

UNIVERSIDADE ESTADUAL PAULISTA “JÚLIO DE MESQUITA FILHO”  
FACULDADE DE CIÊNCIAS FARMACÊUTICAS

JULIANA MARTINEZ

**POLIMORFISMOS DE INSERÇÃO/DELEÇÃO NO CROMOSSOMO X: ANÁLISE  
DE 32 MARCADORES NA POPULAÇÃO DO ESTADO DE SÃO PAULO (BRASIL)**

ARARAQUARA – SP

2018

UNIVERSIDADE ESTADUAL PAULISTA “JÚLIO DE MESQUITA FILHO”  
FACULDADE DE CIÊNCIAS FARMACÊUTICAS

JULIANA MARTINEZ

**POLIMORFISMOS DE INSERÇÃO/DELEÇÃO NO CROMOSSOMO X: ANÁLISE  
DE 32 MARCADORES NA POPULAÇÃO DO ESTADO DE SÃO PAULO (BRASIL)**

Tese apresentada à Faculdade de Ciências Farmacêuticas, Universidade Estadual Paulista “Júlio de Mesquita Filho”, como parte dos requisitos para obtenção do título de Doutora em Biociências e Biotecnologia Aplicadas à Farmácia. Área de concentração: Biologia Molecular.

Orientadora: Prof.<sup>a</sup> Dr.<sup>a</sup> Regina Maria Barretto Cicarelli

ARARAQUARA – SP

2018

**Ficha Catalográfica**

Elaborada Por Diretoria Técnica de Biblioteca e Documentação  
Faculdade de Ciências Farmacêuticas  
UNESP – Campus de Araraquara

**M385p** Martinez, Juliana  
Polimorfismos de inserção/deleção no cromossomo x: análise de 32 marcadores na população do estado de São Paulo (Brasil) / Juliana Martinez. – Araraquara, 2018.  
107 f. : il.

Tese (Doutorado) – Universidade Estadual Paulista. “Júlio de Mesquita Filho”. Faculdade de Ciências Farmacêuticas. Programa de Pós-Graduação em Biociências e Biotecnologia Aplicadas á Farmácia. Área de pesquisa em Biologia Molecular.

Orientadora: Regina Maria Barretto Cicarelli.

1. Genética forense. 2. Identificação humana. 3. Teste de paternidade. 4. Indel. 5. Polimorfismos de inserção deleção. 6. Cromossomo X. 7. X-Indel. 8. São Paulo, SP – Brasil. I. Cicarelli, Regina Maria Barreto, orient. II. Título.

Ficha Catalográfica elaborada por Maria Irani Coito CRB-8/4.440.

**CAPES: 40500005**

JULIANA MARTINEZ

POLIMORFISMOS DE INSERÇÃO/DELEÇÃO NO CROMOSSOMO X: ANÁLISE DE 32  
MARCADORES NA POPULAÇÃO DO ESTADO DE SÃO PAULO

Tese de Doutorado apresentada à Faculdade de  
Ciências Farmacêuticas da Universidade Estadual  
Paulista – UNESP, Campus de Araraquara como  
requisito para a obtenção do título de Doutor(a) em  
Biotecnologia e Biotecnologia aplicadas à Farmácia

Araraquara, 30 de novembro de 2017.

BANCA EXAMINADORA



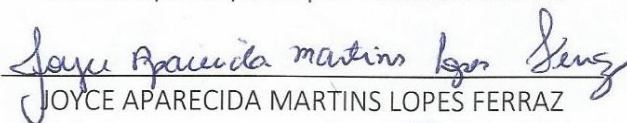
---

REGINA MARIA BARRETTO CICARELLI

---

LEONOR GUSMÃO

Membro participante por videoconferência



JOYCE APARECIDA MARTINS LOPES FERRAZ



RAQUEL MANTUANELLI SCAREL CAMINAGA

---

CELSO TEIXEIRA MENDES JÚNIOR

Membro participante por videoconferência

Dedico a todos aqueles que em si  
carregam um pedaço de mim.

Amo vocês!

## Agradecimentos

---

À **VIDA**, por me dar tanta saúde e energia para que eu possa ir em busca dos meus sonhos e por me permitir estar próxima de todos aqueles que amo imensamente.

Aos meus pais, **RAMON** e **LUCIANA**, e à minha irmã, **MARIA EDUARDA**, por me amarem incondicionalmente e não medirem esforços para que meus sonhos sejam alcançados.  
Obrigada por todo apoio!

Ao meu marido **FÁBIO**, pelo cumplicidade e compreensão diária. Eu não poderia desejar melhor companheiro que você. Te amo!

Aos meus avós **PEDRO** e **MARIA ALICE**, por cuidarem de mim com tanto carinho.

Aos meus filhos peludos **NERO** e **KIARA**, que me permitiram conhecer o que é o verdadeiro amor!

À orientadora Prof.<sup>a</sup> Dr.<sup>a</sup> **REGINA MARIA BARRETTO CICARELLI**, por ter acreditado em mim e tornado esse trabalho possível.

Aos amigos de pós-graduação, **FLÁVIA**, **FERNANDA**, **ANDRÉA**, **BIANCA**, **ISABELA** e **DANILO**, por toda troca de conhecimento e por tornarem os meus dias tão agradáveis.

À **FACULDADE DE CIÊNCIAS FARMACÊUTICAS** por ter oportunizado essa grande conquista.

À **COORDENAÇÃO DE APERFEIÇOAMENTO DE PESSOAL DE NÍVEL SUPERIOR (CAPES)** pelo importante auxílio financeiro que contribuiu para a realização deste.

À todos vocês minha eterna gratidão!

“E o futuro é uma astronave que tentamos pilotar  
Não tem tempo, nem piedade, nem tem hora de chegar  
Sem pedir licença muda nossa vida  
Depois convida a rir ou chorar

Nessa estrada não nos cabe conhecer ou ver o que virá  
O fim dela ninguém sabe bem ao certo onde vai dar  
Vamos todos numa linda passarela  
De uma aquarela que um dia enfim...  
Descolorirá”

**Aquarela** (Toquinho)

# CAPÍTULO I



MARTINEZ, J. **Polimorfismos de inserção/deleção no cromossomo X: análise de 32 marcadores na população do estado de São Paulo (Brasil)**. 2017. 107p. Tese (Doutorado) – Faculdade de Ciências Farmacêuticas, Universidade de Estadual Paulista Júlio de Mesquita Filho, Araraquara.

## RESUMO

Na rotina da genética forense, o uso exclusivo dos marcadores STRs (*Short Tandem Repeats*) em situações que a amostra biológica apresenta-se degradada pode gerar um resultado final estatístico inconclusivo, tornando fundamental a análise de marcadores adicionais para a resolução do caso. Utilizado como método complementar, os polimorfismos InDels (inserção/deleção) têm mostrado grande potencial para superar as limitações dos marcadores tradicionais. A análise de regiões do cromossomo X também vem ganhando significativa importância nesses estudos, especialmente nos casos em que a análise dos autossômicos não é suficiente. Nessa perspectiva, este trabalho teve por objetivo geral caracterizar a população do estado de São Paulo para 32 polimorfismos de inserção/deleção no cromossomo X (32 X-InDels) e avaliar a utilidade desse multiplex na resolução de casos forenses. Para tanto, buscou-se identificar a diversidade genética desses polimorfismos nessa população, a segregação dos alelos entre os genitores (pai e mãe) para as suas respectivas filhas e a eficiência desse painel na amplificação de DNA extraído de amostras ósseas. Para identificar a diversidade genética, foram analisados os perfis genotípicos de 500 indivíduos não aparentados nascidos no estado de São Paulo. Todos os marcadores mostraram-se polimórficos para a população, sendo MID3701 o que apresentou maior diversidade e somente MID2637 se mostrou pouco informativo. O marcador MID1361 apresentou-se em desequilíbrio de Hardy-Weinberg e uma variante alélica foi identificada em seu alelo curto. O painel demonstrou alta eficiência forense, confirmado pelo poder acumulado de discriminação (0,999999999993 em mulheres e 0,99999993 em homens) e pelo valor acumulado da chance média de exclusão (0,999996 em trios e 0,9995 em duos). No comparativo com outras populações, valores significativos da distância genética foram obtidos, verificando-se que São Paulo está mais próximo à três departamentos colombianos e às populações européias. A proporção de ancestralidade identificada foi 41,8% para europeus, 31,6% para africanos e 26,6% para nativo-americanos. Na análise de segregação realizada em 101 trios, o padrão de transmissão esperado entre pai-mãe/filha foi o observado, o que confirma a baixa taxa de mutação desses marcadores. Por fim, os 32 X-InDels apresentaram as características necessárias para a análise de amostras biológicas em baixa concentração e/ou degradadas, mas algumas dificuldades na amplificação podem ser encontradas a depender das condições ambientais a que as amostras foram expostas. Conclui-se que o conhecimento acerca dos marcadores de inserção/deleção no cromossomo X pode ser ampliado, uma vez que na literatura ainda há pouco material disponível sobre o assunto; entretanto os dados deste trabalho já demonstram seu potencial como método alternativo para a análise de amostras forenses, pois foram identificados elevados valores de poder de discriminação e exclusão, baixa taxa de mutação e um elevado potencial de amplificação de amostras biológicas degradadas.

**Palavras-chave:** Genética forense. Identificação humana. Teste de paternidade. Polimorfismos de inserção deleção. Indel. Cromossomo X. X-Indel. São Paulo. Brasil.

MARTINEZ, J. **X chromosome insertion/deletion polymorphisms: analysis of 32 markers in São Paulo state population (Brazil)**. 2017. 107p. Thesis (Doctoral) – Faculdade de Ciências Farmacêuticas, Universidade de Estadual Paulista Júlio de Mesquita Filho, Araraquara.

#### ABSTRACT

In forensic genetics routine, the exclusive use of STRs (Short Tandem Repeats) markers when the biological sample is degraded can generate an inconclusive final statistical result, making essential the analysis of additional markers for case resolution. Used as a complementary method, InDels (insertion/deletion) polymorphisms have shown great potential to overcome the limitations of traditional markers. Polymorphisms in the X chromosome is also gaining significant importance in these studies, especially in those cases in which the analysis of the autosomal markers is not enough. In this perspective, this study aimed to characterize the São Paulo state population for 32 X chromosome insertion/deletion polymorphisms (32 X-InDels) and to evaluate the utility of this multiplex in the resolution of forensic cases. Therefore, it was analyzed the genetic diversity of this population, the alleles segregation between the parents and their respective daughters, and the amplification efficiency of this panel in DNA extracted from human bones. To identify genetic diversity, the genotypic profiles of 500 unrelated individuals born in São Paulo state was analysed. All markers were polymorphic for the population, with MID3701 being the most diverse, and MID2637 the less informative. The MID1361 marker was in Hardy-Weinberg disequilibrium and an allelic variant was identified in its short allele. The panel showed high forensic efficiency, confirmed by the accumulated power of discrimination (0.9999999999993 in females and 0.99999993 in males) and by the accumulated mean exclusion chance (0.999996 in trios and 0.99995 in duos). Comparing with other populations, significant values of genetic distance were obtained and São Paulo is closer to three Colombian departments and to European populations. The ethnic contributions identified 41.8% of Europeans, 31.6% of Africans and 26.6% of Native Americans admixture. In segregation analysis performed in 101 trios, the expected transmission pattern between parent/daughter was observed, which confirms the low mutation rate of these markers. Finally, the 32 X-InDels presented the necessary characteristics for the analysis of degraded biological samples, but some amplification difficulties can be found depending on the environmental conditions in which the samples were exposed. It can be concluded that knowledge about the X chromosome insertion/deletion markers can be expanded, because there is still little information available in the literature; meanwhile data from this work demonstrate its potential as an alternative method for the analysis of forensic samples.

**Keywords:** Forensic genetics. Human identification. Kinship testing. Insertion deletion polymorphism. Indel. X chromosome. X-Indel. São Paulo. Brazil.

**LISTA DE FIGURAS**

|            |  |    |
|------------|--|----|
| Figura 1.  | Esquema comparando os produtos de PCR e o processo de amplificação de marcadores STR, SNP e InDel. Fonte: adaptado de Pereira e Gusmão, 2012.....  | 24 |
| Figura 2.  | Representação esquemática dos cromossomos sexuais X e Y. Fonte: adaptado de Wijchers; Festenstein, 2011.....   | 28 |
| Figura 3.  | Distribuição dos 30 marcadores X-STRs de uso forense. Fonte: Szibor et al., 2007.....  | 32 |
| Figura 4.  | Ideograma do cromossomo X indicando a localização dos 32 marcadores X-InDel selecionados. Fonte: Pereira et al., 2012.....   | 45 |
| Figura 5.  | Eletroferograma de uma amostra feminina amplificada com o multiplex 32 X-InDels. Fonte: Dados da pesquisa.....   | 52 |
| Figura 6.  | Eletroferograma parcial de três amostras femininas obtidas com o multiplex 32 X-InDels. Em destaque a variante alélica do alelo curto do marcador MID1361. Fonte: Dados da pesquisa.....   | 54 |
| Figura 7.  | Eletroferograma parcial de dois casos de análise de segregação obtidos com o multiplex 32 X-InDels. Em destaque a variante alélica do alelo curto do marcador MID1361. Fonte: Dados da pesquisa.....   | 56 |
| Figura 8.  | SNPs identificados nas regiões dos <i>primers</i> de amplificação do marcador MID1361. Fonte: SNPdb.....   | 60 |
| Figura 9.  | Gráfico MDS bidimensional elaborado a partir da distância genética ( $F_{ST}$ ) entre Portugal, Espanha (Maiorca), Dinamarca, Angola e Moçambique, Somália, Groelândia, diversas populações da Colômbia, Nativo-americano, Macau, Bangladesh e Taiwan. Fonte: Dados da pesquisa.....             | 68 |
| Figura 10. | Representação esquemática da mistura estimada para a população de São Paulo (SP) utilizando 32 marcadores X-InDels obtidos através da análise do STRUCTURE baseado na contribuição da população africana (AFR), europeia (EUR) e nativo-americana (NAM) ( $K=3$ ). Fonte: Dados da pesquisa..... | 70 |
| Figura 11. | Eletroferograma dos marcadores 32 X-InDels obtidos a partir de uma amostra referência fermina 9947A (Promega, Wisconsin, EUA). Fonte: Dados da pesquisa.....   | 74 |
| Figura 12. | Eletroferograma dos marcadores 32 X-InDels obtidos a partir de uma amostra referência feminina 9947A (Promega, Wisconsin, EUA). Fonte: Dados da pesquisa.....  | 75 |
| Figura 13. | Eletroferograma dos marcadores 32 X-InDels obtidos a partir de DNA extraído de amostra óssea OS_1.1. Fonte: Dados da pesquisa.....   | 82 |

|            |  |    |
|------------|--|----|
| Figura 14. | Eletroferograma dos marcadores 32 X-InDels obtidos a partir de DNA extraído de amostra óssea OS_1.2. Fonte: Dados da pesquisa.....     | 83 |
| Figura 15. | Eletroferograma dos marcadores 32 X-InDels obtidos a partir de DNA extraído de amostra óssea OS_2.1. Fonte: Dados da pesquisa.....     | 84 |
| Figura 16. | Eletroferograma dos marcadores 32 X-InDels obtidos a partir de DNA extraído de amostra óssea OS_2.2. Fonte: Dados da pesquisa.....     | 85 |
| Figura 17. | Eletroferograma dos marcadores 32 X-InDels obtidos a partir de DNA extraído de amostra óssea OS_3. Fonte: Dados da pesquisa.....       | 86 |
| Figura 18. | Eletroferograma dos marcadores 32 X-InDels obtidos a partir de DNA extraído de amostra óssea OS_4. Fonte: Dados da pesquisa.....       | 87 |
| Figura 19. | Eletroferograma dos marcadores 32 X-InDels obtidos a partir de DNA extraído de amostra óssea OS_5-dente. Fonte: Dados da pesquisa..... | 88 |
| Figura 20. | Eletroferograma dos marcadores 32 X-InDels obtidos a partir de DNA extraído de amostra óssea OS_7. Fonte: Dados da pesquisa.....       | 89 |

**LISTA DE TABELAS**

|            |   |    |
|------------|---|----|
| Tabela 1.  | Frequência alélica e diversidade gênica (DG) dos marcadores 32 X-InDels na população de São Paulo. Número de cromossomos analisados: 782. Fonte: Dados da pesquisa.....                             | 57 |
| Tabela 2.  | Parâmetros de diversidade genética e probabilidade de HWE para os marcadores 32 X-InDels na população feminina de São Paulo. Tamanho da amostra: 282. Fonte: Dados da pesquisa.....                 | 59 |
| Tabela 3.  | Desequilíbrio de ligação dos pares de marcadores analisados na população masculina de São Paulo que obtiveram valor de $p \leq 0,05$ . Tamanho da amostra: 218. Fonte: Dados da pesquisa.....       | 61 |
| Tabela 4.  | Frequência haplotípica observada na população masculina de São Paulo para os marcadores em desequilíbrio de ligação. Tamanho da amostra: 218. Fonte: Dados da pesquisa.....                         | 64 |
| Tabela 5.  | Parâmetros estatísticos de interesse forense para os marcadores 32 X-InDels na população de São Paulo. Número de cromossomos analisados: 782. Fonte: Dados da pesquisa.....                         | 65 |
| Tabela 6.  | Amostras usadas na comparação entre as populações, número de cromossomos analisados (N) e respectivas referências.....  | 67 |
| Tabela 7.  | Distância genética ( $F_{ST}$ ) entre as populações (abaixo da diagonal) e o valor de $p$ correspondente (acima da diagonal). Fonte: Dados da pesquisa.....   | 71 |
| Tabela 8.  | Amplificação dos marcadores 32 X-InDels em amostra comercial 9947A (Promega, Wisconsin, EUA) diluída para diferentes concentrações. Fonte: Dados da pesquisa.....                                   | 73 |
| Tabela 9.  | Informações relacionadas aos casos de investigação de paternidade <i>post-mortem</i> analisados no Laboratório de Investigação de Paternidade da UNESP de Araraquara. Fonte: Dados da pesquisa..... | 77 |
| Tabela 10. | Concentração de DNA utilizada na reação de amplificação dos marcadores 32 X-InDels e número de marcadores amplificados para cada amostra. Fonte: Dados da pesquisa.....                             | 78 |

**LISTA DE ABREVIATURAS E SIGLAS**

|                    |  |
|--------------------|--|
| AMOVA              | <i>Analysis of molecular variance</i> , análise de variância molecular                 |
| AS-InDel           | Inserção/deleção autossômico   |
| AS-STR             | <i>Short Tandem Repeat</i> autossômico   |
| DNA                | <i>Deoxyribonucleic acid</i> , ácido desoxirribonucleico                               |
| dbSNP              | <i>Single Nucleotide Polymorphism database</i>   |
| DG                 | Diversidade gênica   |
| EUA                | Estados Unidos da América  |
| GHEP-ISFG          | Grupo de Línguas Espanhola e Portuguesa da Sociedade Internacional de Genética Forense |
| HID                | <i>Human Identification</i>  |
| InDel              | Inserção/deleção   |
| LD                 | <i>Linkage Disequilibrium</i> , desequilíbrio de ligação                               |
| MID                | <i>Marshfield identification</i>   |
| MCMC               | <i>Markov chain Monte Carlo</i> , cadeia de Markov Monte Carlo                         |
| MDS                | <i>Multidimensional scaling</i> , escalonamento multidimensional                       |
| MEC                | <i>Mean exclusion chance</i> , chance média de exclusão                                |
| MEC <sub>T</sub>   | Chance média de exclusão em trios envolvendo filha                                     |
| MEC <sub>D</sub>   | Chance média de exclusão em duo pai/filha  |
| MEC <sub>PGM</sub> | Chance média de exclusão em casos deficientes envolvendo mãe, filha e avó paterna      |
| MSY                | <i>Male-specific region of the Y</i>   |
| PCR                | <i>Polymerase Chain Reaction</i> , Reação em Cadeia da Polimerase                      |
| PAR                | <i>Pseudoautosomal region</i> , região pseudoautossômica                               |
| PD                 | Poder de discriminação   |
| PD <sub>F</sub>    | Poder de discriminação em mulheres   |
| PD <sub>M</sub>    | Poder de discriminação em homens   |
| RMSP               | Região Metropolitana de São Paulo  |

## Lista de Abreviaturas e Siglas

---

|         |   |
|---------|---|
| RFU     | <i>Relative fluorescence unit</i> , unidade relativa de fluorescência |
| SNP     | <i>Single Nucleotide Polymorphism</i>                                 |
| STR     | <i>Short Tandem Repeat</i>  |
| SRY     | <i>Sex-determining region Y</i>                                       |
| SBE     | Reação de <i>single-base extension</i>                                |
| UNESP   | Universidade Estadual Paulista Júlio Mesquita Filho                   |
| X-InDel | Inserção/deleção no cromossomo X                                      |
| X-STR   | <i>Short Tandem Repeat</i> no cromossomo X                            |

**LISTA DE SÍMBOLOS**

|               |   |
|---------------|---|
| ~             | Aproximadamente                                   |
| %             | Porcento  |
| ng            | Nanogramas  |
| n°            | Número  |
| °C            | Graus Célsius                                     |
| ®             | Marca registrada                                  |
| ml            | Mililitros  |
| µL            | Microlitros                                       |
| cm            | Centímetros                                       |
| mg            | Miligramas  |
| D'            | Valor normalizado do desequilíbrio de ligação     |
| <i>p</i> -LD  | Valor de <i>p</i> do desequilíbrio de ligação     |
| <i>p</i> -HWE | Valor de <i>p</i> 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

|   |            |
|---|------------|
| <b>1. Introdução.....</b>   | <b>17</b>  |
| <b>2. Revisão da literatura.....</b>  | <b>19</b>  |
| 2.1 – Marcadores genéticos para identificação humana.....   | 20         |
| 2.2 – Cromossomo X.....   | 26         |
| 2.3 – Colonização do estado de São Paulo.....   | 35         |
| 2.4 – Justificativa e hipótese estudo.....  | 38         |
| <b>3. Objetivos.....</b>  | <b>39</b>  |
| <b>4. Materiais e Métodos.....</b>  | <b>41</b>  |
| 4.1 – Aspectos éticos da pesquisa.....  | 42         |
| 4.2 – Sujeitos da pesquisa.....   | 42         |
| 4.3 – Coleta das amostras biológicas e extração do DNA.....   | 43         |
| 4.4 – Análise de concentração das amostras ósseas.....  | 44         |
| 4.5 – Amplificação do DNA e genótipo dos X-InDels.....  | 44         |
| 4.6 – Forma de análise dos resultados e parâmetros estatísticos.....  | 48         |
| <b>5. Resultados e Discussão.....</b>   | <b>50</b>  |
| 5.1 – Padronização e otimização dos materiais e métodos.....  | 51         |
| 5.2 – Variante alélica identificada na população de São Paulo.....  | 53         |
| 5.3 – Variabilidade genética da população de São Paulo.....   | 57         |
| 5.4 – Desequilíbrio de ligação entre os marcadores X-InDels.....  | 61         |
| 5.5 – Parâmetros estatísticos de eficiência forense.....  | 64         |
| 5.6 – Comparação genética entre as populações e análise de ancestralidade.....                                  | 66         |
| 5.7 – Análise de segregação dos marcadores X-InDels.....  | 72         |
| 5.8 – Amplificação dos marcadores X-InDels em amostras biológicas em baixa<br>concentração e/ou degradadas..... | 72         |
| <b>6. Conclusões.....</b>   | <b>90</b>  |
| <b>7. Referências.....</b>  | <b>92</b>  |
| <b>Anexo I – Aprovação do Comitê de Ética em Pesquisa com Seres Humanos.....</b>                                | <b>100</b> |

# Introdução

A análise do ácido desoxirribonucleico (DNA) consiste em uma poderosa metodologia amplamente aceita por toda comunidade forense. Pode ser utilizada para auxiliar na identificação humana, especialmente em casos envolvendo vítimas de desastres em massa, pessoas desaparecidas e agressores de crimes violentos, assim como na investigação de paternidade, tendo como principal objetivo a comparação genética entre dois indivíduos ou entre um determinado material biológico e um indivíduo (THOMPSON; ZOPPIS; MCCORD, 2012; ZIĘTKIEWICZ et al., 2012).

A sequência de DNA do ser humano é extremamente semelhante entre as pessoas ao redor do mundo (~99,7%) e a pequena fração nessa sequência que é variável é o que nos torna únicos. Quando em um determinado locus há uma variação entre os alelos identificados em indivíduos de uma mesma população (com pelo menos dois alelos de frequência igual ou superior a 1%), essa variação é conhecida como polimorfismos genético (KARKI et al., 2015). A análise dos polimorfismos genéticos é o que possibilita que o DNA seja utilizado na identificação humana e em testes de paternidade (BUTLER, 2005; HOLMES; RUSSELL; WALLEY, 2003).

Nessa análise são considerados marcadores genéticos os polimorfismos herdáveis identificados em diversas populações (DAVEY et al., 2011). Entre os possíveis marcadores, os 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 disponível pode estar degradado e/ou em baixa quantidade, dificultando a solução do caso com a investigação apenas dos marcadores nucleares autossômicos. Nesse contexto, a análise de outros marcadores e de outras regiões polimórficas pode complementar a metodologia tradicional de forma eficaz (ALVAREZ-CUBERO et al., 2012).

