture.

The results obtained for the 32 X-InDels showed 41.8% European, 31.6% African, and 26.6% Native American contributions to the São Paulo population. Since the markers used in this study were not selected for ancestry inferences, the proportion of European admixture is most likely underestimated, with an overestimation of the African and Native admixtures [24].

Expected mutation rates for InDels are of an order of magnitude of 10^{-9} [25], much lower than the observed for STRs, for which the mutation rate is around 10^{-4} to 10^{-3} per locus per generation [26]. Therefore, data on a very high number of meiosis are necessary to obtain reliable estimates of InDel mutation rates.

Segregation analysis performed in 101 father/mother/daughter trios showed the expected transmission pattern from parents to daughters, with no mutations observed in 202 meiosis (6464 allelic transmissions). For X-InDels, no mutations were previously found in 214 trios from Somalia and Portugal (for the same set of X-InDels) [8, 18], or in 97 trios from Germany (for 26 X-InDels) [12]. Considering all the results for the 32 X-InDel set, no mutations were found in 20,160 allelic transmissions, corresponding to a 95% upper bound limit of 0.00018 for the X-InDels mutation rate.

In conclusion, the 32 X-InDel markers presented high diversities in the São Paulo population and a high forensic efficiency. The results of genetic distance analysis and ancestry estimates showed a closer proximity of this State to the European populations than to Native American or African populations. Segregation analysis confirmed the low mutation rate expected for InDel markers on the X chromosome. Therefore, this system can be an efficient complementary tool to the analysis of autosomal markers, in kinship investigations and forensic cases involving this population.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Gomes C, Magalhães M, Alves C, Amorim A, Pinto N, Gusmão L (2012) Comparative evaluation of alternative batteries of genetic markers to complement autosomal STRs in kinship investigations: autosomal indels vs. X-chromosome STRs. Int J Legal Med 126: 917–921. https://doi.org/10.1007/s00414-012-0768-5
2. Pinto N, Magalhães M, Conde-Sousa E, Gomes C, Pereira R, Alves C, Gusmão L, Amorim A (2013) Assessing paternities with inconclusive STR results: the suitability of bi-allelic markers. Forensic Sci Int Genet 7:16–21. https://doi.org/10.1016/j.fsigen.2012.05.002
3. Szibor R, Krawczak M, Hering S, Edelmann J, Kuhlisch E, Krause D (2003) Use of X-linked markers for forensic purposes. Int J Legal Med 117:67–74. https://doi.org/10.1007/s00414-002-0352-5
4. Szibor R (2007) X-chromosomal markers: past, present and future. Forensic Sci Int Genet 1:93–99. https://doi.org/10.1016/j.fsigen.2007.03.003
5. Tillmar AO, Kling D, Butler JM, Parson W, Prinz M, Schneider PM, Egeland T, Gusmão L (2017) DNA Commission of the International Society for Forensic Genetics (ISFG): guidelines on the use of X-STRs in kinship analysis. Forensic Sci Int Genet 29: 269–275. https://doi.org/10.1016/j.fsigen.2017.05.005
6. Pinto N, Gusmão L, Amorim A (2011) X-chromosome markers in kinship testing: a generalisation of the IBD approach identifying situations where their contribution is crucial. Forensic Sci Int Genet 5:27–32. https://doi.org/10.1016/j.fsigen.2010.01.011
7. Freitas NSC, Resque RL, Ribeiro-Rodrigues EM, Guerreiro JF, Santos NPC, Ribeiro-dos-Santos Â, Santos S (2010) X-linked insertion/deletion polymorphisms: forensic applications of a 33-markers panel. Int J Legal Med 124:589–593. https://doi.org/10.1007/s00414-010-0441-9
8. Pereira R, Pereira V, Gomes I, Tomas C, Morling N, Amorim A, Prata MJ, Carracedo Á, Gusmão L (2012) A method for the analysis of 32 X chromosome insertion deletion polymorphisms in a single PCR. Int J Legal Med 126:97–105. https://doi.org/10.1007/s00414-011-0593-2
9. THE 1000 GENOMES PROJECT CONSORTIUM (2015) A global reference for human genetic variation. Nature 526:68–74. https://doi.org/10.1038/nature15393
10. Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and windows. Mol Ecol Resour 10:564–567. https://doi.org/10.1111/j.1755-0998.2010.02847.x