# Revisão da Literatura

## 2.1 – Marcadores genéticos para identificação humana

### 2.1.1 – A história do DNA e da genética forense

Ao descrever pela primeira vez a estrutura da molécula de DNA, Watson e Crick (1953) provavelmente não poderiam imaginar os desdobramentos científicos que ocorreriam com o seu achado. A análise do DNA atualmente é utilizada em diferentes áreas de aplicação, tais como a medicina genética, a indústria alimentícia e em situações forenses como a resolução de crimes e a identificação de relações de parentesco (TILLMAR, 2010).

A primeira vez que se percebeu que a sequência de DNA poderia ser utilizada na identificação dos indivíduos foi em 1984, quando o geneticista inglês Alec Jeffreys, testando sondas de hibridização para detectar múltiplos minissatélites do DNA humano, percebeu que o padrão de repetição desses minissatélites variava de um indivíduo para o outro, produzindo um perfil genético específico e único, que ficou conhecido como *DNA fingerprinting* ou impressão digital do DNA, em português (JEFFREYS; WILSON; THEIN, 1985).

Devido ao seu potencial na discriminação dos indivíduos, em 1986 a impressão digital do DNA foi utilizada pela primeira vez em um caso criminal na Inglaterra, em que se comparou o perfil genético de um indivíduo acusado de estupro e assassinato com a amostra de sêmen encontrada na vítima, contribuindo com a identificação do verdadeiro autor do crime (GILL; JEFFREYS; WERRETT, 1985).

As investigações realizadas no caso de O.J. Simpson também consistem em um bom exemplo acerca do uso do DNA na resolução de crimes. Em nenhum outro julgamento os resultados dos exames de DNA foram tão intensamente examinados, discutidos e questionados pela defesa (NATIONAL RESEARCH COUNCIL, 1996), na qual ficou evidente a importância da adequada execução da cadeia de custódia para que as provas periciais permaneçam robustas e confiáveis.

Assim, a utilização do DNA na rotina das análises forenses busca responder questões como: o suposto pai é o pai biológico do filho (a)? A amostra encontrada na cena do crime pertence a um determinado suspeito?

No caso da primeira pergunta o vínculo biológico é determinado pelo índice de paternidade calculado através da razão de verossimilhança (ou *likelihood ratio*) com base em duas hipóteses. Essa análise consiste no contraste de hipóteses mutuamente exclusivas, ao qual se identifica a probabilidade de evidência genéticas assumindo que o suposto pai é o pai biológico do filho (a) versus a probabilidade do mesmo resultado assumindo que pai biológico é qualquer homem ao acaso na população (GJERTSON et al., 2007).

Já a resposta para a segunda pergunta é baseada na probabilidade de se identificar, na população, perfis genéticos idênticos de indivíduos não aparentados, sendo que o cálculo dessa probabilidade depende de estudos genéticos populacionais, no qual se identificam frequências alélicas, poder de discriminação e dependência entre os marcadores genéticos (NATIONAL RESEARCH COUNCIL, 1996).

### **2.1.2 – Polimorfismos do DNA como marcadores genéticos**

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

Polimorfismos de sequência consistem na substituição de um ou mais nucleotídeos na sequência do DNA. Dentre esses polimorfismos, os *Single Nucleotide Polymorphisms* (SNPs) são a forma mais frequente de variação genética (BUTLER, 2005).

Os polimorfismos de comprimento são originados pela inserção ou deleção de um ou mais nucleotídeo ao longo da sequência, resultando em variações no comprimento dessa. Nesse grupo, além das inserções ou deleções de alguns pares de bases (pb), também deve-se destacar as repetições consecutivas, como os *Short Tandem Repeats* (STRs) (BUTLER, 2005).

Todos esses polimorfismos estão amplamente espalhados ao longo do genoma humano e apresentam características que os tornam poderosas ferramentas na resolução de casos criminais e na análise de relações de parentesco (TILLMAR, 2010).

#### **2.1.2.1 – *Short Tandem Repeats* (STRs)**

Os marcadores STRs, também conhecidos como microssatélites, são constituídos de unidades de 2 a 7 nucleotídeos (por exemplo, GATA) que se repetem consecutivamente (ou em tandem) por dezenas de vezes ao longo do genoma (BUTLER, 2007).

A utilização de marcadores STRs é amplamente aceita na prática forense. Na maioria dos casos, a análise de menos de 20 STRs autossômicos é suficiente para a identificação humana ou para a resolução de casos de trio (suposto pai/mãe/filho[a]) na investigação de paternidade (GE; EISENBERG; BUDOWLE, 2012).

Eles são predominantemente multialélicos e apresentam elevada taxa de mutação, o que os torna marcadores altamente informativos na discriminação entre os indivíduos. Além disso, podem ser facilmente amplificados em uma Reação em Cadeia da Polimerase (PCR) em multiplex e apresentam um protocolo de rápida e fácil execução técnica, envolvendo poucos passos entre a amplificação e a análise por eletroforese capilar. Em outras palavras, esses marcadores apresentam boas características para uso forense (PEREIRA, 2011; TILLMAR, 2010; AMBRÓSIO; BRAGANHOLI; CICARELLI, 2015).

Entretanto, a alta taxa de mutação frequentemente gera mutações espontâneas entre os indivíduos ao longo das gerações, o que pode dificultar a conclusão em casos complexos de reconstrução para a identificação de parentesco (PEREIRA, 2011; PINTO et al., 2013a).

Além disso, na genotipagem de STRs, os fragmentos de a serem analisados possuem longo comprimento (100-400pb para os STRs-padrão e 80-250pb para os mini-STRs), o que diminui a possibilidade de amplificação completa do fragmento em amostras de DNA degradado ou em baixa concentração (PINTO et al., 2013a).

A exposição ao meio ambiente, através do contato com a água, oxigênio, radiação ultravioleta e enzimas nucleases, degrada as moléculas de DNA, dividindo-o aleatoriamente em pequenos pedaços (BUTLER, 2011). Dixon e colaboradores (2005) sugerem que, durante a morte celular por apoptose ou necrose, o nucleossomo pode oferecer proteção aos 146pb que estão enovelados às histonas. Portanto, para a análise de material biológico degradado, polimorfismos analisados em amplicons de tamanho reduzido podem apresentar maior sucesso na amplificação completa do fragmento.

Assim, quando a análise de microssatélites não é suficiente para a conclusão, a prática comum consiste em ampliar o número de marcadores analisados. Nesse contexto, há autores que sugerem os marcadores bialélicos como complementar, tais como os polimorfismos de base única (SNPs) ou de inserção/deleção (InDels) (PINTO et al., 2013a).

#### **2.1.2.2 – *Single Nucleotide Polymorphisms (SNPs)***

Os SNPs consistem na variação de uma única base na sequência do DNA em um ponto específico. Normalmente apresentam-se como marcadores bialélicos (referência ou alternativo) e, apesar da baixa taxa de mutação, são altamente abundantes no genoma, existindo milhões de alelos alternativos em cada indivíduo (BUTLER, 2005).

Essa baixa taxa de mutação torna os SNPs uma interessante ferramenta nas investigações de parentesco. Outra vantagem é que esses polimorfismos podem ser analisados

através de fragmentos (ou amplicons) curtos (menos de 100pb), o que aumenta as chances de sucesso de amplificação em amostra de DNA degradado (BUTLER, 2005).

A análise desses marcadores pode ser facilmente automatizada e realizada em multiplex, na qual mais de um polimorfismo é analisado concomitantemente. Entretanto, uma desvantagem significativa é que seu protocolo de amplificação e análise é complexo, envolvendo passos de purificação e reação de *single-base extension* (SBE), o que aumenta significativamente seu tempo de execução e custo (PEREIRA; GUSMÃO, 2012).

### 2.1.2.3 – *Insertion/Deletion Polymorphisms (InDel)*

Polimorfismos InDel consistem na inserção ou na deleção de um ou mais nucleotídeos no genoma. Os primeiros estudos relacionados a esses marcadores ocorreram em 2002, quando Weber e colaboradores (2002) identificaram e caracterizaram 2.000 InDels com diferentes variações de comprimento no genoma humano e calcularam sua frequência alélica em europeus, africanos, japoneses e nativo-americanos.

Associado a esse trabalho, a *Marshfiel Clinic*<sup>1</sup> criou um banco de dados online de polimorfismos de inserção/deleção, no qual estão disponíveis informações relacionadas ao tipo de polimorfismos (bialélico ou multialélico), diferenças no comprimento dos alelos, *primers* de amplificação, posição no cromossomo, além das frequências alélicas em diferentes populações (PEREIRA; GUSMÃO, 2012). Informações sobre os InDels também foram incluídas no *Single Nucleotide Polymorphism database*<sup>2</sup> (dbSNP).

Em 2006, Mills e colaboradores (2006) iniciaram um mapa da variação de InDels no genoma, o qual continha mais de 415.000 polimorfismos descritos. Mais recentemente com o projeto 1000 Genomas foram descritos mais de 3,6 milhões desses polimorfismos (THE 1000 GENOMES PROJECT CONSORTIUM, 2015).

Utilizados como complementação à análise dos STRs, os InDels também podem ser analisados em multiplex e superam as limitações da análise tradicional em diversos casos de investigação de paternidade, pois apresentam baixa taxa de mutação, o que favorece verificar o vínculo biológico entre dois indivíduos (TORRES et al., 2014; AMBRÓSIO; BRAGANHOLI; CICARELLI, 2015).

---

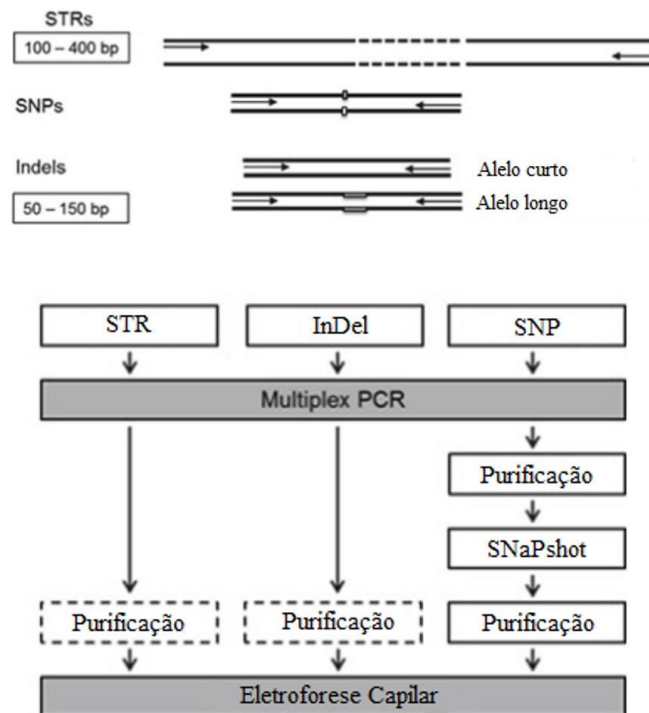
<sup>1</sup> Endereço eletrônico: <http://www.marshfieldclinic.org/mgs>

<sup>2</sup> Endereço eletrônico: <https://www.ncbi.nlm.nih.gov/snp/>



Além disso, as técnicas utilizadas na identificação humana para esses marcadores geram amplicons de tamanho reduzido (50-150pb), o que facilita a análise de material biológico comprometido (TORRES et al., 2014).

**Figura 1:** Esquema comparando os produtos de PCR e o processo de amplificação de marcadores STR, SNP e InDel.



**Fonte:** adaptado de Pereira e Gusmão, 2012.

Como é possível observar na figura 1, os InDels combinam os aspectos mais interessantes dos marcadores STRs e SNPs. Assim como os SNPs, os InDels geram amplicons de tamanho reduzido quando comparados com os STRs e mini-STRs. Além disso, assim como os STRs, possuem um simples e rápido processo de amplificação, o que reduz significativamente o tempo e custo da reação relativamente aos SNPs (PEREIRA; GUSMÃO, 2012).

Apesar de ser uma importante ferramenta de complementação das metodologias tradicionais, para a análise de InDels em cromossomos autossômicos existe atualmente apenas um kit comercial (DIPplex Kit® - Qiagen, Hilden, Alemanha), sendo sua eficácia na identificação humana já comprovada em diversas populações, inclusive no Brasil (PEPINSKI

et al., 2013; SAIZ et al., 2014; SEONG et al., 2014; TORRES et al., 2014; FERREIRA PALHA et al., 2015; WANG et al., 2016; TOMAS et al., 2016; INÁCIO et al., 2017).

A fim de aumentar a possibilidade de utilização de InDels na prática forense, estudos que buscam a padronização de novos marcadores vêm sendo elaborados. Pereira e colaboradores (2009; 2017) descreveram um novo multiplex para a identificação humana no qual é possível analisar 38 InDels autossômicos em uma única reação. Os amplicons gerados na PCR possuem no máximo 160pb 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 dos parâmetros de interesse forense desses marcadores em africanos, europeus e asiáticos mostrou sua eficiência na discriminação entre os indivíduos dessas populações.

Partindo desse multiplex de 38 InDels, Manta e colaboradores (2012) observaram um bom poder de discriminação desses marcadores em indivíduos da população do Rio de Janeiro e em nativo-americanos. Como ferramenta complementar na análise de material biológico degradado, os autores obtiveram sucesso na amplificação de DNA extraído de amostra óssea, o que contribuiu com a resolução de um caso em que apenas o uso de STRs foi insuficiente.

A análise do multiplex 38 InDels também é utilizada como ferramenta complementar no Laboratório de Investigação de Paternidade da Faculdade de Ciências Farmacêuticas da UNESP de Araraquara. O uso desses marcadores InDel, associado aos STRs autossômicos, já permitiu a resolução de 11 casos de investigação de paternidade *post-mortem*.

Em 10 casos de difícil resolução esse multiplex confirmou a exclusão da paternidade, sendo que em um deles nenhum marcador STR pôde ser amplificado. Em outra situação, a não amplificação de 4 STRs somada à incompatibilidade em 2 marcadores entre suposto pai e filho gerou um índice acumulado de paternidade menor que 10.000, o que não permitia a elaboração de um laudo conclusivo. Com a análise dos InDels, o índice subiu para 1.412.687,37, favorecendo a inclusão da paternidade e a solução do caso.

Além dos cromossomos autossômicos, também é possível a utilização de marcadores genéticos nos cromossomos sexuais (X e Y) e no DNA mitocondrial como análise complementar a dos marcadores tradicionais, sendo que a seleção de um ou outro irá depender das particularidades do caso.

## 2.2 – Cromossomo X

### 2.2.1 – Evolução e características dos cromossomos sexuais

Nos mamíferos, com algumas poucas exceções, o sexo é determinado geneticamente pelo sistema de cromossomos sexuais XY, no qual as fêmeas herdam um cromossomo X de cada um dos genitores (XX) e os machos herdam o cromossomo X materno e o Y paterno (XY).

A teoria consenso afirma que os cromossomos sexuais XY iniciaram seu processo evolutivo há mais de 300 milhões de anos a partir de um par de autossômicos homólogos (chamados de proto-cromossomos sexuais) que adquiriram um locus relacionado à determinação do sexo (OHNO, 1967; ELLEGREN, 2011).

Acredita-se que uma barreira de recombinação se desenvolveu entre esses proto-cromossomos sexuais, isolando as regiões determinantes do sexo (ROSS et al., 2005). Com o passar do tempo, a falta de recombinação entre os cromossomos X e Y, associada ao acúmulo de mutações, deleções e retrotransposons, levou rapidamente à degeneração de partes do Y, tornando os cromossomos sexuais heteromórficos (CHARLESWORTH; CHARLESWORTH; MARAIS, 2005; ELLEGREN, 2011).

O cromossomo Y possui aproximadamente 60 megabases (Mb) e contém alguns poucos genes. Apresenta uma elevada extensão de heterocromatina e um alto número de repetições simples sem função codificante, aparentemente, sem efeito fenotípico. A sequência completa da região eurocromática revelou 178 unidades transcritas, mas muitos são pseudogenes, sendo que o Y codifica apenas 45 proteínas únicas (SKALETSKY et al., 2003). A função mais importante desse cromossomo nos mamíferos é a determinação do sexo, na qual o gene SRY (do inglês *sex-determining region Y*) é o responsável pelo desenvolvimento das genitais masculinas (GRAVES, 2006).

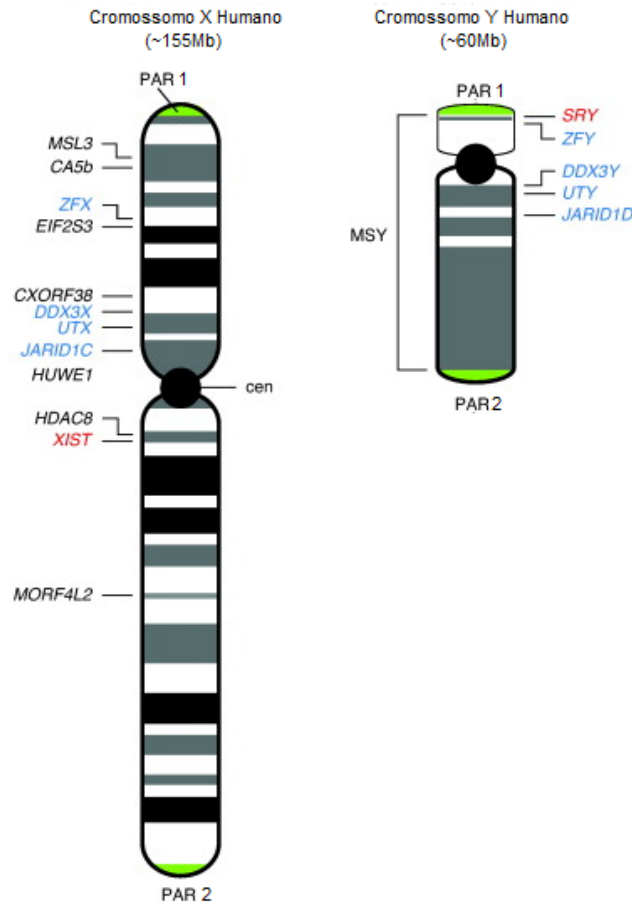
O cromossomo X é muito maior, com ~155Mb ao longo de sua sequência, sendo altamente rico em regiões repetitivas e com 2.516.025 variantes mapeadas. Apresenta 840 pseudogenes e 830 genes codificantes descritos, sendo em sua maioria de pequeno comprimento e com alta variedade de funções específicas e generalizadas (THE 1000 GENOMES PROJECT CONSORTIUM, 2015).

Os processos evolucionários erradicaram a maioria dos traços da relação ancestral entre os cromossomos X e Y, havendo poucos genes homólogos ativos em ambos os cromossomos. Mais precisamente Ross e colaboradores (2005) identificaram que apenas 54 genes descritos no cromossomo X apresentam seu homólogo funcional no Y.

Como consequência dessa degeneração, uma dosagem desigual de transcrições gênicas entre homens e mulheres poderia ocorrer. No entanto, para que se tornasse proporcional os produtos gênicos transcritos, um dos cromossomos X nas mulheres é silenciado de forma estável ao longo da vida (CHARLESWORTH, 1991; GRAVES, 2006).

A reativação desse cromossomo ocorre somente nas células germinativas XX, para que ocorra a recombinação entre eles. Nos homens a recombinação não ocorre ao longo de quase todo o comprimento do cromossomo X, sendo esse processo restrito às regiões pseudoautossômicas (PAR, *pseudoautosomal region*) 1 e 2, localizadas nas extremidades de ambos os cromossomos sexuais (HELENA MANGS; MORRIS, 2007).

No extremo do braço curto dos cromossomos X e Y está localizada a PAR1, com 2,7Mb. A homologia entre eles é mantida por uma recombinação obrigatória na meiose masculina. Assim, os genes nessa região estão presentes em duas cópias em homens e mulheres e não estão sujeitos à inativação no cromossomo X. No extremo do braço longo desses cromossomos localiza-se o PAR2, com 0,33Mb. A recombinação entre eles não ocorre em todas as meioses e alguns dos genes dessa região estão sujeitos à inativação no cromossomo X (ROSS et al., 2005). A região intermediária entre as PARs no Y é especificamente masculina, corresponde a 95% do cromossomo e não pode ser recombinada com o X, sendo chamada de MSY (do inglês *male-specific region of the Y*) (Figura 2) (WIJCHERS; FESTENSTEIN, 2011).

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

Nota: Observe a diferença de tamanho entre os cromossomos. Em vermelho estão representados os genes específicos da determinação do sexo; em azul os genes que possuem um par homólogo no Y e que não estão sujeitos à inativação no X; em preto genes que não possuem um homólogo no Y e que não estão sujeitos à inativação no X; em verde as regiões pseudoautossomais (PAR); MSY (*male-specific region of the Y*).

**Fonte:** adaptado de Wijchers; Festenstein, 2011.

Devido ao seu processo evolutivo e às características descritas, os marcadores genéticos no cromossomo X têm potencial de apresentar parâmetros forenses eficientes em casos de investigação complexa de parentesco (SZIBOR, 2007; PEREIRA et al., 2012). Por esse motivo, esse cromossomo vem ganhando significativa importância em estudos de genética forense nos últimos anos.

### 2.2.2 – Aplicação dos marcadores genéticos do cromossomo X

O padrão de herança do cromossomo X o torna altamente interessante para os estudos de genética médica, de populações e ancestralidade. Esse cromossomo se desloca entre ambos os sexos a cada geração, o que possibilita investigar o padrão migratório de homens e mulheres ao longo da história (SCHAFFNER, 2004; PEREIRA et al., 2012).

O reduzido tamanho efetivo da população ( $N_e$ ) (três quartos dos cromossomos autossômicos) torna o cromossomo X mais sensível aos efeitos da deriva genética e da subestrutura de populações, o que justifica a redução da sua diversidade gênica (SCHAFFNER, 2004; PEREIRA et al., 2012).

Além disso, a recombinação somente nas mulheres aumenta os níveis de desequilíbrio de ligação e o estado de hemizigose nos homens permite a identificação direta de haplótipos (SCHAFFNER, 2004; PEREIRA et al., 2012).

A combinação desses aspectos torna o cromossomo X uma fonte singular de informações sobre a variação genética das populações. Na genética médica o estudo de seus marcadores é especialmente importante para a compreensão de doenças e traços associados ao padrão de herança desse cromossomo, como por exemplo hemofilia e a distrofia muscular de Duchenne (SZIBOR, 2007). Na base de dados OMIM (*Online Mendelian Inheritance in Man*) (HAMOSH et al., 2005) encontram-se descritas, até o momento, 1.274 doenças ou traços com herança mendeliana associadas a esse cromossomo.

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 (SZIBOR et al., 2003).

Por outro lado, em casos de paternidade em duos envolvendo suposto pai e filha quando uma ou duas incompatibilidades são observadas e a hipótese de que essas sejam uma mutação é coerente, 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).

Os casos deficientes de paternidade, em que o material biológico do suposto pai não está disponível, são a maior possibilidade de contribuição do cromossomo X. Esse consiste em uma importante ferramenta na exclusão de paternidade entre duas supostas irmãs ou meias-irmãs. Como há apenas uma possibilidade de 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 em que é necessária a análise de ossadas ou de corpos em decomposição, a combinação de métodos sofisticados de análise do DNA é fundamental, pois a ação ambiental sobre esse material reduz a chance de amplificação do perfil genotípico completo.

Nesse contexto, a análise do cromossomo X tende a ser mais eficiente do que a dos autossômicos. Como os seus marcadores apresentam maior MEC, em situações em que um número menor de marcadores é analisado, pode-se ainda haver poder estatístico suficiente para a análise do caso (SZIBOR et al., 2003; SZIBOR, 2007).

Já quando precisamos identificar a proveniência de um vestígio biológico, ou seja, identificar se determinado vestígio pertence a um indivíduo, os marcadores no cromossomo X tendem a ser igualmente informativos aos autossômicos, independente de se tratar de uma identificação de um homem ou de uma mulher (SZIBOR et al., 2003).

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 homocigótica para todos os loci (SZIBOR et al., 2003).

### 2.2.3 – Marcadores genéticos do cromossomo X

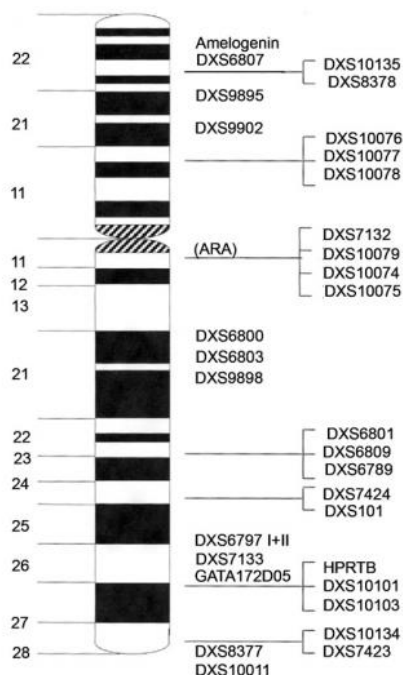
Devido ao seu alto poder de individualização e praticidade, a análise de STRs é amplamente utilizada na rotina das ciências forenses. Durante a era da tecnologia por PCR, diversos desses marcadores foram estabelecidos, sendo muitos deles localizados no cromossomo X. Entretanto, enquanto um número crescente de artigos tem sido publicado com informações desses marcadores em cromossomos autossômicos e Y, a aplicação forense do cromossomo X ganhou atenção somente nos últimos anos (SZIBOR et al., 2003; SZIBOR, 2007).

O cromossomo X apresenta diversos tipos de polimorfismos, que não se diferenciam dos descritos nos demais. Nele também são encontradas 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 heterozigidade 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 deficientes de paternidade em que amostra da suposta avó está sendo analisada. 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 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 (Figura 3).



**Figura 3:** Distribuição dos 30 marcadores X-STRs de uso forense.

**Fonte:** Szibor et al., 2007.

É importante destacar que os marcadores X-STR utilizados na identificação humana e testes de paternidade estão localizados em regiões não recombinantes entre os cromossomos sexuais X e Y, sendo a recombinação desses rescrita às mulheres, de modo a garantir a sua diversidade (MARTINS, 2011).

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-STR e permaneceu no mercado entre 2003 e 2008, mas não foi disponibilizado no Brasil. Em 2005, a empresa Biotype (Dresden, Alemanha) lança um novo kit, Mentype® Argus X-8, com a amplificação de 8 marcadores mais amelogenina, mas esse 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 agosto de 2017).

Em 2006, o Grupo de Línguas Espanhola e Portuguesa da Sociedade Internacional de Genética Forense (GHEP-ISFG) desenvolveu um sistema multiplex com a análise de 10 X-STRs *in house* (*decaplex*), na qual foram caracterizadas 15 populações ibéricas e latino americanas (GUSMÃO et al., 2009). Até essa data o Investigator Argus X-12 (Qiagen, Hilden,

Alemanha) ainda não existia e a elaboração do *decaplex* possibilitou a utilização de marcadores X-STR nas populações da América Latina (MARTINS, 2011). Além disso, o Investigator Argus X-12 (Qiagen, Hilden, Alemanha) trata-se de um painel com marcadores ligados, o que o torna interessante para análise das relações biológicas, sempre que seja necessária a reconstrução de haplótipos. Por outro lado, o decaplex (GUSMÃO et al., 2009) consiste em marcadores independentes, que são mais informativos nos estudos de identificação.

Estudos populacionais analisando marcadores X-STRs vêm sendo produzidos em diversas partes do mundo (MARTINS et al., 2010; NOTHNAGEL et al., 2012; DONG et al., 2014; ISRAR et al., 2014) e um banco de dados<sup>3</sup> foi desenvolvido pelo grupo de pesquisa da Prof<sup>a</sup>. Dr<sup>a</sup>. Regina Cicarelli com a finalidade de compilar os dados genéticos populacionais desses marcadores na população brasileira (MARTINS et al., 2014).

Entretanto, quando tratamos de marcadores de inserção/deleção para o cromossomo X pouco se tem publicado. Buscando a padronização de marcadores X-InDels para uso forense, um grupo brasileiro (FREITAS et al., 2010) desenvolveu uma metodologia utilizando 33 marcadores em uma única reação de PCR em multiplex com amplicons que variam de 100 a 300pb. Os autores demonstraram a possibilidade de aplicação da metodologia em casos de identificação humana e testes de paternidade em Belém-PA e concluíram que se tratam de marcadores robustos e eficientes, que podem facilmente ser adotados na rotina forense.

A eficiência do painel 33 X-InDels foi recentemente analisada na população argentina (CAPUTO et al., 2017) e os autores identificaram que 5 marcadores e 2 blocos haplotípicos apresentaram heterozigidade esperada abaixo de 20%, o que indica que esses são pouco informativos para essa população. Além disso, o poder de discriminação e a chance média de exclusão ficaram 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, que não consiste no principal contribuidor 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 forenses, em especial aqueles em que o DNA encontra-se degradado.

---

<sup>3</sup> Endereço eletrônico: [www.bgbx.com.br](http://www.bgbx.com.br)

---

Explorando ainda mais a potencialidade desses marcadores, Ibarra e colaboradores (2014) utilizaram o multiplex 32 X-InDels para ampliar o conhecimento atual sobre a base genética da população da Colômbia. Por meio da análise de distância genética e proporção de ancestralidade, esses marcadores possibilitaram a caracterização das misturas da população colombiana, apontando diferenças significativas na contribuição ancestral entre as diferentes regiões geográficas do país.

### 2.3 – Colonização do estado de São Paulo

O litoral paulista já havia sido visitado por navegadores portugueses e franceses no início do século XVI, mas sua colonização iniciou apenas em 1532, quando Martim Afonso de Souza fundou a Vila de São Vicente, localizada na atual Baixada Santista, marcando o início da fixação dos portugueses na costa brasileira (BUENO, 2009).

Nessa época, a principal atividade de subsistência no Brasil era a agricultura, e os índios que viviam na região foram mantidos como escravos para a implementação da lavoura de cana-de-açúcar (BUENO, 2009).

Aqui é importante destacar que não há registros de mulheres portuguesas no Brasil nessa época, sendo as índias as primeiras matriarcas dos paulistas, o que tornava a população predominantemente mestiça de lusos e índios (mameluco). Por essas características, durante os três primeiros séculos de colonização, o número de índios e mamelucos superou o de europeus (MOTA, 2003; SÃO PAULO, 2017).

O rigoroso frio e as dificuldades de acesso ao litoral paulista inviabilizaram o plantio da cana-de-açúcar na Vila de São Vicente. Além disso, notícias sobre mineração de ouro e ferro no sertão brasileiro desviaram o foco dos colonos da agricultura. Assim, no final do século XVI iniciaram-se as viagens de reconhecimento do interior do país, conhecidas como “entradas” ou “bandeiras” dos sertanistas, em busca de metais e indígenas escravos (BUENO, 2009). Ao contrário do que se imagina, esse movimento de interiorização não foi exercido prioritariamente por europeus. Segundo relatos, numa bandeira de quatro mil homens, menos de mil eram brancos ou mestiços, sendo os restantes Tupi-Guarani (MOTA, 2003).

Ao longo do século XVI e XVII, as atividades econômicas dos colonos da região assessoravam-se por uma base sólida de escravos indígenas, que foram aprisionados ao longo das expedições pelo sertão. Nessa época, quase não se encontram escravos africanos devido ao seu alto preço e à pobreza da região, sendo que o termo “negros da terra” utilizado na época referia-se, na verdade, aos índios, que não eram objetos de compra e venda, mas somente de aprisionamento (MONTEIRO, 1994).

Já no interior do estado, a economia açucareira ganhou grande importância, levando a ocupação do solo até a metade do território atual. Foi nesse período, especialmente no final do século XVIII, que a crise da escravidão indígena obrigou São Paulo a intensificar a importação de escravos africanos, transformando os tijupares em senzalas (MONTEIRO, 1994).

O final do período colonial e começo do império, a partir de 1882, é marcado pela substituição da cana-de-açúcar pelas plantações de café, que se mostraram mais promissoras e

deram ao estado destaque no cenário nacional devido à produtividade de suas terras roxas (MOTA, 2003).

A economia cafeeira propiciou a construção de estradas de ferro e de rodagem, expandindo a ligação São Paulo-Santos para o interior paulista. Essa nova conjuntura alterou o mapa colonial e acelerou significativamente a imigração em massa de italianos, espanhóis e, mais tarde, japoneses que, junto aos portugueses, alteraram a estruturação do interior paulista (MOTA, 2003).

A abolição da escravatura (1888), o início do período republicano (1889) e o surto de industrialização trouxeram a modernização ao estado. Além da liderança econômica paulista, o início do século XX foi marcado pela elevação da concentração populacional. Em 1900, o estado representava 12,2% da população nacional, alcançando, em 1970, 19,1% e chegando à 21,8% em 2000 (DEDECCA et al., 2010).

A imigração de estrangeiros foi um grande contribuinte para esse aumento, especialmente de italianos, sírios, judeus e libaneses, com dados que indicam que entre 1908 e 1920 chegaram em Santos 190 mil imigrantes subvencionados e 340 mil espontâneos (MOTA, 2003).

Até 1920, 84% de sua população ainda se concentrava no interior do estado. Esse quadro começou a se modificar com a instalação de indústrias na capital, que culminou no aumento significativo da imigração interestadual a partir dos anos 40, especialmente de Minas Gerais e da região Nordeste (MONTALI et al., 2010).

A expansão da Região Metropolitana de São Paulo (RMSP) superou as taxas nacionais, com crescimento populacional anual de 5,5% na década de 40, 5,9% nos anos 50 e 5,4% na década de 60 (MONTALI et al., 2010). Durante a década de 70, São Paulo recebeu um total de 3.325.468 imigrantes provenientes de outros estados. Desses, 66% dirigiram-se para a região metropolitana e 34% para o interior. A partir das décadas seguintes é possível observar uma redução no número de imigrantes interestaduais, porém um aumento no fluxo migratório para o interior paulista (CUNHA et al., 2000).

O interior do estado reforçou nos anos 90 seu potencial de absorção migratória, o que culminou em um saldo migratório positivo de 850 mil pessoas nos anos 80, para 1,1 milhão nos anos 90 (DEDECCA et al., 2010). Nesse processo a RMSP passou a assistir uma desaceleração de seu crescimento populacional e, em contrapartida, as cidades médias do interior assistiram à aceleração de seu crescimento (MONTALI et al., 2010).

O estado de São Paulo conta atualmente com 645 municípios, divididos em 15 regiões administrativas e em 6 regiões metropolitanas, sendo essas a de São Paulo, Campinas, Baixada Santista, Vale do Paraíba e Litoral Norte, Sorocaba e Ribeirão Preto. Representando 21,6% de toda população nacional, abriga mais de seis milhões de imigrantes interestaduais e estrangeiros de mais de 70 nacionalidades, e leva o título de estado mais populoso do Brasil, sendo que metade de seus habitantes está centralizada na RMSP (IGBE, 2010; SÃO PAULO, 2017).

Logo, devido a seu contexto histórico, a população do estado de São Paulo é considerada a maior mistura multicultural e étnica do país, abrangendo principalmente descendentes de italianos e portugueses, mas também com forte influência de ameríndios, africanos, espanhóis, alemães, árabes e japoneses, além do alto número de imigrantes da região nordeste (ALMEIDA PRADO OLIVEIRA E SOUSA et al., 2014; SÃO PAULO, 2017).

## 2.4 – Justificativa e hipótese do estudo

Na rotina forense, tanto no contexto da identificação humana quanto nas investigações de relações biológicas, há situações em que a análise dos 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.

A análise do cromossomo sexual X também vem ganhando destaque devido seu padrão de herança genética. O pai transfere 100% do seu perfil genético para a sua filha, o que não ocorre com os autossômicos, nos quais apenas 50% da informação genética paterna é herdada. Essa característica dá a esse cromossomo maior capacidade de excluir em uma investigação de paternidade.

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.

O estado de São Paulo é considerado a maior mistura multicultural e étnica do país, portanto os indivíduos naturais desse devem apresentar frequências genotípicas particulares, tornando relevante estudos que identifiquem a variabilidade genética de marcadores com interesse forense nessa região.

Tendo em vista a escassez de informações na literatura sobre marcadores X-InDels em populações brasileiras e a eficiência forense dos marcadores descritos por Pereira e colaboradores (2012), estudos relacionados as variações no cromossomo X na população de São Paulo permitem compreender a diversidade desses marcadores em altamente miscigenadas e permitem possam auxiliar na resolução de casos forenses em nosso país.

# Objetivos



**Objetivo Geral**

Caracterizar a população do estado de São Paulo para 32 polimorfismos de inserção/deleção no cromossomo X e avaliar a utilidade desse multiplex na resolução de casos forenses.

**Objetivos Específicos**

- Padronizar e otimizar os custos relacionados ao protocolo de amplificação dos 32 polimorfismos de inserção/deleção do cromossomo X.
- Realizar a caracterização genética e estimar a eficiência forense desses 32 polimorfismos na população do estado de São Paulo;
- Comparar a distância genética entre São Paulo e outras populações;
- Identificar a contribuição ancestral da população desse estado;
- Analisar a segregação dos 32 X-InDels usando informação genética de trios (pai, mãe e filha);
- Identificar a eficiência de amplificação desse multiplex em amostras de DNA em baixa concentração e/ou degradadas (ossadas).

# **Materiais e Métodos**

#### **4.1 – Aspectos éticos da pesquisa**

Por se tratar de uma pesquisa com seres humanos, o projeto foi aprovado pelo Comitê de Ética em Pesquisa com Seres Humanos da Faculdade de Ciências Farmacêuticas (UNESP) de Araraquara, conforme parecer nº 1.425.912/2016 (Anexo 1).

Todos os participantes foram esclarecidos quanto aos preceitos éticos da pesquisa, aceitaram participar dessa e assinaram o Termo de Consentimento Livre e Esclarecido.

#### **4.2 – Sujeitos da pesquisa**

Para a caracterização genética dos 32 polimorfismos de inserção/deleção, foram coletadas amostras sanguíneas de 500 indivíduos (282 mulheres e 218 homens) não aparentados, nascidos no estado de São Paulo. A seleção desses e a coleta de amostras biológicas foi realizada entre alunos e funcionários da Faculdade de Ciências Farmacêuticas da UNESP de Araraquara e doadores de sangue do Hemocentro de São José do Rio Preto.

Considerando a naturalidade, os participantes selecionados nasceram em 156 diferentes cidades do estado de São Paulo, sendo que há pelo menos 1 participante de cada Região Administrativa e Metropolitana.

Para a análise de segregação foram selecionados 101 trios envolvendo pai, mãe e filha em que o vínculo genético biológico foi confirmado pelo Laboratório de Investigação de Paternidade da Faculdade de Ciências Farmacêuticas da UNESP de Araraquara através da análise de 20 marcadores STR autossômicos.

Para analisar a eficiência de amplificação desses marcadores em amostras biológicas degradadas, foram selecionadas 10 ossadas pertencentes a 8 indivíduos falecidos, que foram exumados a pedido judicial para análise de investigação de paternidade pelo Laboratório supracitado.

#### **Critérios de Inclusão**

##### **a) Para a análise da diversidade genética**

- Indivíduos não aparentados;
- Nascidos no estado de São Paulo;
- Idade mínima de 18 anos;
- Indivíduos que aceitaram participar do estudo.

##### **b) Para análise de segregação**

- Vínculo genético biológico confirmado;
- Indivíduos que aceitaram participar do estudo.

### **Critérios de Exclusão**

- Parentesco auto-declarado entre indivíduos da amostra selecionada para a análise da diversidade genética;
- Vínculo genético biológico não confirmado para a análise de segregação;
- Indivíduos que não concordaram em participar do estudo;

### **4.3 – Coleta das amostras biológicas e extração do DNA**

As amostras de sangue foram coletadas em cartão Whatman FTA<sup>®</sup> (GE Healthcare, Little Chalfont, Reino Unido) e armazenadas a 8°C até o momento da extração do DNA, que foi realizada a partir de uma adaptação feita na metodologia proposta por Singer-Sam, Tanguay e Riggs (1989) e Walsh, Metzger, e Higuchi (1991).

Com o auxílio da caneta *Harris Micro Punch*<sup>®</sup> (GE Healthcare, Little Chalfont, Reino Unido), foram coletados 2 discos de 1,2mm do centro da gota de sangue do cartão e depositados em tubo de 0,2ml. Os discos foram agitados por 5 segundos em 50µL de água ultrapura (Milli-Q<sup>®</sup> - Millipore Corporation, Massachusetts, EUA). O sobrenadante foi descartado e os discos incubados em 50µL de água ultrapura (Milli-Q<sup>®</sup> - Millipore Corporation, Massachusetts, EUA) por 30 minutos em temperatura ambiente. O sobrenadante foi descartado e os discos foram incubados em 100µL de Chelex 100 (Bio-Rad Laboratories, Califórnia, EUA) na concentração de 5% por pelo menos 1 hora à 56°C. Após esse período, a solução foi agitada e incubada por 8 minutos à 100°C. A solução foi centrifugada a 13.000rpm por 3 minutos para que a resina se concentrasse no fundo do tubo e o sobrenadante foi utilizado para a amplificação do DNA.

A extração de DNA em amostras ósseas foi realizada em duplicada utilizando a técnica de total desmineralização. Para tanto os ossos foram serrados em pequenos fragmentos de aproximadamente 10cm e posteriormente macerados no Freezer/Mill<sup>®</sup> Spex 6770. Em 200mg do pó resultante da maceração foram adicionados 6ml de EDTA 0,5M (pH 8,0) e 200µL de proteinase K (Sigma) 20mg/ml, sendo a solução mantida em agitação a 56°C por 18 horas. Após esse período, essa foi transferida para o tubo Amicon Ultra-15 *Centrifugal Filter Units* (Millipore Corporation, Massachusetts, EUA) e centrifugada em temperatura ambiente pelo tempo necessário para que ficasse retido na coluna ~500µL da solução. Ao sobrenadante retido foi adicionado o tampão de ligação do QIAquick PCR *Purification Kit* (Qiagen, Hilden, Alemanha) na proporção 1:1. A solução foi transferida para uma nova coluna disponível no Kit e centrifugada a 13.000rpm por 60 segundos. Foi adicionado 750µL do tampão de lavagem e

repetida a centrifugação. O DNA retido na coluna foi transferido em 40 $\mu$ L de tampão de eluição.

#### 4.4 – Análise de concentração das amostras ósseas

A concentração de DNA extraído de amostras ósseas foi quantificada para todas as duplicatas através de PCR em tempo real utilizando o sistema Plexor<sup>®</sup> HY (Promega, Wisconsin, EUA), conforme as instruções do fabricante:

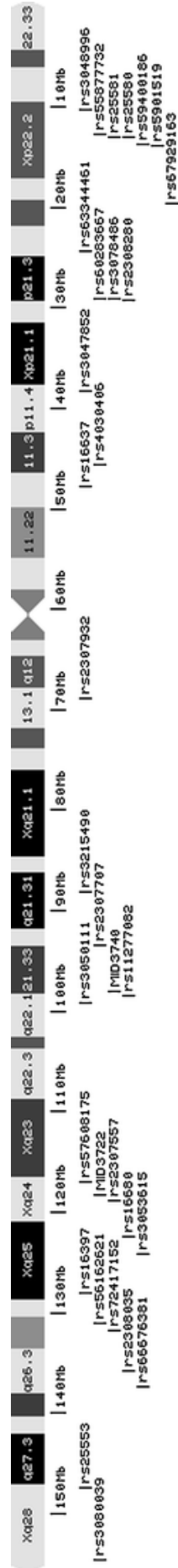
| <b>Componentes</b>  | <b>Volume</b>                |
|---|------------------------------|
| Plexor <sup>®</sup> HY 2X Master Mix  | 10,0 $\mu$ L                 |
| Plexor <sup>®</sup> HY 20X Primer/IPC Mix   | 1,0 $\mu$ L                  |
| Água disponível no Plexor <sup>®</sup> HY   | 7,0 $\mu$ L                  |
| DNA padrão disponível no Plexor <sup>®</sup> HY<br>OU<br>DNA resultante da extração | 2,0 $\mu$ L                  |
| <b>Volume final da reação</b>   | <b>20,0<math>\mu</math>L</b> |

#### 4.5 – 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<sup>®</sup>, VIC<sup>®</sup>, PET<sup>®</sup>, NED<sup>®</sup> da Applied Biosystems<sup>®</sup> (Thermo Fischer Scientific, Massachusetts, EUA), conforme sequência descrita por Pereira e colaboradores (2012). A distribuição dos 32 marcadores ao longo do cromossomo X está disposto na figura 4. Os *primers* foram aliquotados em uma única solução conforme as concentrações descritas pelos autores.

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

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



*a) Para amostras sanguíneas*

Na PCR em multiplex para amostras sanguíneas um controle positivo (amostra comercial 9947A, Promega, Wisconsin, EUA) e um negativo foram incluídos em todas as reações. O protocolo utilizado foi modificado do originalmente proposto por Pereira e colaboradores (2012), utilizando:

| <b>Reação em Multiplex</b>  | <b>Volume</b>       |
|---|---------------------|
| Multiplex PCR Master Mix 2x<br>(Qiagen, Hilden, Alemanha)                               | 2,5µL               |
| Solução de <i>primers</i> 10x   | 0,5µL               |
| Água ultrapura autoclavada<br>(Milli-Q® - Millipore Corporation,<br>Massachusetts, EUA) | q.s.p. 5,0µL        |
| Solução de DNA extraído por<br>Chelex 100<br>(Bio-Rad Laboratories, Califórnia, EUA)    | 1,0µL (~5ng de DNA) |
| <b>Volume final da reação</b>   | <b>5,0µL</b>        |

Os fragmentos foram amplificados em termociclador Veriti® Thermal Cycler (Thermo Fischer Scientific, Massachusetts, EUA), conforme o ciclo:

|                |              | Temperatura | Tempo  |
|----------------|--------------|-------------|--------|
| Incubação      |              | 90°C        | 15 min |
| 29 ciclos      | Desnaturação | 94°C        | 30 seg |
|                | Hibridização | 60°C        | 90 seg |
|                | Extensão     | 72°C        | 45 seg |
| Extensão final |              | 72°C        | 60 min |

*b) Para amostras ósseas*

Para a PCR em multiplex para amostras ósseas foi utilizado como controle positivo uma amostra sanguínea com genótipo previamente conhecido. Seguiu-se o protocolo descrito por Pereira e colaboradores (2012), utilizando:

| <b>Reação em Multiplex</b>  | <b>Volume</b>                                |
|---|--|
| Multiplex PCR Master Mix 2x<br>(Qiagen, Hilden, Alemanha)   | 5,0 $\mu$ L                                  |
| Solução de <i>primers</i> 10x   | 1,0 $\mu$ L                                  |
| Água ultrapura autoclavada<br>(Milli-Q <sup>®</sup> - Millipore Corporation,<br>Massachusetts, EUA) | q.s.p. 10,0 $\mu$ L                          |
| Solução de DNA  | Variável conforme<br>concentração da amostra |
| <b>Volume final da reação</b>   | <b>10,0<math>\mu</math>L</b>                 |

Os fragmentos foram amplificados em termociclador Veriti<sup>®</sup> Thermal Cycler (Thermo Fischer Scientific, Massachusetts, EUA), conforme o ciclo:

|                |              | Temperatura | Tempo  |
|----------------|--------------|-------------|--------|
| Incubação      |              | 90°C        | 15 min |
| 30 ciclos      | Desnaturação | 94°C        | 30 seg |
|                | Hibridização | 60°C        | 90 seg |
|                | Extensão     | 72°C        | 45 seg |
| Extensão final |              | 72°C        | 60 min |

#### 4.5.1 – Preparação para eletroforese capilar

O protocolo original de Pereira e colaboradores (2012) utilizou como padrão de corrida o GeneScan<sup>®</sup> 500 LIZ<sup>®</sup> (Thermo Fischer Scientific, Massachusetts, EUA). Para as amostras desse estudo foi realizada a adequação dos eletroferogramas para o padrão de corrida GeneScan<sup>®</sup> 600 LIZ<sup>®</sup> (Thermo Fischer Scientific, Massachusetts, EUA). Assim, os produtos amplificados de todas as amostras foram preparados para a corrida eletroforética como segue:



|   |               |
|---|---------------|
| Hi-Di <sup>®</sup> <i>Formamide</i><br>(Thermo Fischer Scientific, Massachusetts, EUA)                          | 8,7µl         |
| Padrão de corrida GeneScan <sup>®</sup> 600 LIZ <sup>®</sup><br>(Thermo Fischer Scientific, Massachusetts, EUA) | 0,3µl         |
| Produto da amplificação   | 1,0µl         |
| <b>Volume final</b>   | <b>10,0µL</b> |

Para eletroforese de *Human Identification* (HID) no Analisador Genético ABI 3500 utilizou-se POP-4 *Polymer* e capilar de 36cm (todos da Applied Biosystems<sup>®</sup>, Thermo Fischer Scientific, Massachusetts, EUA).

Os eletroferogramas resultantes foram analisados e genotipados com o auxílio do *software* GeneMapper<sup>®</sup> ID-X v1.2.1 da Applied Biosystems<sup>®</sup> (Thermo Fischer Scientific, Massachusetts, EUA).

#### 4.6 – Forma de análise dos resultados e parâmetros estatísticos

##### 4.6.1 – Análise intrapopulacional

A frequência alélica e diversidade gênica foram calculadas para a população total dos indivíduos nascidos no estado de São Paulo utilizando o *software* Arlequin v3.5 (EXCOFFIER; LISCHER, 2010). 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, com o auxílio do mesmo *software*.

A análise do desequilíbrio de ligação foi feita apenas usando os haplótipos masculinos. Nesse caso, o número de passos de Cadeia de Markov utilizados foi 1.000.000 e o número de passos de dememoração foi 10.000.

Para essas análises a correção para múltiplos testes porposta 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.

Também para a população total, os parâmetros de interesse forense foram calculados com o auxílio da ferramenta de cálculo do banco de dados internacional de X-STRs<sup>4</sup>: PDF, poder de

<sup>4</sup> Endereço eletrônico: <http://www.chrx-str.org/>

discriminação feminino;  $PD_M$ , poder de discriminação masculino;  $MEC_T$ , chance média de exclusão em trios envolvendo filha;  $MEC_D$ , chance média de exclusão em duo pai/filha, todos segundo Desmarais et al. (1998); e  $MEC_{PGM}$ , chance média de exclusão em casos deficientes envolvendo mãe, filha e avó paterna, segundo Krüger et al. (1968).

A análise de segregação foi feita através da observação da transmissão alélica entre os genitores (pai e mãe) para a filha. Com essa observação foi possível identificar possíveis casos de mutação e determinar se os marcadores apresentaram a transmissão parental esperada.

O perfil proveniente do DNA extraído de ossadas foi analisado considerando a dificuldade para a reprodução da metodologia, quantidade de marcadores amplificados e artefatos encontrados.