11. Pinto JC, Prereira V, Marques SL, Amorim A, Alvarez L, Prata MJ (2015) Mirandese language and genetic differentiation in Iberia: a study using X chromosome markers. Ann Hum Biol 42:20–25. https://doi.org/10.3109/03014460.2014.944215
12. Edelmann J, Kohl M, Dressler J, Hoffmann A (2016) X-chromosomal 21-indel marker panel in German and Baltic populations. Int J Legal Med 130:357–360. https://doi.org/10.1007/s00414-015-1221-3
13. Ferragut JF, Bentayebi K, Pereira R, Castro JA, Amorim a, Ramon C, Picornell A (2017) Genetic portrait of Jewish populations based on three sets of X-chromosome markers: Indels, Alu insertions and STRs. Forensic Sci Int Genet 31:e5–e11. https://doi.org/10.1016/j.fsigen.2017.09.008
14. Ibarra A, Restrepo T, Rojas W, Castilho A, Amorim A, Martinez B, Burgos G, Ostos H, Álvarez K, Camacho M, Suarez Z, Pereira R Gusmão L (2014) A. et al. Evaluating the X chromosome-specific diversity of Colombian populations using insertion/deletion) polymorphisms. PLoS One 9:e87202. https://doi.org/10.1371/journal.pone.0087202
15. Caputo M, Amador MA, Santos S, Corach D (2017) Potential forensic use of a 33 X-Indel panel in the Argentinean population. Int J Legal Med 131:107–112. https://doi.org/10.1007/s00414-016-1399-z
16. Pereira V, Tomas C, Sanchez JJ, Syndercombe-Court D, Amorim A, Gusmão L, Prata MJ, Morling N (2015) The peopling of Greenland: further insights from the analysis of genetic diversity using autosomal and X-chromosomal markers. Eur J Hum Genet 23:245–251. https://doi.org/10.1038/ejhg.2014.90
17. Resque RL, Freitas NS, Rodrigues EM, Guerreiro JF, Santos NP, Ribeiro dos Santos A, Zago MA Santos S (2010) Estimates of interethnic admixture in the Brazilian population using a panel of 24 X-linked insertion/deletion markers. Am J Hum Biol 22:849–852. https://doi.org/10.1002/ajhb.21089
18. Pereira V, Moncada E, Diez IE, Tomas C, Amorim A, Morling N, Gusmão L, Prata MJ (2011) Genetic characterization of Somali and Iraqi populations using a set of 33 X-chromosome Indels. Forensic Sci Int Genet 3:e137–e138. https://doi.org/10.1016/j.fsigss.2011.08.069
19. Zarrabeitia MT, Alonso A, Martin J (2006) Study of six X-linked tetranucleotide microsatellites: population data from five Spanish regions. Int J Legal Med 120:147–150. https://doi.org/10.1007/s00414-005-0057-7
20. Ossa H, Aquino J, Sierra S, Ramírez A, Carvalho EF, Gusmão L (2015) Analysis of admixture in Native American populations from Colombia. Forensic Sci Int Genet, Supplement Series 5:e332–e333. https://doi.org/10.1016/j.fsigss.2015.09.132
21. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959
22. Moura RR, Coelho AVC, Balbino VQ, Crovella S, Brandão LAC (2015) Meta-analysis of Brazilian genetic admixture and comparison with other Latin America countries. Am J Hum Biol 27:674–680. https://doi.org/10.1002/ajhb.22714
23. Wang S, Ray N, Rojas W, Parra MV, Bedoya G, Gallo C, Poletti G, Mazzotti G, Hill K, Hurtado AM, Camrena B, Nicolini H, Klitz W, Barrantes R, Molina JA, Freimer NB, Bortolini MC, Salzano FM, Petzl-Erler AL, Tsuneto LT, Dipierri JE, Alfaro EL, Bailliet G, Bianchi NO, Lop E, Rothhammer F, Excoffier L, Ruiz-Linares A (2008) Geographic patterns of genome admixture in Latin American Mestizos. PLoS Genet 4:e1000037. https://doi.org/10.1371/journal.pgen.1000037
24. Aquino JG, Jannuzzi J, Carvalho EF, Gusmão L (2015) Assessing the suitability of different sets of indels in ancestry estimation. Forensic Sci Int 5:e34–e36. https://doi.org/10.1016/j.fsigss.2015.09.014
25. Campbell CD, Eichler EE (2013) Properties and rates of germline mutations in humans. Trends Genet 29:575–584. https://doi.org/10.1016/j.tig.2013.04.005
26. Sun JX, Helgason A, Masson G, Ebenesersdóttir SS, Li H, Mallick S, Gnerre S, Patterson N, Kong A, Reich D, Stefansson K (2012) A direct characterization of human mutation based on microsatellites. Nat Genet 44(10):1161–1165. 10.1038/ng.2398