#### 4.6.2 – Análise comparativa

A fim de se obter uma visão mais completa da diversidade genética da população de São Paulo, os resultados foram comparados com dados de outras populações reportados na literatura para os 32 marcadores X-InDels.

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<sup>5</sup> (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); a opção “*use population information to test for migrants*” foi utilizada no modelo *admixture*.

---

<sup>5</sup> Endereço eletrônico: <http://www.statsoft.com.br/>

# **Resultados**

# **Discussão**

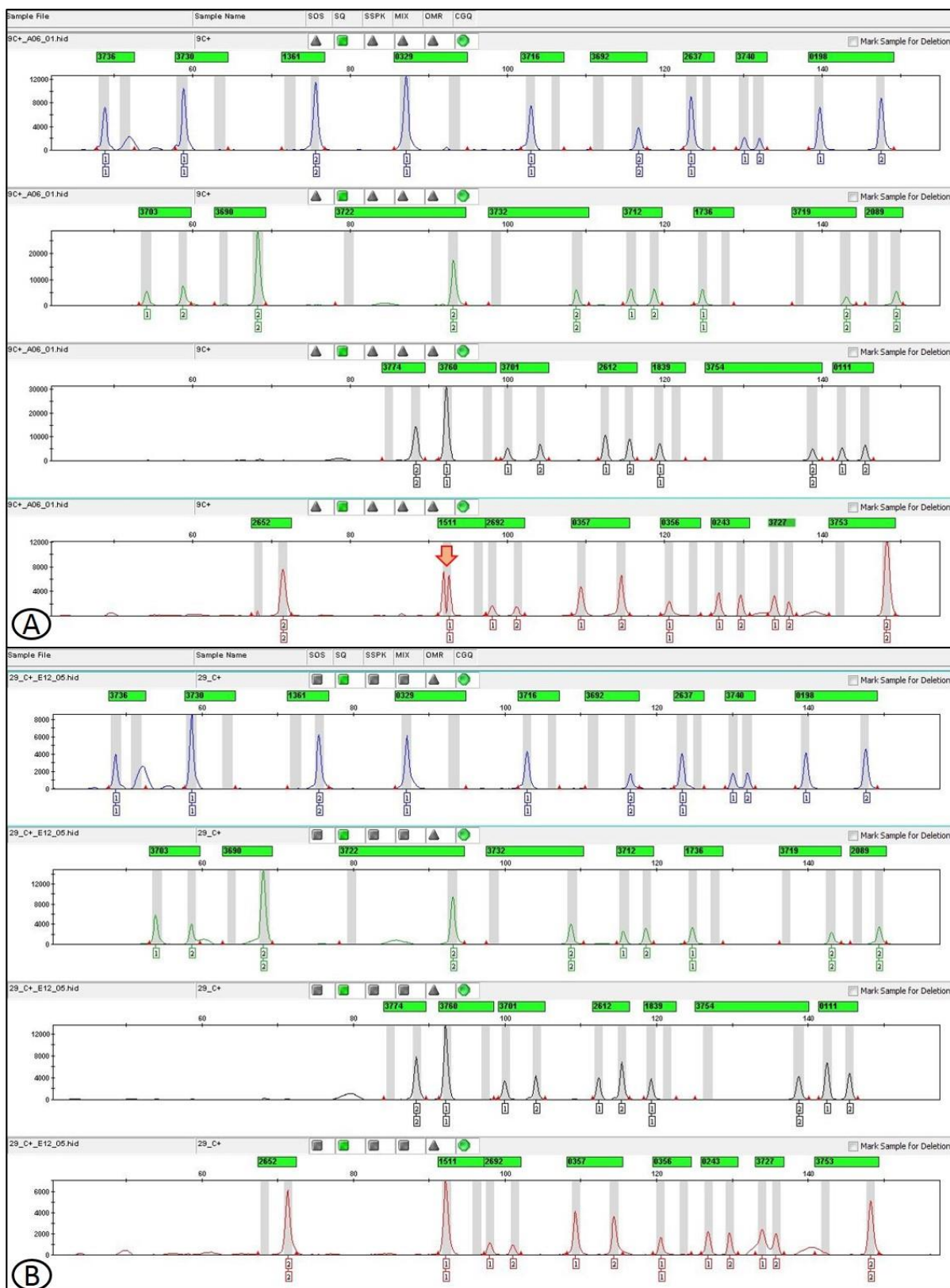
### 5.1 – Padronização e otimização dos materiais e métodos

O protocolo original proposto por Pereira e colaboradores (2012) sugere realizar a amplificação dos 32 X-InDels em uma reação com volume final de 10 $\mu$ L. Para a padronização da metodologia, os primeiros genótipos realizados seguiram as condições recomendadas pelos autores.

Tendo em vista a importância da otimização de custos dos métodos realizados na rotina dos laboratórios, sem afetar sua sensibilidade, especificidade, fácil execução e interpretação dos dados, testou-se a redução do volume final da reação para 5 $\mu$ L, mantendo a concentração estipulada do Multiplex PCR Master Mix 2x (Qiagen, Hilden, Alemanha) e *primers*.

Para a comparação dos resultados, foram realizadas concomitantemente reações seguindo o protocolo descrito pelos autores e reações seguindo o protocolo adaptado, utilizando, em ambas, as mesmas amostras de DNA.

Observando os eletroferogramas da mesma amostra feminina (Figuras 5 e 6) é possível notar que houve adequada amplificação dos marcadores nas amostras com volume final de reação 10 $\mu$ L e 5 $\mu$ L, respectivamente, sendo que neste último caso as 32 regiões também puderam ser genotipadas com a mesma precisão que o protocolo proposto pelos autores.

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

A: Perfil obtido a partir de uma PCR seguindo protocolo padrão descrito por Pereira e colaboradores (2012), com volume final de 10 $\mu$ L e 30 ciclos. A seta indica ocorrência de *pull-up*. B: Perfil obtido a partir do protocolo modificado, com volume final de 5 $\mu$ L e 29 ciclos. 1, alelo curto; 2, alelo longo.

**Fonte:** Dados da pesquisa.

O protocolo original sugere a realização da PCR com 30 ciclos. Nessas condições, o marcador MID1511 (Figura 5) comumente apresentava o artefato conhecido como *pull-up* em amostras femininas.

*Pull-up* ocorre quando o pico de uma determinada fluorescência excede a linha de base e extrapola para outra fluorescência (BUTLER, 2005). Na figura 5 é possível identificar que o pico do marcador MID3760 extrapolou a linha de base; esse fenômeno é visualizado como um pico duplo no marcador MID1511 (seta indicativa). A ocorrência de *pull-up* normalmente apresenta-se associada ao excesso de uma determinada fluorescência (BUTLER, 2005), assim, ao reduzir a PCR para 29 ciclos, eliminou-se a ocorrência desse artefato.

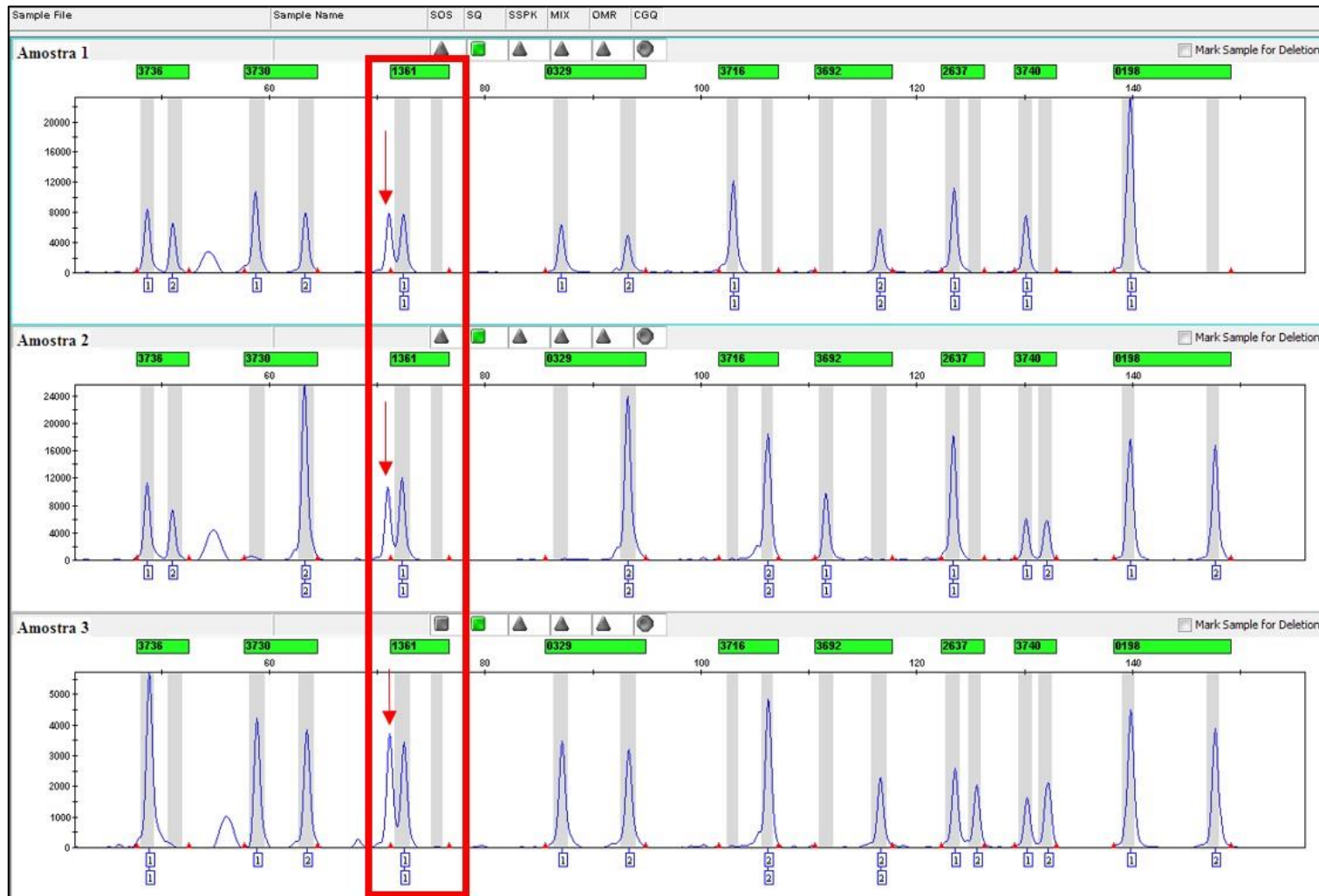
Esse fenômeno era observado somente nos perfis femininos, pois os homens apresentam-se em hemizigose para o cromossomo X e, portanto, ocorre a amplificação de somente um alelo, o que reduz em 50% a quantidade de amplicons gerados no processo de amplificação e, portanto, não ocorria o excesso de fluorescência nesses casos.

A modificação realizada no protocolo mostrou-se sensível, específica, reprodutível e de fácil execução. Somando-se a essas características há a redução do custo da reação em 50%, passando de R\$ 4,75 para R\$ 2,37 o valor da PCR realizada para cada amostra. Essa redução é relevante no orçamento dos laboratórios, pois nas análises de biologia molecular um número grande de amostras é necessário para que os dados tenham poder estatístico plausíveis de serem extrapolados para a população.

## 5.2 – Variante alélica identificada na população de São Paulo

Foram obtidos perfis genotípicos de 500 indivíduos não aparentados nascidos no estado de São Paulo, sendo 282 mulheres e 218 homens. Na análise dos genótipos foi observado em três amostras femininas um alelo variante no marcador MID1361, provavelmente resultante da deleção de uma base associada ao alelo curto (denominado no eletroferograma de alelo 1), conforme é possível observar na figura 6.

**Figura 6:** Eletroferograma parcial de três amostras femininas obtidas com o multiplex 32 X-InDels. Em destaque a variante alélica do alelo curto do marcador MID1361.



1, alelo curto; 2, alelo longo.

**Fonte:** Dados da pesquisa.

Dessas três mulheres que apresentaram a variante alélica, em duas (chamadas aqui de 61M e 64M) também foi possível analisar o perfil genotípico de suas respectivas filhas, buscando identificar se houve transmissão dessa variante para os seus descendentes.

O primeiro caso, denominado de 61F-61M-61P<sup>6</sup>, consistia em um trio simples (pai, mãe e filha) em que a filha recebeu o alelo longo esperado do pai e a variante alélica da mãe, conforme é possível observar na figura 7.

No segundo caso, 64F1-64F2-64M-64P<sup>7</sup>, foram analisados a segregação em duas filhas para os mesmos genitores. Nessa situação, a primeira filha (64F1) recebeu o alelo longo esperado do pai e o alelo curto da mãe e a segunda filha (64F2) recebeu o alelo longo esperado do pai e a variante alélica da mãe (Figura 7).

Esses dois casos confirmam que o pico deslocado à esquerda do alelo curto no marcador MID1361 trata-se efetivamente de uma variante alélica identificada na população de São Paulo. Isso porque a transferência do alelo dos genitores para os seus descendentes não seria possível se estivéssemos diante de um artefato.

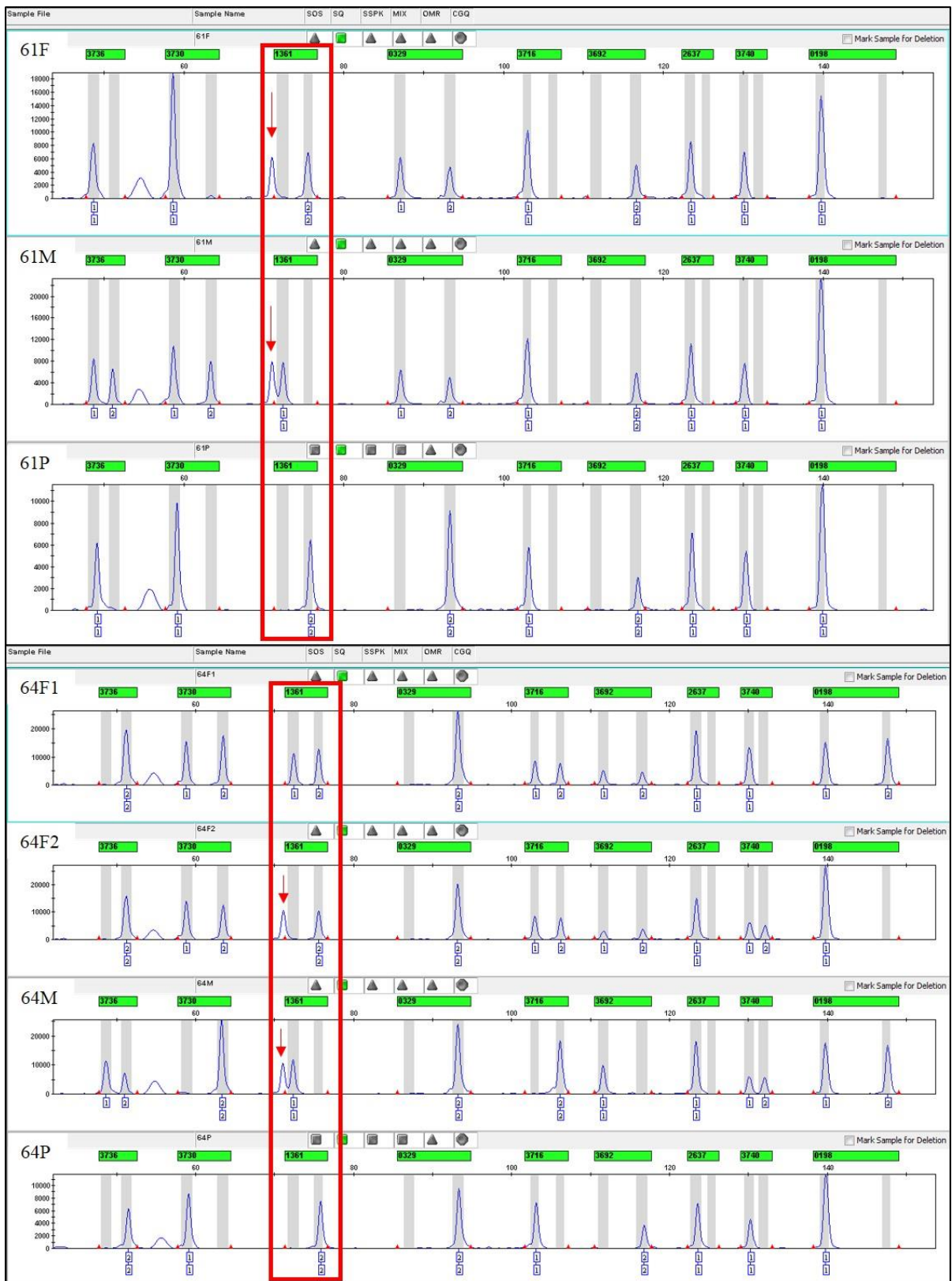
---

<sup>6</sup> 61F: filha do caso 61; 61M, mãe do caso 61; 61P, pai do caso 61.

<sup>7</sup> 64F1, primeira filha do caso 64; 64F2, segunda filha do caso 64; 64M, mãe do caso 64; 64P, pai do caso 64.



**Figura 7:** Eletroferograma parcial de dois casos de análise de segregação obtidos com o multiplex 32 X-InDels. Em destaque a variante alélica do alelo curto do marcador MID1361.



1, alelo curto; 2, alelo longo. 61F, filha do caso 61; 61M, mãe do caso 61; 61P, pai do caso 61; 64F1, primeira filha do caso 64; 64F2, segunda filha do caso 64; 64M, mãe do caso 64; 64P, pai do caso 64.

**Fonte:** Dados da pesquisa.

O marcador MID1361 apresenta predominantemente os alelos “-/ACA”, mas no dbSNP já consta registrado um padrão trialélico com a combinação de InDels “-/AA/ACA”. Entretanto, o padrão encontrado para as três amostras desse estudo consiste na deleção de 1pb no alelo curto esperado, o que se configura, então, como um padrão tetralélico para esse marcador.

Apesar dessa deleção não estar associada no dbSNP ao marcador MID1361, que se encontra na posição 119614552, logo ao seu lado, na posição 119614551, já foi observado pelo Projeto 1000 Genomas a deleção de um alelo T (rs202070714) (THE 1000 GENOMES PROJECT CONSORTIUM, 2015). Se a deleção desse T aparecer na sequência do alelo curto, então o amplicon teria -1bp que o alelo curto esperado e geraria no eletroferograma o efeito observado nesses indivíduos. Trata-se de um evento muito raro que ainda não possui dados de frequência estipulados em nenhuma população, mas é possível que seja a variante alélica encontrada nesse trabalho.

### 5.3 – Variabilidade genética da população de São Paulo

As frequências alélicas para os 32 X-InDels foram calculadas separadamente para homens e mulheres e não se observou distância genética significativas entre os sexos ( $F_{ST} \leq 0,0007$ ;  $p = 0,194 \pm 0,004$ ). Por esse motivo a frequência alélica da população de São Paulo foi calculada considerando a amostra total (Tabela 1).

**Tabela 1:** Frequência alélica e diversidade gênica (DG) dos marcadores 32 X-InDels na população de São Paulo. Número de cromossomos analisados: 782.

| Locus   | Número rs  | Alelo Curto |        | Alelo Longo | DG     |
|---------|------------|-------------|--------|-------------|--------|
| MID3727 | rs3050111  | 0,1880      |        | 0,8120      | 0,3057 |
| MID3753 | rs72417152 | 0,1304      |        | 0,8696      | 0,2271 |
| MID3736 | rs56162621 | 0,5908      |        | 0,4092      | 0,4841 |
| MID3730 | rs3215490  | 0,3082      |        | 0,6918      | 0,4270 |
| MID1361 | rs2307557  | 0,0038*     | 0,2289 | 0,7673      | 0,3594 |
| MID329  | rs25553    | 0,2110      |        | 0,7890      | 0,3334 |
| MID3716 | rs63344461 | 0,5742      |        | 0,4258      | 0,4896 |
| MID3692 | rs67929163 | 0,1944      |        | 0,8056      | 0,3136 |
| MID2637 | rs3053615  | 0,8951      |        | 0,1049      | 0,1880 |
| MID3740 |            | 0,3836      |        | 0,6164      | 0,4735 |
| MID198  | rs16637    | 0,6010      |        | 0,3990      | 0,4802 |
| MID3703 | rs59400186 | 0,3107      |        | 0,6893      | 0,4289 |
| MID3690 | rs60283667 | 0,3696      |        | 0,6304      | 0,4666 |
| MID3722 |            | 0,3760      |        | 0,6240      | 0,4698 |
| MID3732 | rs11277082 | 0,2404      |        | 0,7596      | 0,3657 |

|         |            |        |        |        |
|---------|------------|--------|--------|--------|
| MID3712 | rs55877732 | 0,2826 | 0,7174 | 0,4060 |
| MID1736 | rs2307932  | 0,3824 | 0,6177 | 0,4729 |
| MID3719 | rs3078486  | 0,2059 | 0,7941 | 0,3274 |
| MID2089 | rs2308280  | 0,3427 | 0,6573 | 0,4511 |
| MID3774 | rs5901519  | 0,3645 | 0,6356 | 0,4639 |
| MID3760 | rs66676381 | 0,8146 | 0,1854 | 0,3025 |
| MID3701 | rs4030406  | 0,4859 | 0,5141 | 0,5000 |
| MID2612 | rs3048996  | 0,4297 | 0,5703 | 0,4907 |
| MID1839 | rs2308035  | 0,6586 | 0,3414 | 0,4503 |
| MID3754 | rs57608175 | 0,3133 | 0,6867 | 0,4308 |
| MID111  | rs16397    | 0,5652 | 0,4348 | 0,4921 |
| MID2652 | rs3080039  | 0,7072 | 0,2928 | 0,4147 |
| MID1511 | rs2307707  | 0,6100 | 0,3900 | 0,4764 |
| MID2692 | rs3047852  | 0,5870 | 0,4130 | 0,4855 |
| MID357  | rs25581    | 0,3721 | 0,6279 | 0,4679 |
| MID356  | rs25580    | 0,4054 | 0,5946 | 0,4827 |
| MID243  | rs16680    | 0,8056 | 0,1944 | 0,3136 |

\*variante alélica

**Fonte:** Dados da pesquisa.

Todos os marcadores mostram-se polimórficos para a população estudada, sendo o marcador MID3701 o que apresentou maior diversidade (0,5). Nessa população, apenas o marcador MID2637 se mostrou pouco informativo, com menos de 20% de diversidade. A média da diversidade gênica foi 0,414 ( $\pm 0,084$ ), estando um pouco acima da média identificada para europeus (Portugueses = 0,388; Miranda do Douro = 0,376; Espanhóis/Zamora = 0,386), asiáticos/Macau (0,347) e para a população de Angola e Moçambique (0,408) (PEREIRA et al., 2012; PINTO et al., 2015).

Entre as amostras masculinas, não foi identificado nenhum perfil idêntico. A diversidade haplotípica consiste na probabilidade de dois indivíduos selecionados ao acaso em uma população apresentarem haplótipos diferentes para um conjunto de marcadores. Para os 32 X-InDels, a diversidade haplotípica foi igual a 1,0 ( $\pm 0,0004$ ), o que significa que para esse conjunto de marcadores há uma probabilidade baixa de dois indivíduos do sexo masculino apresentarem o mesmo haplótipo.

Trinta e um marcadores apresentaram-se em equilíbrio de Hardy-Weinberg após correção de Bonferroni para múltiplos testes. O marcador MID1361 se mostrou em desequilíbrio mesmo após correção, conforme tabela 2.

**Tabela 2:** Parâmetros de diversidade genética e probabilidade de HWE para os marcadores 32 X-InDels na população feminina de São Paulo. Tamanho da amostra: 282.

| Locus   | Ho     | He     | p-HWE   |
|---------|--------|--------|---------|
| MID3727 | 0,2766 | 0,3058 | 0,1192  |
| MID3753 | 0,2092 | 0,2310 | 0,1225  |
| MID3736 | 0,4787 | 0,4858 | 0,8083  |
| MID3730 | 0,4610 | 0,4354 | 0,3391  |
| MID1361 | 0,3121 | 0,3673 | 0,0014* |
| MID329  | 0,3333 | 0,3189 | 0,5728  |
| MID3716 | 0,4965 | 0,4944 | 1,0000  |
| MID3692 | 0,2908 | 0,2968 | 0,6934  |
| MID2637 | 0,1702 | 0,1849 | 0,1917  |
| MID3740 | 0,4823 | 0,4751 | 0,8992  |
| MID198  | 0,4645 | 0,4881 | 0,4598  |
| MID3703 | 0,4575 | 0,4315 | 0,3361  |
| MID3690 | 0,4929 | 0,4709 | 0,4516  |
| MID3722 | 0,4645 | 0,4726 | 0,8016  |
| MID3732 | 0,3475 | 0,3592 | 0,6208  |
| MID3712 | 0,3972 | 0,4009 | 0,8814  |
| MID1736 | 0,5000 | 0,4832 | 0,6237  |
| MID3719 | 0,3262 | 0,3515 | 0,2372  |
| MID2089 | 0,4539 | 0,4626 | 0,7948  |
| MID3774 | 0,4645 | 0,4532 | 0,6988  |
| MID3760 | 0,2731 | 0,2991 | 0,1625  |
| MID3701 | 0,4929 | 0,5008 | 0,8103  |
| MID2612 | 0,5284 | 0,4881 | 0,1807  |
| MID1839 | 0,4114 | 0,4476 | 0,1878  |
| MID3754 | 0,4220 | 0,4366 | 0,5887  |
| MID111  | 0,4291 | 0,4881 | 0,0515  |
| MID2652 | 0,3972 | 0,4071 | 0,7702  |
| MID1511 | 0,5106 | 0,4825 | 0,3884  |
| MID2692 | 0,4220 | 0,4845 | 0,0362  |
| MID357  | 0,4114 | 0,4664 | 0,0550  |
| MID356  | 0,4362 | 0,4804 | 0,1375  |
| MID243  | 0,2908 | 0,3058 | 0,4348  |

Ho, heterozigose observada; He, heterozigose esperada; p-HWE, valor de p para o equilíbrio de Hardy-Weinberg, com nível de significância de 0,0016 (após correção de Bonferroni para múltiplos testes); \*marcador em desequilíbrio.

**Fonte:** Dados da pesquisa.

Conforme proposto por Szibor e colaboradores (2003), é recomendável que os marcadores do cromossomo X utilizados para aplicações forenses apresentem-se em equilíbrio

de Hardy-Weinberg. 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, as frequências de seus alelos permanecem inalteradas ao longo das gerações e as proporções genóticas atingem um equilíbrio estável (ANDREWS, 2010).

Foi observado desequilíbrio de Hardy-Weinberg em MID1361 na população de São Paulo e na população alemã por Edelman e colaboradores (2016). Não há na literatura alelo nulo descrito que justifique o desequilíbrio encontrado e em 10 amostras aleatórias foi realizado uma PCR com a redução da temperatura de hibridização para 54°C, mas essas mantiveram-se homocigotas para o alelo longo. Ao realizar uma busca no SNPdb nas regiões que correspondem os *primers* desse marcador, dois SNPs foram identificados (Figura 8).

**Figura 8:** SNPs identificados nas regiões dos *primers* de amplificação do marcador MID1361.

|   |  |
|---|--|
| rs914334630 [ <i>Homo sapiens</i> ]                     |  |
| AGTACAGGGAAACTCAGTCCTTAA[A/C]AAGGGAGCCCACTGACTGAACGCC   |  |
| Chromosome:   | X:119614514  |
| Validated:  | no info  |
| HGVS:   | NC_000023.11:g.119614514C>A, NG_012565.1:g.83857G>T                              |
| rs749840100 [ <i>Homo sapiens</i> ]                     |  |
| ACTTTTAGCTCAGGTAGCCTTCACA[A/G]TGACCATGGCAAGGCAATCACTGAT |  |
| Chromosome:   | X:119614574  |
| Validated:  | no info  |
| HGVS:   | NC_000023.10:g.118748537A>G, NC_000023.11:g.119614574A>G, NG_012565.1:g.83797T>C |

**Fonte:** SNPdb.

O SNP rs914334630 encontra-se na região do *primer forward* e o rs749840100 na região do *primer reverse*. São eventos raros, com dados de frequência não descritos na literatura e ainda não validados pelo Projeto 1000 Genomas, o que não nos permite afirmar se eles estão presentes na população de São Paulo. Entretanto, se qualquer uma dessas substituições ocorrer nos participantes desse estudo, aumenta-se a chance da presença de alelo nulo para esse marcador e, conseqüentemente, a ocorrência de desequilíbrio de Hardy-Weinberg.

#### 5.4 – Desequilíbrio de ligação entre os marcadores X-InDels

Quando estudamos marcadores ligados, que estão localizados próximo 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. Já LD consiste na dependência de alelos em diferentes loci, não necessariamente no mesmo cromossomo, e é definido como uma associação não randômica de alelos na população. Quando os marcadores em análise estão em LD, não se pode fazer inferência às frequências haplotípicas baseada na frequência alélica, sendo que a frequência haplotípica observada deve ser calculada 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. A análise resultou em 496 pareamentos, sendo na tabela 3 demonstrado apenas os marcadores que apresentaram valor de  $p \leq 0,05$ .

**Tabela 3:** Desequilíbrio de ligação dos pares de marcadores analisados na população masculina de São Paulo que obtiveram valor de  $p \leq 0,05$ . Tamanho da amostra: 218.

| Par de Indels    | Distância em pb <sup>a</sup> | p-LD    |
|------------------|------------------------------|---------|
| MID3703-MID3722  | 104444857                    | 0,0450  |
| MID3716-MID3719  | 4805824                      | 0,0003  |
| MID357- MID356   | 5187                         | 0,0000* |
| MID3740- MID2089 | 68748573                     | 0,0327  |
| MID3703- MID3774 | 97700                        | 0,0000* |
| MID3703- MID3760 | 123658495                    | 0,0028  |
| MID3774- MID3760 | 123560795                    | 0,0366  |
| MID3719- MID2089 | 117035                       | 0,0000* |
| MID3690- MID3719 | 56862                        | 0,0000* |
| MID3727- MID3701 | 53626288                     | 0,0085  |
| MID3774- MID3701 | 31730201                     | 0,0274  |
| MID3774- MID1839 | 121886920                    | 0,0441  |
| MID3736-MID3754  | 14073560                     | 0,0333  |
| MID3712-MID3754  | 104329791                    | 0,0095  |
| MID3722- MID111  | 9802227                      | 0,0047  |
| MID329- MID2652  | 7168177                      | 0,0103  |
| MID3774- MID2652 | 140752961                    | 0,0074  |
| MID1839- MID2652 | 18866041                     | 0,0007  |

|                  |           |         |
|------------------|-----------|---------|
| MID3736-MID1511  | 37583541  | 0,0207  |
| MID3740- MID1511 | 4514540   | 0,0091  |
| MID3722- MID1511 | 24764151  | 0,0120  |
| MID1736- MID1511 | 24658526  | 0,0096  |
| MID2652- MID1511 | 61169955  | 0,0040  |
| MID2089- MID2692 | 9104728   | 0,0497  |
| MID3692- MID2652 | 135045709 | 0,0000* |
| MID2692-MID243   | 84107713  | 0,0430  |

p-LD, valor de p para o teste de desequilíbrio de ligação; \*, valor de p estatisticamente significativa para o teste de desequilíbrio de ligação, com nível de significância de 0,0001 (após correção de Bonferroni para múltiplos testes); <sup>a</sup>, distância em pares de bases (pb) entre os marcadores.

**Fonte:** Dados da pesquisa.

Após correção de Bonferroni para múltiplas análises, foi observado valor de *p*-LD indicando desequilíbrio de ligação para 5 pares de marcadores. Em concordância com dados de outras populações, os marcadores MID356-MID357 se mostraram em desequilíbrio, sendo este evento também identificado em descendentes de judeus (FERRAGUT et al., 2017), na população da Somália e Iraque (PEREIRA et al., 2011), na população da Argentina (CAPUTO et al., 2017), Colômbia (IBARRA et al., 2014), Alemanha e países Bálticos (EDELDMANN 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 na população de Belém, PA, Brasil (FREITAS et al., 2010). Como é possível observar na tabela 3, estes marcadores apresentam-se muito próximos no cromossomo X (~5,2Kb de distância) e, portanto, devem ser tratados como haplótipos quando forem utilizados juntos.

Os marcadores MID3703-MID3774 parecem também estar associados, sendo esse desequilíbrio identificado em colombianos (IBARRA et al., 2014), africanos, europeus e em populações miscigenadas do Brasil (RESQUE et al., 2010).

O desequilíbrio entre MID3719-MID2089 também foi identificado em descendentes de judeus, na Espanha (Maiorca) (FERRAGUT et al., 2017), Portugal (PEREIRA et al., 2012), Somália (PEREIRA et al., 2011; PEREIRA et al., 2012), Iraque (PEREIRA et al., 2011), Alemanha, países Bálticos (EDELDMANN et al., 2016) e Colômbia (IBARRA et al., 2014). E o desequilíbrio em MID3690-MID3719 foi observado em Angola, Moçambique (PEREIRA et al., 2012), Colômbia (IBARRA et al., 2014), Alemanha e países Bálticos (EDELDMANN et al., 2016).

Na verdade, MID3690-MID3719-MID2089 estão distantes uns dos outros somente ~170Kb, formando um *cluster* no cromossomo. Por esse motivo Pereira e colaboradores (2012) sugerem que esses três marcadores devem ser analisados em um único bloco haplotípico.

Por fim, os marcadores MID3692-MID2652 se apresentaram em desequilíbrio somente para a população de São Paulo. Esse evento nunca foi relatado na literatura, mas Ribeiro-Rodrigues e colaboradores (2009) identificaram que MID2652 se mostrou um marcador informativo de ancestralidade para a população europeia. A proximidade no cromossomo (ou *linkage*) não consiste na única causa do LD, podendo ser ocasionado também por fatores como seleção e miscigenação (TILLMAR, 2010; ZARRABEITIA et al., 2006). Assim, esses marcadores não estão em *linkage*, mas é possível que haja uma subestrutura populacional que favoreça a ocorrência do LD.

Para os marcadores em desequilíbrio, a frequência haplotípica deve ser utilizada no lugar da frequência alélica nas análises estatísticas forenses.

Sabendo-se que o cromossomo X apresenta-se em hemizigose em homens, a frequência dos blocos haplotípicos desses marcadores foi calculada por simples contagem para as 218 amostras masculinas (Tabela 4).



**Tabela 4:** Frequência haplotípica observada na população masculina de São Paulo para os marcadores em desequilíbrio de ligação. Tamanho da amostra: 218.

| <b>BLOCO I</b>   |             |                   |                   |
|------------------|-------------|-------------------|-------------------|
| <b>357</b>       | <b>356</b>  | <b>Frequência</b> |                   |
| 1                | 1           | 0,3807            |                   |
| 1                | 2           | 0,0000            |                   |
| 2                | 1           | 0,0413            |                   |
| 2                | 2           | 0,5780            |                   |
| <b>BLOCO II</b>  |             |                   |                   |
| <b>3703</b>      | <b>3774</b> | <b>Frequência</b> |                   |
| 1                | 1           | 0,0505            |                   |
| 1                | 2           | 0,2523            |                   |
| 2                | 1           | 0,3624            |                   |
| 2                | 2           | 0,3349            |                   |
| <b>BLOCO III</b> |             |                   |                   |
| <b>3692</b>      | <b>2652</b> | <b>Frequência</b> |                   |
| 1                | 1           | 0,0963            |                   |
| 1                | 2           | 0,1330            |                   |
| 2                | 1           | 0,5872            |                   |
| 2                | 2           | 0,1835            |                   |
| <b>BLOCO IV</b>  |             |                   |                   |
| <b>3690</b>      | <b>3719</b> | <b>2089</b>       | <b>Frequência</b> |
| 1                | 1           | 1                 | 0,0046            |
| 1                | 1           | 2                 | 0,0000            |
| 1                | 2           | 1                 | 0,1009            |
| 1                | 2           | 2                 | 0,2431            |
| 2                | 1           | 1                 | 0,1239            |
| 2                | 1           | 2                 | 0,0229            |
| 2                | 2           | 1                 | 0,0642            |
| 2                | 2           | 2                 | 0,4404            |

1, alelo curto; 2, alelo longo.

**Fonte:** Dados da pesquisa.

### 5.5 – Parâmetros estatísticos de eficiência forense

Para identificarmos a representatividade de uma evidência de DNA na resolução de casos forenses, primeiramente deve-se calcular algumas propriedades dos marcadores de DNA selecionados para determinada população, como os parâmetros estatísticos de eficiência forense (TILLMAR, 2010)

Testes de paternidade consistem nas análises mais comuns na rotina da genética forense. A eficiência de um marcador nesses testes é mensurada através 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).

Diante de uma situação de paternidade com o trio (mãe, filha e suposto pai), o valor combinado de  $MEC_T$  para a população de São Paulo foi de 0,999996. Para situações com o duo (suposto pai e filha) o valor combinado de  $MEC_D$  foi 0,9995. 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 (Tabela 5).

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 5:** Parâmetros estatísticos de interesse forense para os marcadores 32 X-InDels na população de São Paulo. Número de cromossomos analisados: 782.

| Locus     | PDF      | PD <sub>M</sub> | MEC <sub>T</sub> | MEC <sub>D</sub> | MEC <sub>PGM</sub> |
|-----------|----------|-----------------|------------------|------------------|--------------------|
| Bloco I   | 0,670419 | 0,519274        | 0,420791         | 0,28689          | 0,230574           |
| Bloco II  | 0,842107 | 0,690346        | 0,628338         | 0,48202          | 0,417683           |
| Bloco III | 0,792257 | 0,595037        | 0,551289         | 0,4017           | 0,360364           |
| Bloco IV  | 0,881023 | 0,716775        | 0,678015         | 0,53652          | 0,494141           |
| MID3727   | 0,470774 | 0,305287        | 0,258687         | 0,15264          | 0,129343           |
| MID3753   | 0,37649  | 0,226836        | 0,201109         | 0,11342          | 0,100554           |
| MID3736   | 0,61635  | 0,483514        | 0,366621         | 0,24176          | 0,183311           |
| MID3730   | 0,580082 | 0,42641         | 0,335497         | 0,21321          | 0,167749           |
| MID1361   | 0,527286 | 0,358902        | 0,297194         | 0,18147          | 0,150253           |
| MID329    | 0,499624 | 0,332958        | 0,277527         | 0,16648          | 0,138764           |
| MID3716   | 0,619317 | 0,488998        | 0,369438         | 0,2445           | 0,184719           |
| MID2637   | 0,322594 | 0,187729        | 0,170108         | 0,09386          | 0,085054           |
| MID3740   | 0,610358 | 0,472916        | 0,361091         | 0,23646          | 0,180546           |
| MID198    | 0,61417  | 0,47959         | 0,364587         | 0,2398           | 0,182293           |
| MID3722   | 0,608194 | 0,469228        | 0,359141         | 0,23461          | 0,17957            |
| MID3732   | 0,530367 | 0,365226        | 0,298531         | 0,18261          | 0,149266           |

|              |                      |                  |                 |               |              |
|--------------|----------------------|------------------|-----------------|---------------|--------------|
| MID3712      | 0,564341             | 0,405483         | 0,323275        | 0,20274       | 0,161637     |
| MID1736      | 0,610009             | 0,472317         | 0,360775        | 0,23616       | 0,180388     |
| MID3760      | 0,46728              | 0,302079         | 0,256453        | 0,15104       | 0,128227     |
| MID3701      | 0,624802             | 0,499604         | 0,374802        | 0,2498        | 0,187401     |
| MID2612      | 0,619907             | 0,490107         | 0,370005        | 0,24505       | 0,185002     |
| MID1839      | 0,596062             | 0,449711         | 0,348591        | 0,22486       | 0,174296     |
| MID3754      | 0,582853             | 0,430286         | 0,337713        | 0,21514       | 0,168857     |
| MID111       | 0,620638             | 0,491493         | 0,37071         | 0,24575       | 0,185355     |
| MID1511      | 0,612029             | 0,475813         | 0,362614        | 0,23791       | 0,181307     |
| MID2692      | 0,617095             | 0,484876         | 0,367324        | 0,24244       | 0,183662     |
| MID243       | 0,479238             | 0,313181         | 0,26414         | 0,15659       | 0,13207      |
| <b>Total</b> | <b>0,99999999993</b> | <b>0,9999993</b> | <b>0,999996</b> | <b>0,9995</b> | <b>0,997</b> |

PD<sub>F</sub>, poder de discriminação feminino; PD<sub>M</sub>, poder de discriminação masculino; MEC<sub>T</sub>, chance média de exclusão em trios envolvendo filha, mãe e suposto pai; MEC<sub>D</sub>, chance média de exclusão em duo pai/filha; MEC<sub>PGM</sub>, chance média de exclusão em casos deficientes envolvendo mãe, filha e suposta avó paterna; Bloco I: MID357-MID356; Bloco II: MID3703-MID3774; Bloco III: MID3692-MID2652; Bloco IV: MID3690-MID3719-MID2089.

**Fonte:** Dados da pesquisa.

O Poder de Discriminação (PD) consiste na probabilidade de dois indivíduos escolhidos ao acaso na população apresentarem genótipos diferentes. Para as mulheres, o valor combinado de PD<sub>F</sub> foi de 0,99999999993 e o valor combinado para os homens foi de PD<sub>M</sub> 0,99999993 (Tabela 5).

Comparando com dados de outras populações, os 32 X-InDels se mostraram tão eficientes em São Paulo quanto foram para Somália, Angola e Moçambique (PEREIRA et al., 2012) e mais eficientes do que para Portugal e Macau (PEREIRA et al., 2012). 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).

## 5.6 – Comparação genética entre as populações e análise de ancestralidade

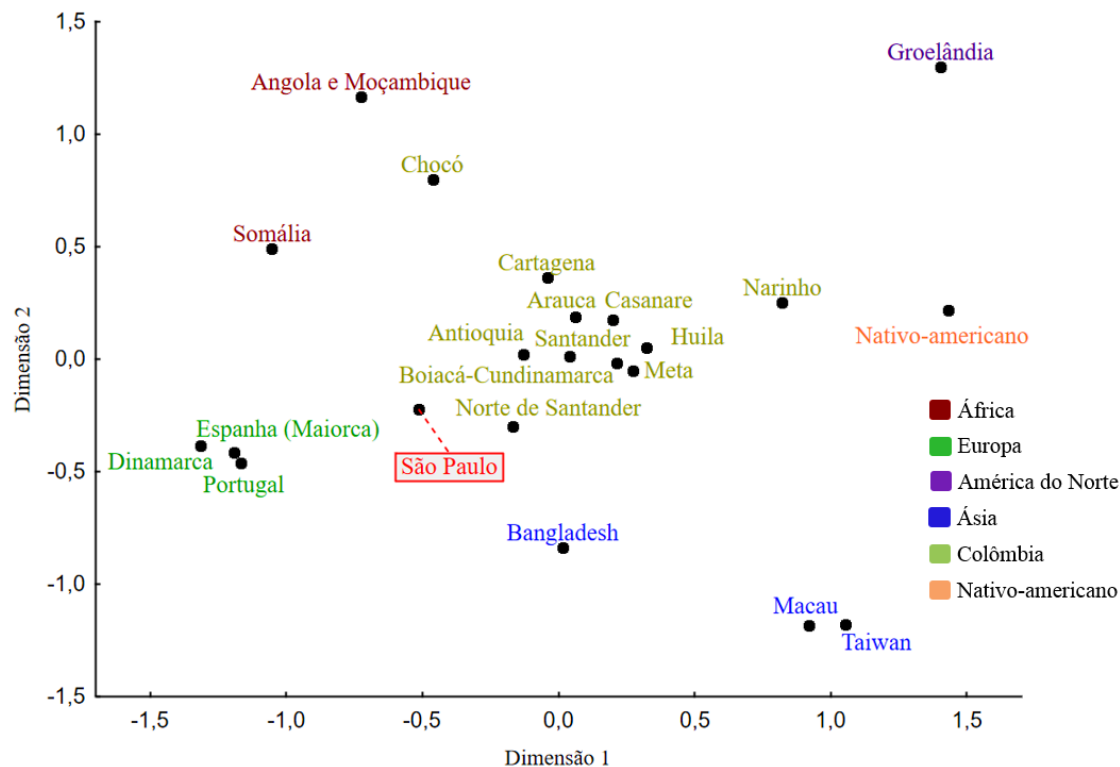
Com os dados obtidos nesse estudo foi realizada a análise da distância genética ( $F_{ST}$ ) em nível haplotípico entre a população de São Paulo e populações estrangeiras com dados disponíveis na literatura para o mesmo conjunto de X-InDels (Tabela 6).

**Tabela 6:** Amostras usadas na comparação entre as populações, número de cromossomos analisados (N) e respectivas referências.

| <b>População</b>        | <b>N</b> | <b>Referência</b>      |
|-------------------------|----------|------------------------|
| <b>EUROPA</b>           |          |                        |
| Portugal                | 324      | Pereira et al. (2012)  |
| Espanha (Maiorca)       | 112      | Ferragut et al. (2017) |
| Dinamarca               | 71       | Pereira et al. (2015)  |
| <b>ÁFRICA</b>           |          |                        |
| Angola e Moçambique     | 116      | Pereira et al. (2012)  |
| Somália                 | 162      | Pereira et al. (2012)  |
| <b>AMÉRICA LATINA</b>   |          |                        |
| Brasil (São Paulo)      | 782      | <b>Este estudo</b>     |
| 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       |                        |
| <b>NATIVO-AMERICANO</b> |          |                        |
| Pasto (Colômbia)        | 83       | Ibarra et al. (2014)   |
| <b>ÁSIA</b>             |          |                        |
| Macau                   | 100      | Pereira et al. (2012)  |
| Bangladesh              | 56       | Pereira et al. (2015)  |
| Taiwan                  | 20       | Pereira et al. (2015)  |
| <b>AMÉRICA DO NORTE</b> |          |                        |
| Groelândia              | 129      | Pereira et al. (2015)  |

Os valores da distância genética ( $F_{ST}$ ) e respectivos valores de probabilidade estão descritos na tabela 8 ao final desse subitem. A fim de melhorar a visualização dos resultados, os dados foram dispostos em um gráfico de análise em escala multidimensional (MDS) de duas dimensões (Figura 9). A adição de uma terceira dimensão ao gráfico MDS não contribui para uma melhoria significativa na visualização dos resultados e o valor de stress (0,060) foi apenas um pouco menor do que o obtido para duas dimensões (0,099).

**Figura 9:** Gráfico MDS bidimensional elaborado a partir da distância genética ( $F_{ST}$ ) entre Portugal, Espanha (Maiorca), Dinamarca, Angola e Moçambique, Somália, Groelândia, diversas populações da Colômbia, Nativo-americano, Macau, Bangladesh e Taiwan.



**Fonte:** Dados da pesquisa.

Na maioria das comparações (Tabela 7), valores significativos foram obtidos entre São Paulo e as outras populações. Esses resultados revelaram que os perfis de X-InDels provenientes desse estado são mais similares à três departamentos colombianos (Antioquia, Norte de Santander e Santander), que são caracterizados por possuir maior componente europeu, seguido dos europeus de Portugal, Espanha e Dinamarca. A população mais distante foi Groelândia, na América do Norte.

Segundo a história da formação da população brasileira, o maior contingente migratório para o Brasil foi de europeus, composto principalmente por portugueses, italianos e espanhóis. Escravos africanos são o segundo maior grupo de imigrantes, trazidos pelos europeus entre os séculos XVI e XIX. Por fim, a imigração asiática também ocorreu no início do século XX, entretanto em menor escala e com pequena contribuição para a formação populacional do país.

Em uma meta-análise, Moura e colaboradores (2015) relataram que em média a contribuição ancestral no sudeste do Brasil para marcadores autossômicos é 67% de europeus, seguido por 23% de africanos e 10% de nativo-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 para 35% quando o cromossomo X é analisado.

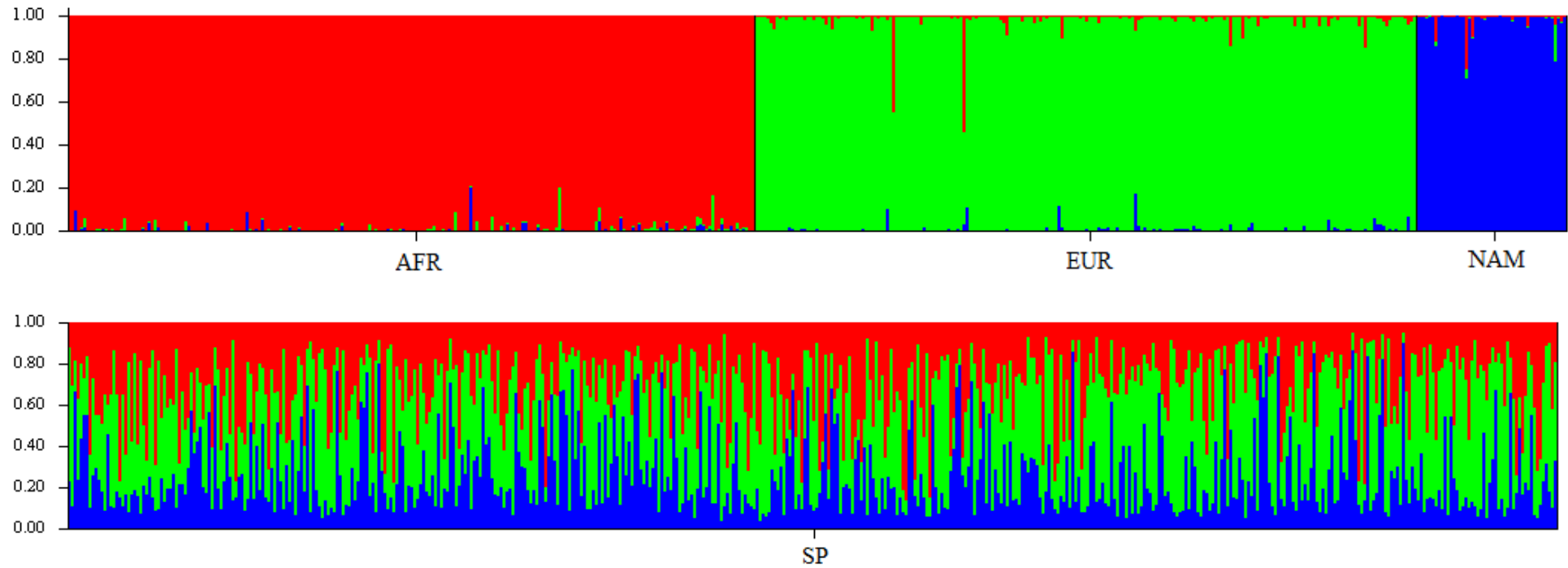
Corroborando com essa informação, Resque e colaboradores (2010) ressaltaram que a contribuição de Europeus para o cromossomo X é maior na região sudeste do Brasil, mas com valores muito próximos aos da contribuição africana, sendo estruturada por 38% de europeus, 35% de africanos e 27% de nativo-americanos.

Baseado nessas informações e no estudo de Ibarra e colaboradores (2014) em que os 32 X-InDels foram utilizados para identificar a proporção de misturas na população colombiana, foi realizada uma análise da contribuição ancestral genética desses marcadores na população de São Paulo.

Para essa análise foram utilizados os dados de africanos e europeus publicados por Pereira e colaboradores (2012). Não há na literatura dados dos marcadores estudados para populações nativas da América do Sul, por esse motivo utilizou-se os dados obtidos para os nativos de Pasto na Colômbia (IBARRA et al., 2014).

A proporção de ancestralidade identificada em marcadores X-InDels para São Paulo foi 41,8% para europeus, 31,6% para africanos e 26,6% para nativo-americanos (Figura 10), conforme sugerido por Wang e colaboradores (2008) e Resque e colaboradores (2010). Entretanto, os 32 X-InDels não se tratam de marcadores de 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.

**Figura 10:** Representação esquemática da mistura estimada para a população de São Paulo (SP) utilizando 32 marcadores X-InDels obtidos através da análise do STRUCTURE baseado na contribuição da população africana (AFR), europeia (EUR) e nativo-americana (NAM) (K=3).



**Fonte:** Dados da pesquisa.

**Tabela 7:** Distância genética ( $F_{ST}$ ) entre as populações (abaixo da diagonal) e o valor de  $p$  correspondente (acima da diagonal).

| Populações                       | Antioquia <sup>1</sup> | Boiacá-Cundinamarca <sup>1</sup> | Huila <sup>1</sup> | Santander <sup>1</sup> | Norte de Santander <sup>1</sup> | Arauca <sup>1</sup> | Casanare <sup>1</sup> | Meta <sup>1</sup> | Cartagena <sup>1</sup> | Chocó <sup>1</sup> | Narinho <sup>1</sup> | Nativo de Pasto <sup>1</sup> | Somália | Angola e Moçambique | Portugal | Espanha (Maiorca) | Dinamarca | Groelândia | Macau  | Bangladesh | Taiwan | São Paulo |        |
|----------------------------------|------------------------|----------------------------------|--------------------|------------------------|---------------------------------|---------------------|-----------------------|-------------------|------------------------|--------------------|----------------------|------------------------------|---------|---------------------|----------|-------------------|-----------|------------|--------|------------|--------|-----------|--------|
| Antioquia <sup>1</sup>           | *                      | 0,0373                           | 0,0077             | 0,4411                 | 0,0238                          | 0,0317              | 0,3281                | 0,0383            | 0,0005                 | 0,0000             | 0,0000               | 0,0000                       | 0,0000  | 0,0000              | 0,0000   | 0,0000            | 0,0000    | 0,0000     | 0,0000 | 0,0000     | 0,0000 | 0,0000    | 0,0002 |
| Boiacá-Cundinamarca <sup>1</sup> | 0,0050                 | *                                | 0,1588             | 0,4341                 | 0,0365                          | 0,0166              | 0,5176                | 0,5836            | 0,0005                 | 0,0000             | 0,0009               | 0,0000                       | 0,0000  | 0,0000              | 0,0000   | 0,0000            | 0,0000    | 0,0000     | 0,0000 | 0,0000     | 0,0015 | 0,0002    | 0,0000 |
| Huila                            | 0,0081                 | 0,0021                           | *                  | 0,1481                 | 0,0353                          | 0,0055              | 0,3158                | 0,2945            | 0,0013                 | 0,0000             | 0,0175               | 0,0000                       | 0,0000  | 0,0000              | 0,0000   | 0,0000            | 0,0000    | 0,0000     | 0,0000 | 0,0000     | 0,0022 | 0,0005    | 0,0000 |
| Santander <sup>1</sup>           | -0,0001                | 0,0000                           | 0,0020             | *                      | 0,0457                          | 0,0567              | 0,5003                | 0,3026            | 0,0008                 | 0,0000             | 0,0004               | 0,0000                       | 0,0000  | 0,0000              | 0,0000   | 0,0000            | 0,0000    | 0,0000     | 0,0000 | 0,0000     | 0,0003 | 0,0001    | 0,0000 |
| Norte de Santander <sup>1</sup>  | 0,0136                 | 0,0124                           | 0,0125             | 0,0108                 | *                               | 0,0204              | 0,1514                | 0,0390            | 0,0065                 | 0,0000             | 0,0002               | 0,0000                       | 0,0000  | 0,0000              | 0,0000   | 0,0000            | 0,0197    | 0,0000     | 0,0000 | 0,0286     | 0,0028 | 0,0430    |        |
| Arauca <sup>1</sup>              | 0,0051                 | 0,0066                           | 0,0089             | 0,0036                 | 0,0159                          | *                   | 0,7802                | 0,0407            | 0,0010                 | 0,0000             | 0,0000               | 0,0000                       | 0,0000  | 0,0000              | 0,0000   | 0,0000            | 0,0000    | 0,0000     | 0,0000 | 0,0000     | 0,0000 | 0,0000    | 0,0000 |
| Casanare <sup>1</sup>            | 0,0024                 | -0,0019                          | 0,0025             | 0,0013                 | 0,0117                          | -0,0067             | *                     | 0,1486            | 0,1035                 | 0,0008             | 0,2004               | 0,0013                       | 0,0001  | 0,0000              | 0,0000   | 0,0000            | 0,0131    | 0,0038     | 0,0000 | 0,0607     | 0,0069 | 0,0074    |        |
| Meta <sup>1</sup>                | 0,0084                 | -0,0014                          | 0,0016             | 0,0014                 | 0,0160                          | 0,0086              | 0,0103                | *                 | 0,0116                 | 0,0000             | 0,0052               | 0,0000                       | 0,0000  | 0,0000              | 0,0000   | 0,0000            | 0,0005    | 0,0000     | 0,0000 | 0,0215     | 0,0030 | 0,0004    |        |
| Cartagena <sup>1</sup>           | 0,0138                 | 0,0132                           | 0,0109             | 0,0101                 | 0,0208                          | 0,0114              | 0,0105                | 0,0122            | *                      | 0,0054             | 0,0000               | 0,0000                       | 0,0000  | 0,0000              | 0,0000   | 0,0000            | 0,0000    | 0,0000     | 0,0000 | 0,0010     | 0,0002 | 0,0000    |        |
| Chocó <sup>1</sup>               | 0,0404                 | 0,0488                           | 0,0442             | 0,0429                 | 0,0539                          | 0,0374              | 0,0426                | 0,0451            | 0,0094                 | *                  | 0,0000               | 0,0000                       | 0,0000  | 0,0116              | 0,0000   | 0,0000            | 0,0000    | 0,0000     | 0,0000 | 0,0000     | 0,0000 | 0,0000    | 0,0000 |
| Narinho <sup>1</sup>             | 0,0346                 | 0,0177                           | 0,0092             | 0,0202                 | 0,0396                          | 0,0268              | 0,0070                | 0,0192            | 0,0245                 | 0,0605             | *                    | 0,0891                       | 0,0000  | 0,0000              | 0,0000   | 0,0000            | 0,0000    | 0,0000     | 0,0000 | 0,0000     | 0,0003 | 0,0002    | 0,0000 |
| Nativo de Pasto <sup>1</sup>     | 0,0737                 | 0,0503                           | 0,0393             | 0,0568                 | 0,0853                          | 0,0593              | 0,0478                | 0,0481            | 0,0614                 | 0,1070             | 0,0058               | *                            | 0,0000  | 0,0000              | 0,0000   | 0,0000            | 0,0000    | 0,0000     | 0,0000 | 0,0000     | 0,0000 | 0,0001    | 0,0000 |
| Somália                          | 0,0489                 | 0,0667                           | 0,0671             | 0,0535                 | 0,0611                          | 0,0502              | 0,0559                | 0,0691            | 0,0292                 | 0,0236             | 0,1014               | 0,1564                       | *       | 0,0000              | 0,0000   | 0,0000            | 0,0000    | 0,0000     | 0,0000 | 0,0000     | 0,0000 | 0,0000    | 0,0000 |
| Angola e Moçambique              | 0,0640                 | 0,0838                           | 0,0766             | 0,0714                 | 0,0845                          | 0,0619              | 0,0652                | 0,0813            | 0,0307                 | 0,0071             | 0,0970               | 0,1433                       | 0,0297  | *                   | 0,0000   | 0,0000            | 0,0000    | 0,0000     | 0,0000 | 0,0000     | 0,0000 | 0,0000    | 0,0000 |
| Portugal                         | 0,0374                 | 0,0588                           | 0,0658             | 0,0472                 | 0,0337                          | 0,0594              | 0,0711                | 0,0622            | 0,0616                 | 0,0841             | 0,1231               | 0,1795                       | 0,0533  | 0,1036              | *        | 0,4610            | 0,2333    | 0,0000     | 0,0000 | 0,0000     | 0,0000 | 0,0000    | 0,0000 |
| Espanha (Maiorca)                | 0,0385                 | 0,0633                           | 0,0683             | 0,0491                 | 0,0352                          | 0,0627              | 0,0711                | 0,0647            | 0,0598                 | 0,0787             | 0,1246               | 0,1886                       | 0,0511  | 0,0941              | 0,000    | *                 | 0,3086    | 0,0000     | 0,0000 | 0,0000     | 0,0000 | 0,0000    | 0,0003 |
| Dinamarca                        | 0,0483                 | 0,0742                           | 0,0764             | 0,0570                 | 0,0433                          | 0,0692              | 0,0794                | 0,0753            | 0,0664                 | 0,0878             | 0,1362               | 0,1993                       | 0,0497  | 0,1000              | 0,002    | 0,0010            | *         | 0,0000     | 0,0000 | 0,0027     | 0,0011 | 0,0006    |        |
| Groelândia                       | 0,1049                 | 0,0919                           | 0,0792             | 0,0946                 | 0,1119                          | 0,0975              | 0,0904                | 0,0855            | 0,0949                 | 0,1368             | 0,0797               | 0,0986                       | 0,1688  | 0,1533              | 0,167    | 0,1697            | 0,1752    | *          | 0,0000 | 0,0000     | 0,0015 | 0,0000    |        |
| Macau                            | 0,0912                 | 0,0649                           | 0,0576             | 0,0781                 | 0,0633                          | 0,0854              | 0,0890                | 0,0670            | 0,0712                 | 0,1115             | 0,0690               | 0,0926                       | 0,1322  | 0,1594              | 0,131    | 0,1381            | 0,1531    | 0,1248     | *      | 0,0004     | 0,2594 | 0,0000    |        |
| Bangladesh                       | 0,0487                 | 0,0347                           | 0,0340             | 0,0405                 | 0,0428                          | 0,0504              | 0,0429                | 0,0335            | 0,0301                 | 0,0594             | 0,0614               | 0,0980                       | 0,0601  | 0,0922              | 0,073    | 0,0752            | 0,0756    | 0,1335     | 0,0555 | *          | 0,0444 | 0,0000    |        |
| Taiwan                           | 0,0942                 | 0,0750                           | 0,0636             | 0,0826                 | 0,0834                          | 0,0899              | 0,1035                | 0,0659            | 0,0755                 | 0,1148             | 0,0794               | 0,1072                       | 0,1341  | 0,1539              | 0,142    | 0,1450            | 0,1612    | 0,1248     | 0,0052 | 0,0662     | *      | 0,0000    |        |
| São Paulo                        | 0,0092                 | 0,0212                           | 0,0237             | 0,0128                 | 0,0096                          | 0,0197              | 0,0264                | 0,0240            | 0,0200                 | 0,0411             | 0,0609               | 0,1055                       | 0,0330  | 0,0620              | 0,012    | 0,0130            | 0,0169    | 0,1091     | 0,0837 | 0,0388     | 0,0922 | *         |        |

<sup>1</sup>, departamentos da Colômbia.

**Fonte:** Dados da pesquisa.



### 5.7 – Análise de segregação dos marcadores X-InDels

A fim de identificar a transferência dos alelos entre os genitores e filha para os marcadores 32 X-InDels, foram analisados 101 casos trios envolvendo pai/mãe/filha. A paternidade e maternidade de todos os casos foi previamente confirmada através da análise de 20 STRs autossômicos.

Todos os marcadores demonstraram o padrão de transmissão esperado. Foram analisadas um total de 202 meioses e nenhuma mutação foi identificada, o que corresponde a uma taxa de mutação menor que 1 em 6.464 transmissões alélicas.

Em Somália e Portugal foram analisados 214 trios e nenhuma mutação foi encontrada. (PEREIRA et al., 2012). Na população alemã também nenhuma mutação foi identificada em 97 trios (EDELMANN et al., 2016).

Considerando as segregações analisadas para todas essas populações, com um total de 412 trios, temos para os X-InDels uma taxa de mutação menor que 1 em 25.204 transmissões alélicas, com 95% de confiança e um limite superior de 0,0001.

Diferente de STRs, em que a taxa de mutação esperada é na ordem de grandeza de  $10^{-3}$  (MARTINEZ et al., 2017), para os marcadores InDels a taxa estimada é menor, com ordem de grandeza de  $10^{-9}$  (CAMPBELL; EICHLER, 2013), assim um número maior de trios é necessário para que se possa fazer essa estimativa para esses marcadores. A baixa taxa de mutação nos InDels os torna ideal na investigação de paternidade, especialmente quando são encontradas uma ou duas incompatibilidades na análise de STR autossômicos (FREITAS et al., 2010).

### 5.8 – Amplificação dos marcadores X-InDels em amostras biológicas em baixa concentração e/ou degradadas

Explorou-se teoricamente ao longo deste trabalho que o painel 32 X-InDels apresenta características promissoras para a amplificação de amostras biológicas em baixa concentração e/ou degradadas. Buscando identificar esse potencial, foi analisado em um primeiro momento a amplificação dos 32 marcadores em uma amostra comercial (9947A, Promega, Wisconsin, EUA) diluída em diferentes concentrações, conforme exposto na tabela 8.

**Tabela 8:** Amplificação dos marcadores 32 X-InDels em amostra comercial 9947A (Promega, Wisconsin, EUA) diluída para diferentes concentrações.

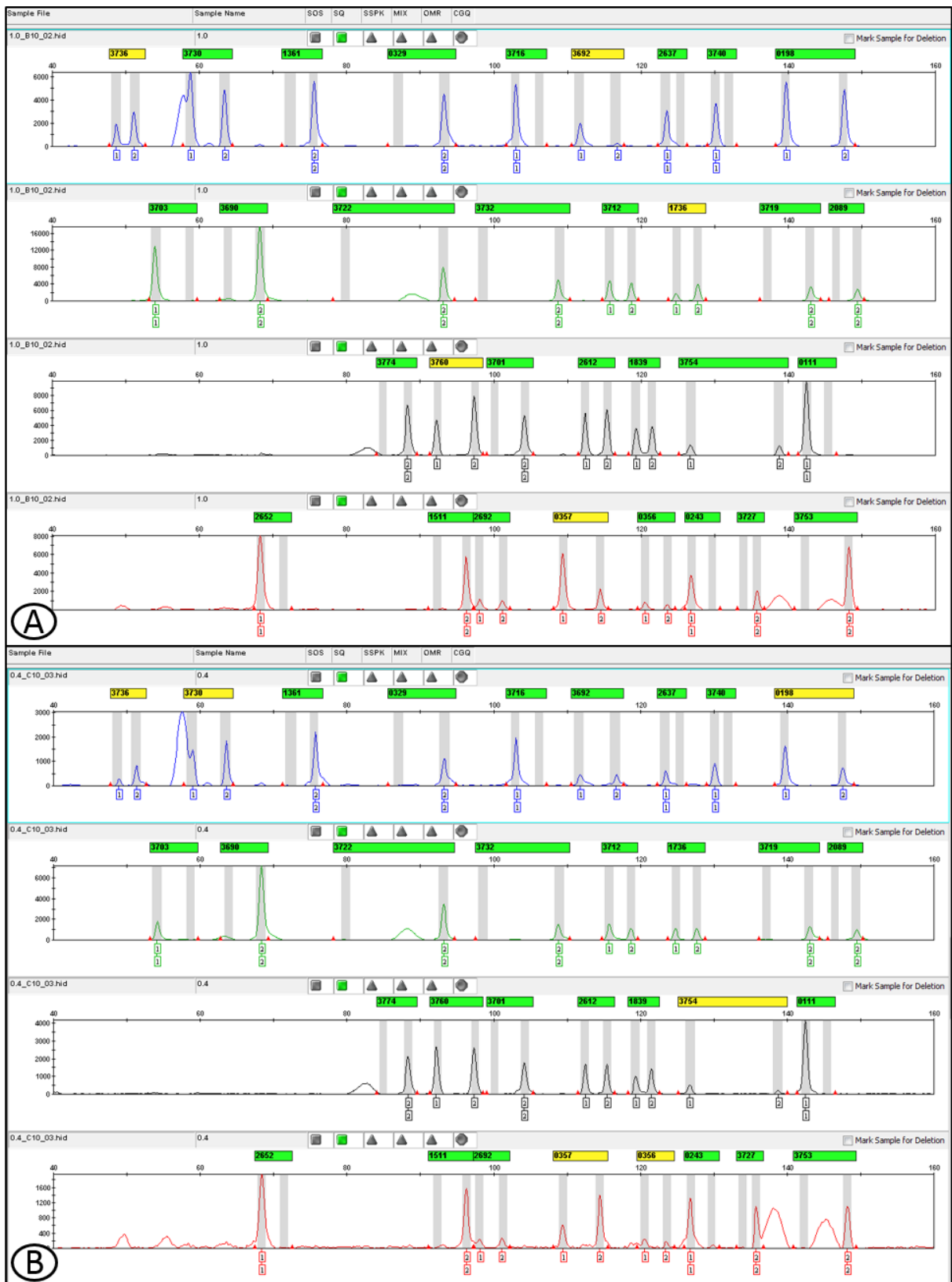
| <b>Quantidade utilizada de DNA em nanogramas (ng) para uma reação de volume final de 10<math>\mu</math>L</b> | <b>Número de marcadores amplificados</b> |
|--|--|
| 2,0  | 32 X-InDels                              |
| 1,0  | 32 X-InDels                              |
| 0,4  | 32 X-InDels                              |
| 0,08   | 23 X-InDels                              |
| 0,016  | 13 X-InDels                              |
| 0,0032   | 0 X-InDels                               |

**Fonte:** Dados da pesquisa.

Os alelos obtidos em cada perfil foram comparados com o eletroferograma da amostra 9947A (Promega, Wisconsin, EUA) disponível no trabalho publicado por Pereira e colaboradores (2012). Conforme esperado, todos os alelos foram facilmente identificados nas concentrações de 2,0ng e 1,0ng e nenhum marcador foi amplificado na concentração de 0,0032ng, devido à quantidade de DNA disponível nessa amostra.

Pereira e colaboradores (2012) destacaram que era possível a obtenção do perfil genotípico com concentrações que variavam de 0,3ng a 5ng de DNA, o que foi confirmado nesse trabalho em que todos os marcadores foram corretamente genotipados com 0,4ng, conforme se observa na figura 11.

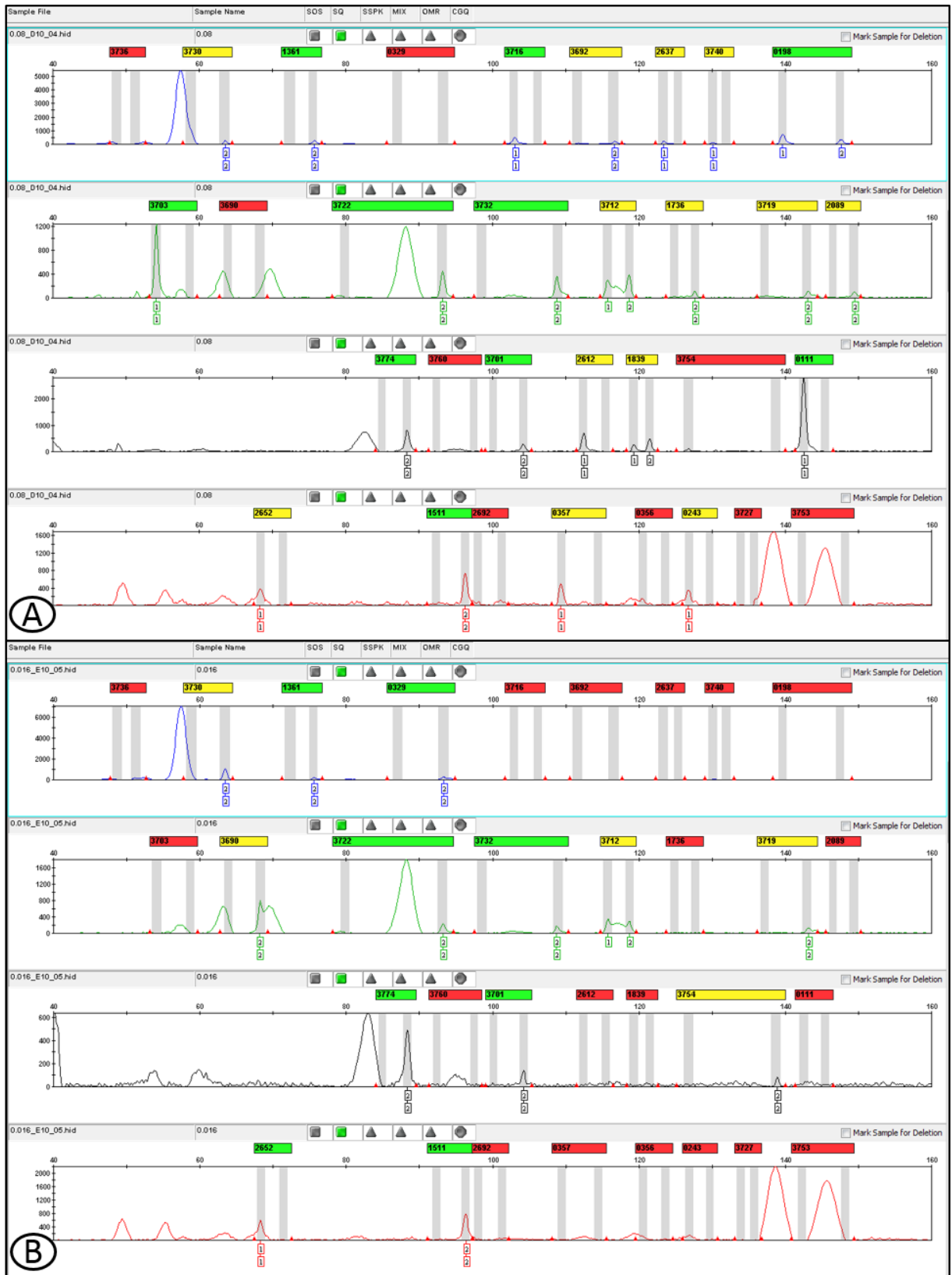
**Figura 11:** Eletroferograma dos marcadores 32 X-InDels obtidos a partir de uma amostra referência feminina 9947A (Promega, Wisconsin, EUA).



A: Perfil obtido com 1ng de DNA; B: Perfil obtido com 0,4ng de DNA.

**Fonte:** Dados da pesquisa.

**Figura 12:** Eletroferograma dos marcadores 32 X-InDels obtidos a partir de uma amostra referênciada feminina 9947A (Promega, Wisconsin, EUA).



A: Perfil obtido com 0,08ng de DNA; B: Perfil obtido com 0,016ng de DNA.  
**Fonte:** Dados da pesquisa.

Poucos marcadores foram amplificados em concentrações muito baixas de DNA. Com 0,08ng foram genotipados corretamente 18 marcadores e 5 amplificaram apenas um alelo (MID3730, MID3692, MID1736, MID2612 e MID357); com 0,016ng, 11 marcadores foram corretamente identificados e 2 amplificaram apenas um alelo (MID3730 E MID3754), conforme figura 12.

Claro que se deve destacar que a obtenção desses perfis foi possível pela utilização de um DNA comercial puro, mas, tratando-se de um painel *in house*, esses dados indicam que ele apresenta um alto poder de amplificação, mesmo utilizando uma baixa quantidade de DNA.

Diante desses aspectos, analisou-se como o painel se comporta perante amostras que comumente são encontradas na rotina das análises forenses. Foram selecionadas 10 diferentes amostras de 8 casos *post-mortem* de investigação de paternidade realizados pelo Laboratório de Investigação de Paternidade da UNESP de Araraquara, descritos na tabela 9.

**Tabela 9:** Informações relacionadas aos casos de investigação de paternidade *post-mortem* analisados no Laboratório de Investigação de Paternidade da UNESP de Araraquara.

| Amostra | Osso analisado  | Idade* <sup>1</sup> | Ano do óbito <sup>1</sup> | Marcadores Analisados <sup>2</sup> |              | Número de marcadores amplificados <sup>2</sup> |                          | Resultado final do laudo <sup>2</sup> |
|---------|-----------------|---------------------|---------------------------|------------------------------------|--------------|--|--------------------------|---------------------------------------|
|         |                 |                     |                           | 20 AS-STRs + Amelogenina           | 38 AS-InDels | 20 AS-STRs + Amelogenina                       | 35 AS-InDels             |                                       |
| OS_1    | cabeça femoral  | 85                  | 2012                      | 20 AS-STRs + Amelogenina           | 38 AS-InDels | 20 AS-STRs + Amelogenina                       | 35 AS-InDels             | Inclusão                              |
| OS_2    | diáfise femoral | 89                  | 2007                      | 20 AS-STRs + Amelogenina           | 38 AS-InDels | 20 AS-STRs + Amelogenina                       | 34 AS-InDels             | Exclusão                              |
|         | cabeça femoral  |                     |                           | 20 AS-STRs + Amelogenina           |              | 0 AS-STRs <sup>3</sup>                         |                          | **                                    |
| OS_3    | cabeça femoral  | 71                  | 2009                      | 15 AS-STRs + Amelogenina           | 38 AS-InDels | 15 AS-STRs + Amelogenina                       | 36 AS-InDels             | Exclusão                              |
| OS_4    | cabeça femoral  | 27                  | 2012                      | 20 AS-STRs + Amelogenina           | 38 AS-InDels | 11 AS-STRs <sup>3</sup>                        | 0 AS-InDels <sup>3</sup> | Inconclusivo                          |
| OS_5    | cabeça femoral  | 54                  | 2007                      | 20 AS-STRs + Amelogenina           |              | 6 AS-STRs + Amelogenina                        |                          | **                                    |
|         | dente           |                     |                           | 20 AS-STRs + Amelogenina           |              | 20 AS-STRs + Amelogenina                       |                          | Inclusão                              |
| OS_6    | cabeça femoral  | 67                  | 2006                      | 20 AS-STRs + Amelogenina           |              | 0 AS-STRs <sup>3</sup>                         |                          | Inconclusivo                          |
| OS_7    | dente           | 67                  | 2011                      | 23 AS-STRs + Amelogenina           |              | 22 AS-STRs + Amelogenina                       |                          | Inclusão                              |
| OS_8    | cabeça femoral  | 81                  | 1988                      | 23 AS-STRs + Amelogenina           |              | 17 AS-STRs + Amelogenina                       |                          | Inclusão                              |

\*Idade que o indivíduo tinha no momento do óbito;

\*\*Dado não utilizado para a elaboração do laudo;

<sup>1</sup>Informações coletadas no Cadastro Nacional de Falecidos [CNF Brasil] ([www.falecidosnobrasil.org.br](http://www.falecidosnobrasil.org.br));

<sup>2</sup>Informações coletadas no Laboratório de Investigação de Paternidade da UNESP de Araraquara;

<sup>3</sup>Não amplificou amelogenina;

AS-STRs: STRs autossômicos;

AS-InDels: InDels autossômicos.

**Fonte:** Dados da pesquisa.

Como é possível observar na tabela 10, foram selecionados casos em que a data do óbito ocorreu há mais de 10 anos e casos em que a data do óbito ocorreu há menos de 10 anos. Todas as amostras pertenciam à indivíduos do sexo masculino que tinham de 27 a 89 anos no momento do óbito.

Em uma primeira análise das quantificações (Tabela 10) pode-se observar que mesmo utilizando a mesma técnica de extração, as concentrações de DNA entre as amostras variaram consideravelmente. Conforme exposto por Rohland e Hofreiter (2007), os rendimentos de DNA extraído de amostras ósseas podem variar mesmo quando se utiliza o mesmo osso e o mesmo método de extração. Tais diferenças são provavelmente atribuíveis à heterogeneidade dentro do osso e, mesmo quando se utiliza um pó homogêneo, as alterações no rendimento nem sempre podem ser evitadas.

**Tabela 10:** Concentração de DNA utilizada na reação de amplificação dos marcadores 32 X-InDels e número de marcadores amplificados para cada amostra.

| Amostra | Osso analisado  | Concentração do DNA em ng <sup>1</sup> | Número de marcadores amplificados |
|---------|-----------------|--|-----------------------------------|
| OS_1.1* | cabeça femoral  | 4,56266                                | 32 X-InDels                       |
| OS_1.2* |                 | 0,93041                                | 12 X-InDels                       |
| OS_2.1* | diáfise femoral | 1,58913                                | 31 X-InDels                       |
| OS_2.2* |                 | 0,87459                                | 25 X-InDels                       |
| OS_2.3  | cabeça femoral  | 0,62873                                | 0 X-InDels                        |
| OS_3    | cabeça femoral  | 0,17639                                | 15 X-InDels                       |
| OS_4    | cabeça femoral  | 0,13799                                | 8 X-InDels                        |
| OS_5    | cabeça femoral  | 0,39451                                | 0 X-InDels                        |
|         | dente           | 0,60815                                | 21 X-InDels                       |
| OS_6    | cabeça femoral  | 0,11131                                | 0 X-InDels                        |
| OS_7    | dente           | 1,94094                                | 20 X-InDels                       |
| OS_8    | cabeça do fêmur | 3,74161                                | 0 X-InDels                        |

\*Refere-se à duplicata da amostra.

ng: nanogramas

<sup>1</sup>Observação: não consiste na mesma extração utilizada para análise de AS-STR e AS-InDel.

**Fonte:** Dados da pesquisa.

Nas amostras OS\_1 e OS\_2, pode-se observar uma grande diferença nas concentrações do DNA nas duplicatas. Nessas, como pelo menos uma das extrações apresentou uma concentração adequada, as duplicatas foram analisadas separadamente. Para as amostras OS\_3, OS\_4, OS\_5 e OS\_6 as concentrações estavam muito baixas em cada duplicata individualmente, o que aumentava a chance de insucesso da amplificação. Por esse motivo elas

foram utilizadas simultaneamente na mesma PCR. Uma das duplicatas das amostras OS\_7 e OS\_8 apresentou concentrações ínfimas de DNA e a amplificação foi realizada utilizando apenas uma das extrações.

A amostra OS\_1.1 (Figura 13) foi a única em que foi possível a identificação de todos os marcadores. Para a duplicata com menor concentração de DNA (OS\_1.2) é interessante observar que os 12 marcadores amplificados são aqueles que geram amplicons mais curtos, com até 100pb (Figura 14). Diante de marcadores AS-STRs, AS-InDels e da amelogenina essa amostra também apresentou uma amplificação satisfatória, apesar de ter sido utilizados extrações diferentes na análise dos marcadores autossômicos e sexuais.

A amostra OS\_2 (Figura 15; Figura 16) em que foi utilizado a diáfise femoral na extração também apresentou diferentes concentrações de DNA entre as duplicatas, o que explica a diferença entre o número de marcadores amplificados para cada uma. Por outro lado, a extração realizada utilizando a cabeça femoral (OS\_2.3) não amplificou nenhum marcador. O mesmo resultado se repetiu para os marcadores STRs e InDels autossômicos e amelogenina.

A concentração de DNA encontrada nas amostras OS\_3 (Figura 17) e OS\_4 (Figura 18) não são muito diferentes, mas em OS\_3 foi possível genotipar 7 marcadores a mais. Por outro lado, não foi possível a amplificação da amelogenina em OS\_4, o que pode ser um indício de que há maior degradação dos cromossomos sexuais nessa amostra. Além disso, aqui vale reforçar que para a amostra comercial 9947A os resultados de amplificação com menos do que 0,4ng de DNA não eram mais fidedignos, portanto resultados da amplificação de material biológico degradado abaixo dessa concentração devem ser analisados com cautela, pois podem não representar o genótipo real do indivíduo.

Nessa perspectiva é interessante observar os resultados da amostra OS\_8 que, apesar do alto rendimento da extração, não amplificou nenhum marcador, o que também reforça que a concentração não é o único fator que influencia o sucesso da amplificação.

A PCR consiste em uma reação enzimática e, portanto, sensível à inibidores. Os inibidores de PCR podem originar da própria amostra, como a hemoglobina, ureia e íons de cálcio, e a principal consequência da inibição parcial ou total da PCR é uma diminuição de sua sensibilidade ou a ocorrência de um resultado falso negativo (SCHRADER et al., 2012).

Amostras ósseas em estágios avançados de degradação apresentam em grande quantidade a presença de inibidores como ácido húmicos (JAKUBOWSKA; MACIEJEWSKA; PAWŁOWSKI, 2012). Esse consiste em uma complexa mistura de diferentes ácidos que são produzidos na biodegradação de material orgânico morto. O ácido húmico, mesmo em



concentrações muito baixas, interage com a molécula de DNA e a polimerase impedindo a reação enzimática (SCHRADER et al., 2012).

Entretanto, curiosamente, para OS\_8 foi possível amplificar 17 marcadores AS-STRs mais a amelogenina. Em situações com altos níveis de inibidores, a utilização de kits comerciais apresenta vantagens quando comparadas aos *in house*, pois são munidos de mecanismos de tolerância aos inibidores de PCR, o permite aumentar também as taxas de sucesso de amplificação dessas amostras.

Assim, não foi possível identificar precisamente o motivo da não amplificação de OS\_8, apesar da adequada concentração de DNA e da amplificação de marcadores AS-STRs e amelogenina. Mas uma atenção especial deve ser dada à extração quando o uso de métodos *in house* é necessário, especialmente porque em todos os casos os marcadores AS-STR apresentaram, de forma geral, maior sucesso na análise quando comparados aos X-InDels.

Um estudo comparando três diferentes métodos de extração de DNA em ossos identificou que o protocolo de total desmineralização, utilizado nesse trabalho, é capaz de extrair notavelmente maiores concentrações de DNA do que as demais técnicas, entretanto também foi o método que apresentou maiores níveis de inibidores. Com esses dados os autores concluíram que a total desmineralização consiste na melhor técnica a ser utilizada para a maioria dos casos de extração de DNA em ossos, entretanto, em situações em que a degradação está em estágio avançado, o tratamento do material com hipoclorito de sódio (NaOCl) antes da extração pode ser uma melhor alternativa (JAKUBOWSKA; MACIEJEWSKA; PAWŁOWSKI, 2012).

Na amostra OS\_6 também não foi possível a amplificação de nenhum marcador X-InDel e também não se obteve nenhum alelo dos AS-STR. De tal modo, em dois casos em que o resultado do exame de paternidade foi inconclusivo (OS\_4 e OS\_6), os marcadores X-InDels não puderam contribuir com informações adicionais. Também vale reforçar que para os quatro casos em que nenhum X-InDels foi amplificado (OS\_2.3, OS\_5 cabeça femoral, OS\_6 e OS\_8), a extração e amplificação foram repetidas e não houve alteração no resultado.

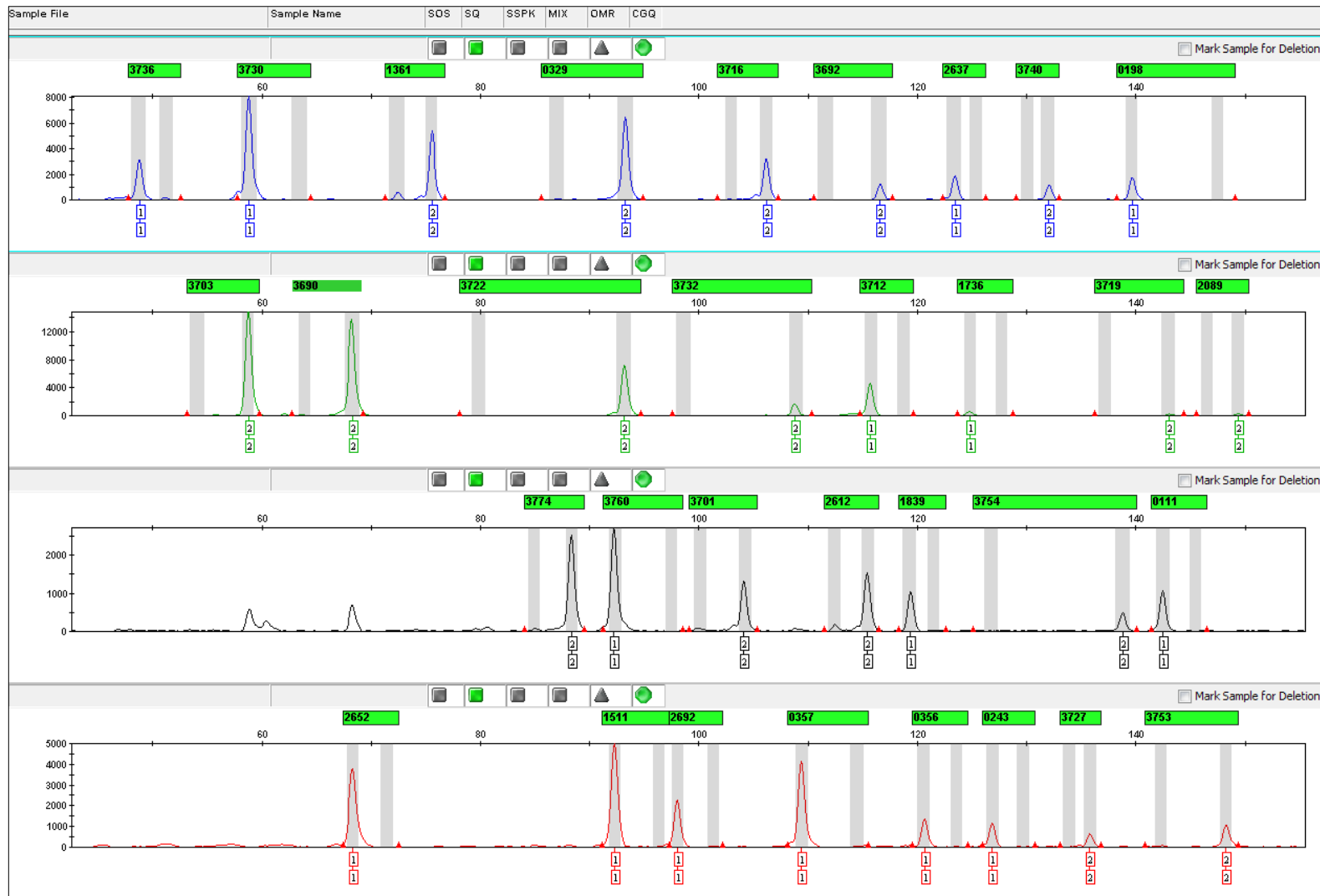
Nas amostras de dente, pode-se observar que foi possível a amplificação de pelo menos 20 marcadores X-InDels (OS\_5-dente; Figura 19 e OS\_7; Figura 20) e que para os STRs somente um marcador em OS\_7 não foi genotipado. O dente consiste em uma fonte valiosa para extração de DNA, pois seu tecido dental rígido age como uma “caixa selada” que protege a polpa, preservando a molécula às condições ambientais extremas. Essa condição colocou os

dentes como um importante objeto de estudo de extração de DNA (GIRISH; RAHMAN; TIPPU, 2010).

Também buscando avaliar a eficiência de amplificação dos marcadores X-InDels em material biológico degradado, Zhang e colaboradores (2015) analisaram 18 X-InDels em amostras degradadas artificialmente, sendo que 12 marcadores são compartilhados com os 32 X-InDels deste estudo. Os autores observaram que mesmo em graus de degradação mais elevados, apenas alelos com 150pb a 200pb não eram amplificados, enquanto que os alelos menores que 150pb eram todos detectados, com picos de 839 RFU (unidade relativa de fluorescência). Este dado reforça a hipótese levantada por Dixon e colaboradores (2005) de que aproximadamente 146pb ficam protegidos das enzimas de degradação.

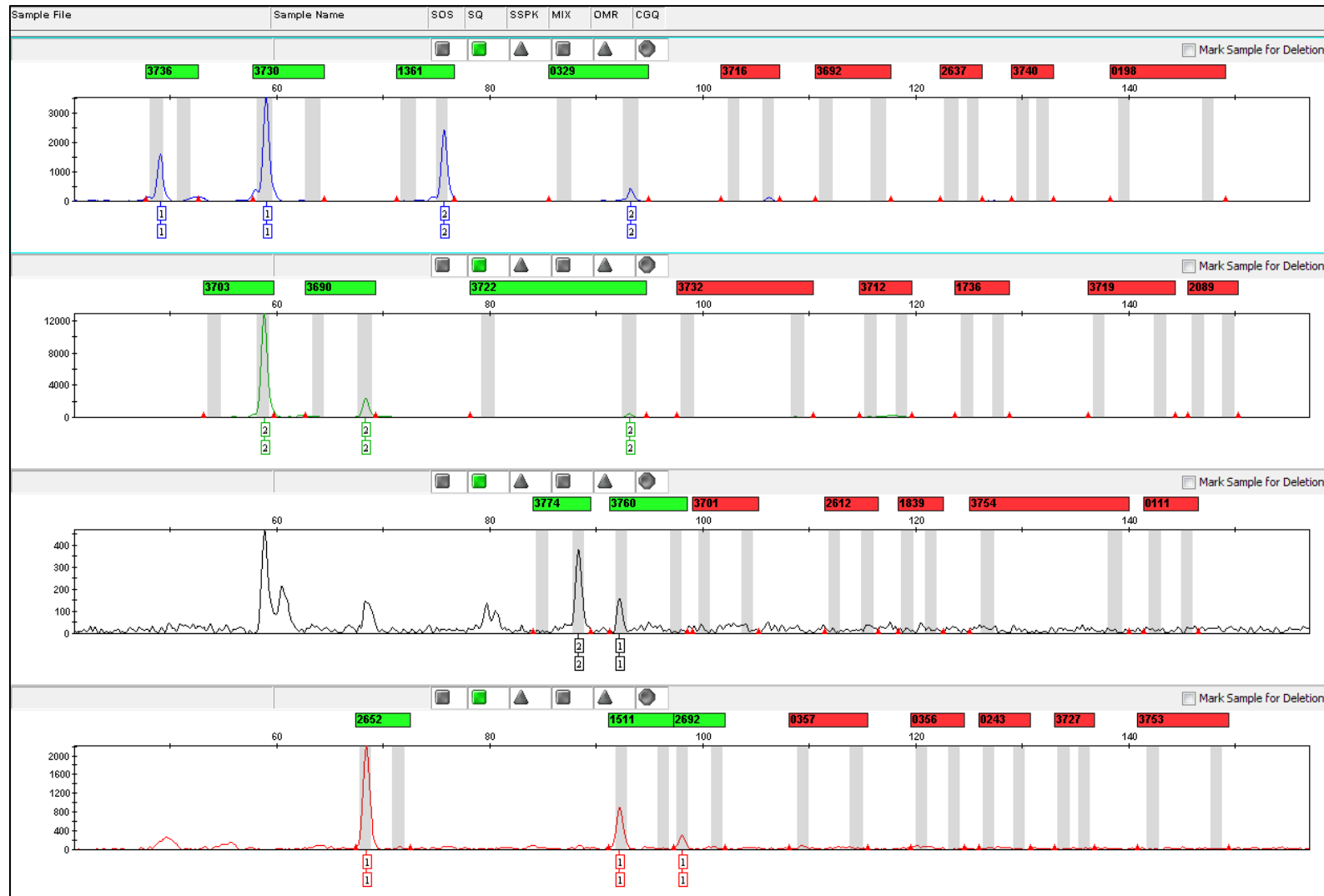
Baseado no exposto é possível concluir que os marcadores 32 X-InDels apresentarem as características necessárias para a análise de amostras forenses e podem ser empregadas com sucesso em casos complexos de parentesco ou identificação de amostras encontradas em locais de crime. Mas não se deve deixar de destacar que algumas dificuldades na amplificação ainda podem ser encontradas a depender das condições ambientais em que a amostra foi exposta.

**Figura 13:** Eletroferograma dos marcadores 32 X-InDels obtidos a partir de DNA extraído de amostra óssea OS\_1.1.



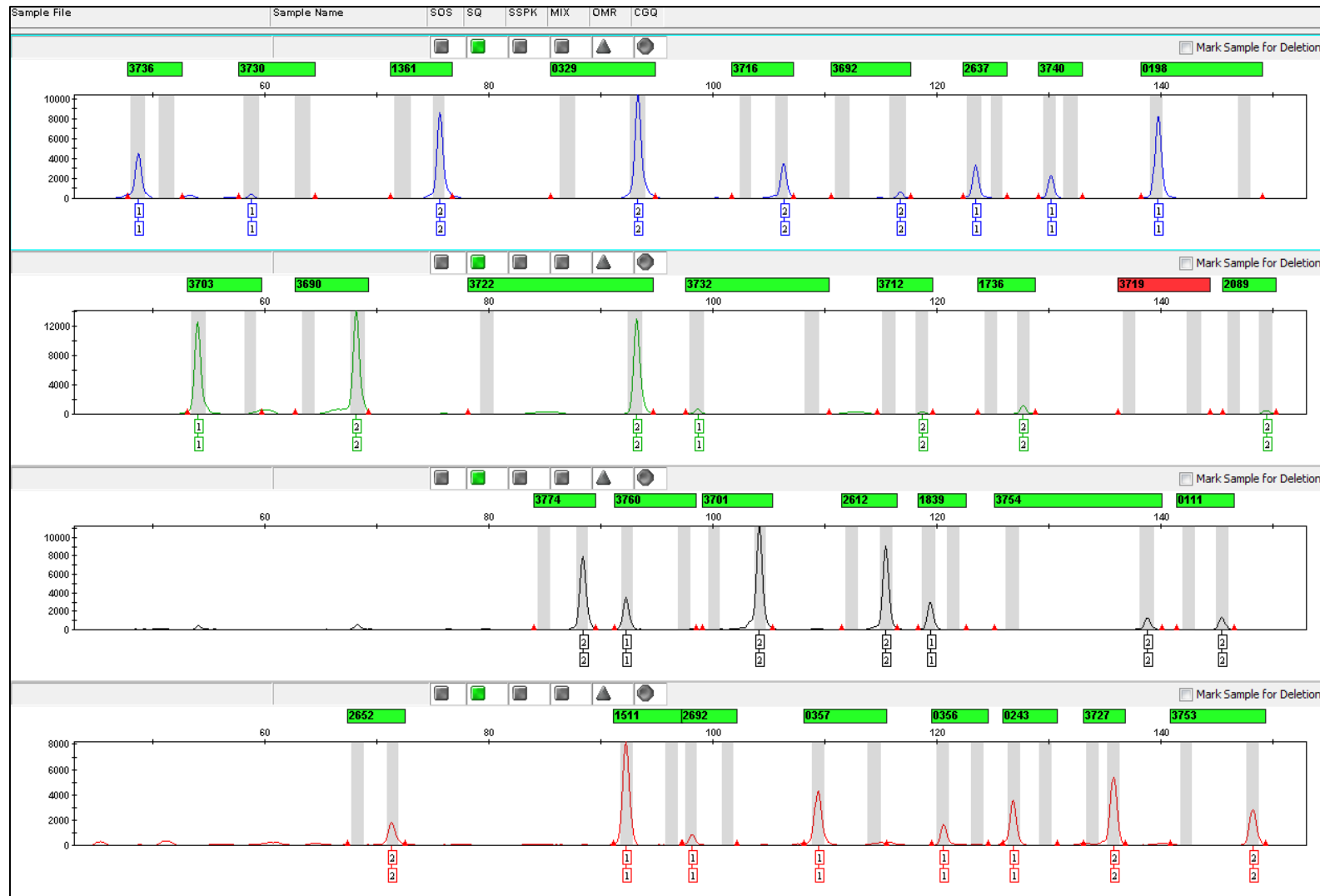
Fonte: Dados da pesquisa.

**Figura 14:** Eletroferograma dos marcadores 32 X-InDels obtidos a partir de DNA extraído de amostra óssea OS\_1.2.



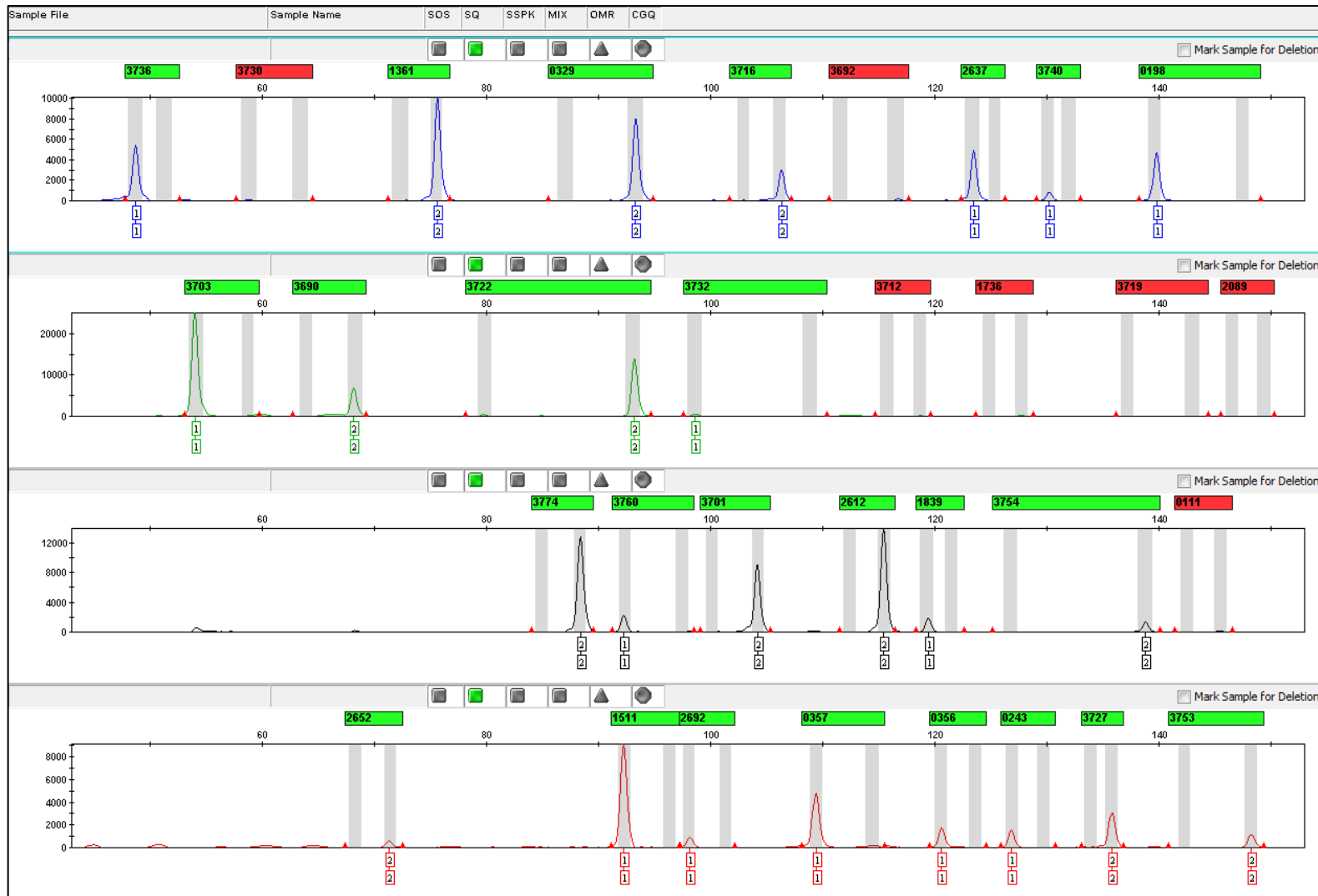
**Fonte:** Dados da pesquisa.

**Figura 15:** Eletroferograma dos marcadores 32 X-InDels obtidos a partir de DNA extraído de amostra óssea OS\_2.1.



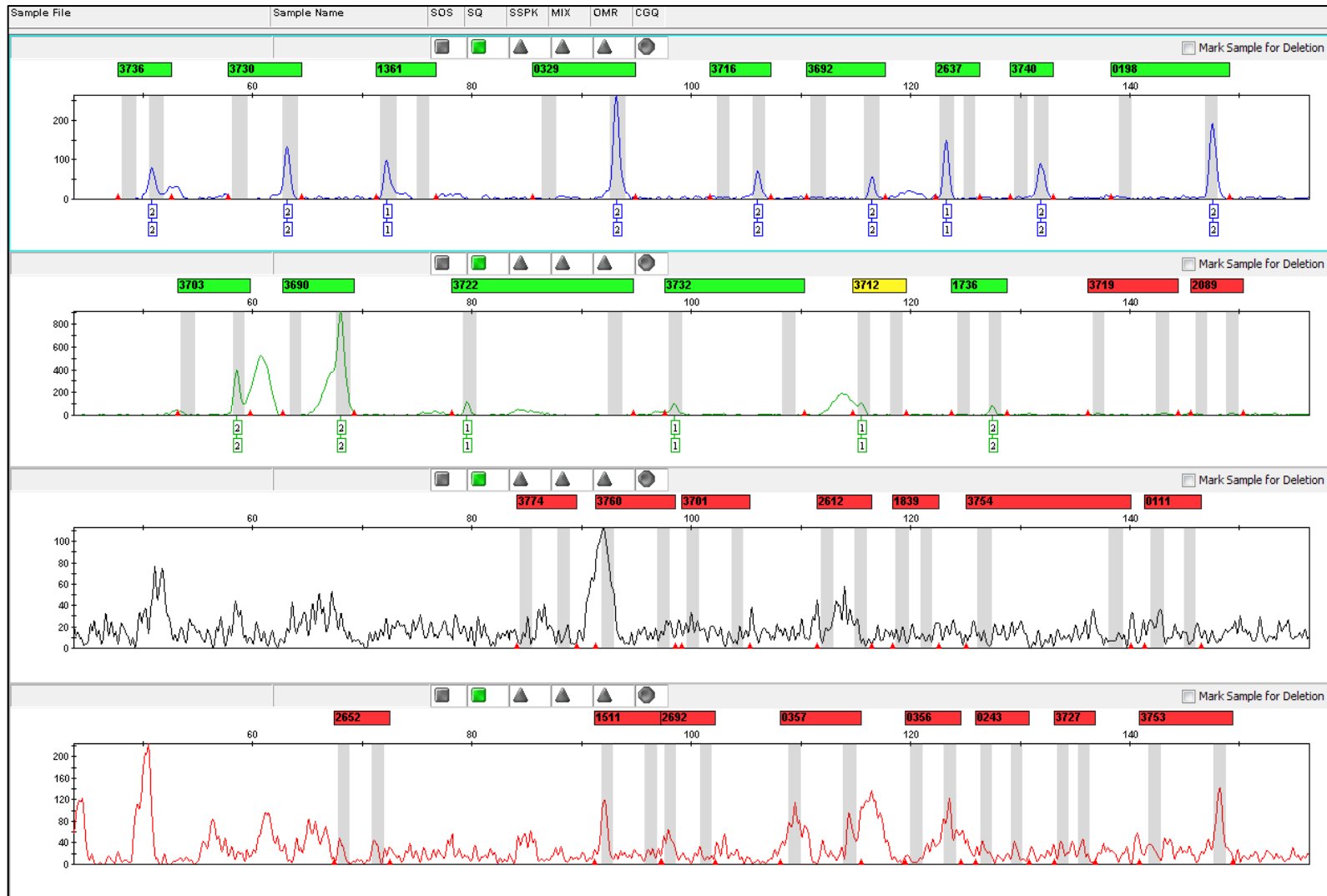
Fonte: Dados da pesquisa.

**Figura 16:** Eletroferograma dos marcadores 32 X-InDels obtidos a partir de DNA extraído de amostra óssea OS\_2.2.



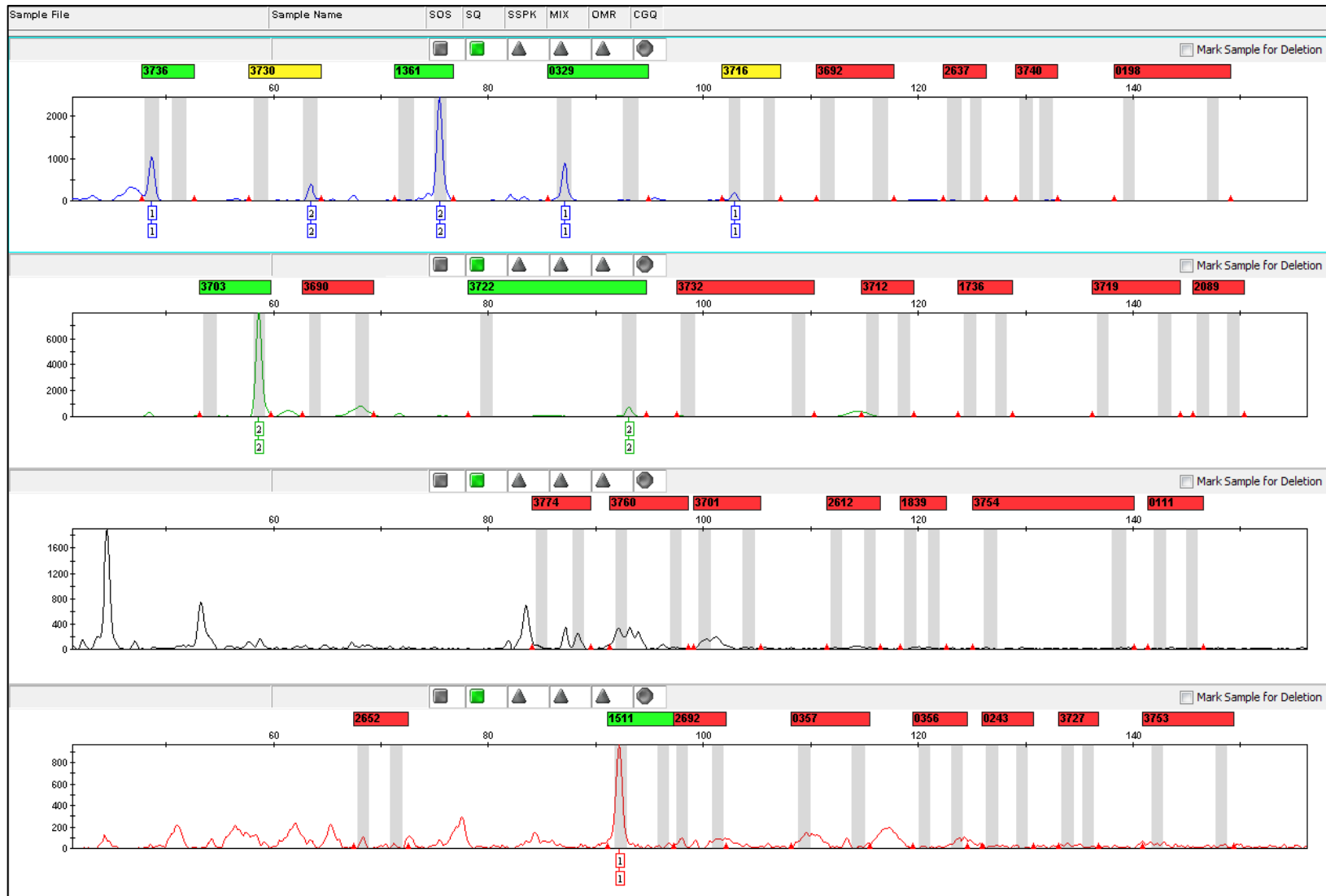
Fonte: Dados da pesquisa.

Figura 17: Eletroferograma dos marcadores 32 X-InDels obtidos a partir de DNA extraído de amostra óssea OS\_3.



Fonte: Dados da pesquisa.

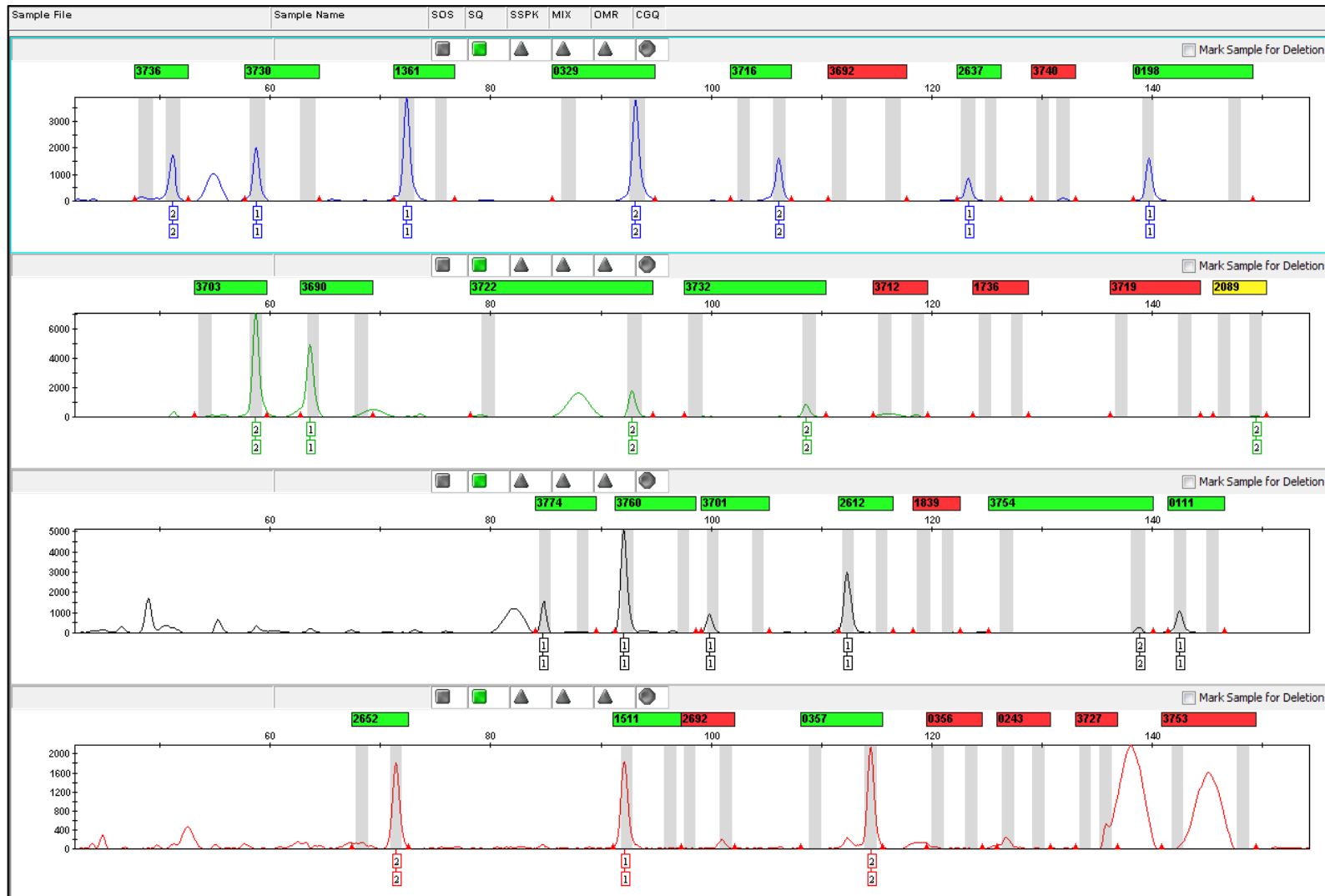
**Figura 18:** Eletroferograma dos marcadores 32 X-InDels obtidos a partir de DNA extraído de amostra óssea OS\_4.



**Fonte:** Dados da pesquisa.

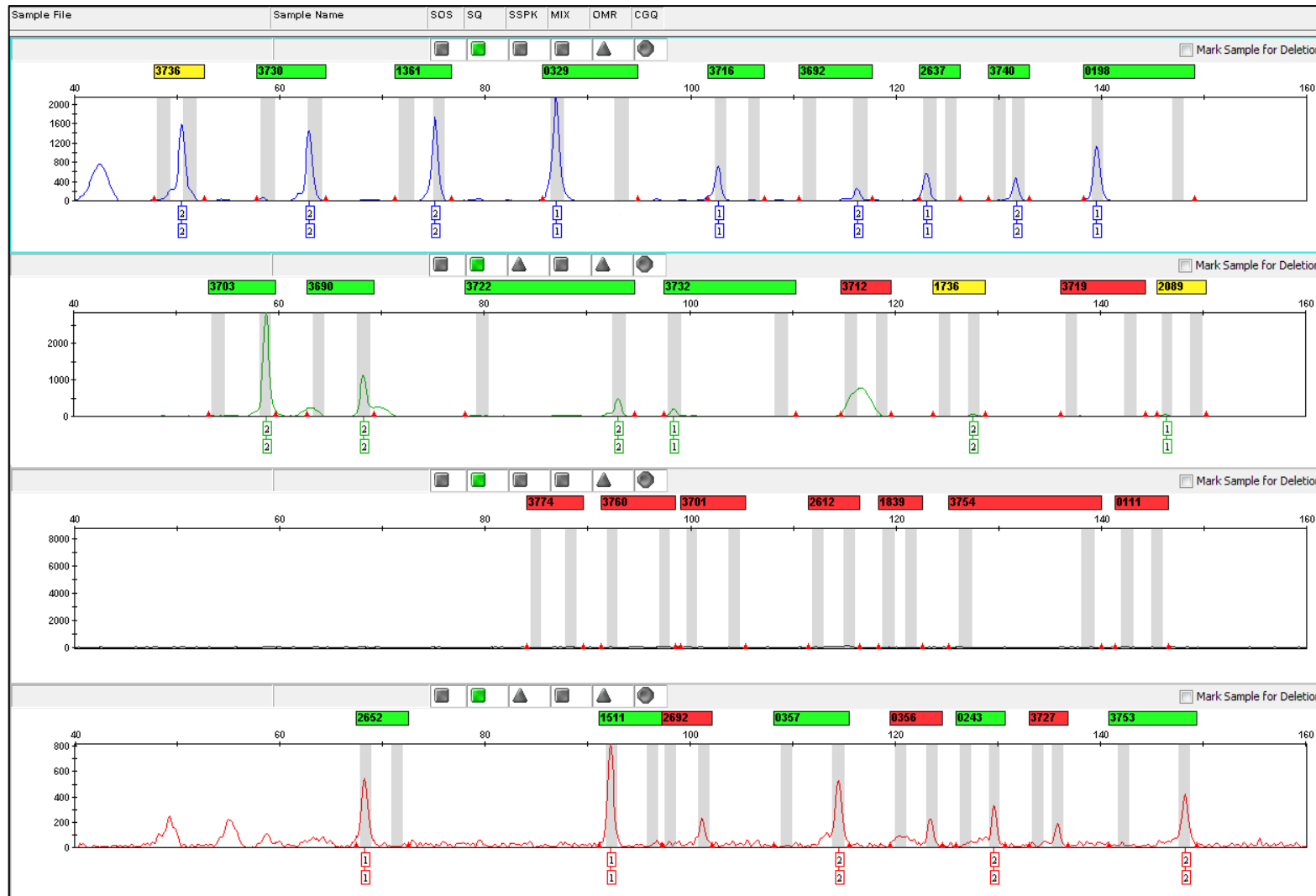


**Figura 19:** Eletroferograma dos marcadores 32 X-InDels obtidos a partir de DNA extraído de amostra óssea OS\_5-dente.



Fonte: Dados da pesquisa.

**Figura 20:** Eletroferograma dos marcadores 32 X-InDels obtidos a partir de DNA extraído de amostra óssea OS\_7.



**Fonte:** Dados da pesquisa.

# Conclusões

O conhecimento acerca dos marcadores de inserção/deleção no cromossomo X ainda precisa ser ampliado, uma vez que na literatura há pouco material disponível sobre o assunto, mas resultados recentes já demonstram seu potencial como método alternativo para a análise de amostras forenses. Com os resultados obtidos neste trabalho pode-se concluir que:

- A metodologia de amplificação dos 32 X-InDels pôde ser padronizada visando a redução de custos, sem perder sua eficiência e especificidade. Essa modificação trará importantes benefícios aos laboratórios, uma vez que as técnicas de biologia molecular tendem a ser dispendiosas;
- Os 32 X-InDels mostraram-se polimórficos na população de São Paulo, o que significa que consistem em marcadores genéticos bastante informativos;
- MID1361 apresentou-se em desequilíbrio de Hardy-Weinberg, possivelmente devido à presença de alelo nulo, e uma variante alélica foi identificada associada ao seu alelo curto, o que sugere a necessidade de uma atenção especial quando esse marcador for utilizado na investigação de relações biológicas;
- Para todos os parâmetros de eficiência forenses, esse painel se mostrou altamente discriminativo na população de São Paulo;
- Na maioria das comparações, valores significativos de distância genética foram obtidos entre São Paulo e as outras populações. Observou-se maior proximidade desse estado com três departamentos colombianos (Antioquia, Norte de Santander e Santander), seguido dos europeus de Portugal, Espanha e Dinamarca. A população mais distante foi Groelândia, na América do Norte;
- A proporção de ancestralidade identificada foi similar ao descrito na literatura, com 41,8% de componente europeu, 31,6% de africano e 26,6% de nativo-americano, o que reafirma a importante contribuição do cromossomo X de nativo-americanos para a formação da população de São Paulo;
- O padrão de transmissão esperado entre pai-mãe/filha foi o observado na análise de segregação, confirmando a baixa taxa de mutação para marcadores InDels no cromossomo X e sua eficiência na investigação de paternidade, especialmente quando há incompatibilidades na análise de STRs autossômicos;
- Para DNA em baixa concentração e/ou degradado, os 32 X-InDels apresentaram as características necessárias para a análise de amostras forenses, mas algumas dificuldades na amplificação ainda podem ser encontradas a depender das condições ambientais às quais o material foi exposto.

# Referências

- ALMEIDA PRADO DE OLIVEIRA E SOUZA, M.L. et al. Population data of 16 autosomal STR loci of the Powerplex ESX 17 System in a Brazilian Population from the State of São Paulo. **Forensic Sci Int Genet.** v. 11. p. 15-7. 2014.
- 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.
- AMBROSIO, I.B.; BRAGANHOLI, D.F.; CICARELLI, R.M.B. **Aplicações forenses do DNA: fundamentos, métodos e limitações.** In: RESENDE, R. R. (Org.). *Biotecnologia aplicada à saúde: fundamentos e aplicações.* 1ed. São Paulo: Blucher, v. 3, p. 845- 73. 2015.
- ANDREWS, C. The Hardy-Weinberg Principle. **Nature Education Knowledge.** v. 3. n. 10. p. 65. 2010.
- BONFERRONI, C.E. **Publicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze.** v. 8. p. 3-62. 1936.
- BOTSTEIN, D. et al. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. **Am J Hum Genet.** v. 32. p. 314–31. 1980.
- BUENO, B.P.S. Dilatação dos confins: caminhos, vilas e cidades na formação da Capitania de São Paulo (1532-1822). **An. mus. paul.** v. 17, n. 2, p. 251-94. 2009.
- 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.
- CAMPBELL, C.D.; EICHLER, E.E. Properties and rates of germline mutations in humans. **Trends Genet.** v. 29. n. 10. p. 575-84. 2013.
- 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.
- CHAKRABORTY, R. **Analysis of Genetic Structure of Populations: Meaning, Methods, and Implications.** p. 189-206. In: MAJUMDER, P.P. (eds) *Human Population Genetics.* Springer, Boston, MA. 1993.
- CHARLESWORTH, B. The evolution of sex chromosomes. **Science.** v. 251. n. 4997. p. 1030-3. 1991.
- CHARLESWORTH, D.; CHARLESWORTH, B.; MARAIS, G. Steps in the evolution of heteromorphic sex chromosomes. **Heredity (Edinb).** v. 95. n. 2. p. 118-28. 2005.

- CUNHA, J.M.P. da. et al. **Dinâmica Migratória no Estado de São Paulo**. In: HOGAN, D.J. et al. Migração e ambiente em São Paulo: aspectos relevantes da dinâmica recente. Campinas: Núcleo de Estudos de População/UNICAMP, 2000. 518p.
- 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.
- DEDECCA, C.S. et al. **Estado de São Paulo: regiões metropolitanas e polos regionais**. v. 1. In: Por dentro do Estado de São Paulo. Campinas: Núcleo de Estudos de População/UNICAMP. 2010. 172p.
- 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**. v. 41. n. 10. p. 6429-32. 2014.
- 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.
- ELLEGREN, H. Sex-chromosome evolution: recent progress and the influence of male and female heterogamety. **Nat Rev Genet**. v. 12. n. 3. p. 157-66. 2011.
- 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 X-chromosome markers: Indels, Alu insertions and STRs. **Forensic Sci Int Genet**. pii: S1872-4973. n. 17. p. 30196-5. 2017.
- FERREIRA PALHA, T.J.B. 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. 10-4. 2015.
- 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.
- GE, J.; EISENBERG, A.; BUDOWLE, B. Developing criteria and data to determine best options for expanding the core CODIS loci. **Investig Genet**. v. 3. 2012.
- GILL, P.; JEFFREYS, A.J.; WERRETT, D.J. Forensic application of DNA 'fingerprints'. **Nature**. v. 318. n. 6046. p. 577-9. 1985.

- GIRISH, K.L.; RAHMAN, F.S.; TIPPU, S.R. Dental DNA fingerprinting in identification of human remains. **J Forensic Dent Sci.** v. 2. n. 2. p. 63–68. 2010.
- GJERTSON, D.W. et al. ISFG: Recommendations on biostatistics in paternity testing. **Forensic Sci Int Genet.** v. 1. n. 3-4. p. 223-31. 2007.
- GRAVES, J.A. Sex chromosome specialization and degeneration in mammals. **Cell.** v. 124. n. 5. p. 901-14. 2006.
- GUSMÃO, L. 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. n. 3. p. 227-34. 2009.
- HAMOSH, A. et al. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. **Nucleic Acids Res.** v. 33. n. Database Issue. p. D514–D517. 2005.
- HARTL, D.L.; CLARK, A.G. **Principles of population genetics.** 4ª ed. Estados Unidos da América: Sinauer Associates. 2006.
- HEARNE, C.M.; TODD, J.A. Tetranucleotide repeat polymorphism at the HPRT locus. **Nucleic Acids Res.** v. 19. n. 19. p. 5450. 1991.
- HELENA MANGS, A.; MORRIS, B.J. The Human Pseudoautosomal Region (PAR): Origin, Function and Future. **Curr Genomics.** v. 8. n. 2. p. 129-36. 2007.
- HOLMES, C.L., RUSSELL, J.A., WALLEY, K.R. Genetic polymorphisms in sepsis and septic shock: role in prognosis and potential for therapy. **Chest.** v. 124. n. 3 p. 1103-15. 2003.
- 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.
- IBGE. Instituto Brasileiro de Geografia e Estatística. **Censo Demográfico de 2010.** Rio de Janeiro: IBGE; 2010.
- INÁCIO, A. et al. Study of InDel genetic markers with forensic and ancestry informative interest in PALOP's immigrant populations in Lisboa. **Int J Legal Med.** v. 131. n. 3. p. 657-60. 2017.
- ISRAR, M. et al. Development and characterization of a new 12-plex ChrX miniSTR system. **Int J Legal Med.** v. 128. n. 4. p. 595-8. 2014.
- JAKUBOWSKA, J.; MACIEJEWSKA, A. PAWŁOWSKI, R. Comparison of three methods of DNA extraction from human bones with different degrees of degradation. **Int J Legal Med.** v. 126. n. 1. p. 173–78. 2012.
- JEFFREYS, A.J.; WILSON, V.; THEIN, S.L. Individual-specific 'fingerprints' of human DNA. **Nature.** v. 316. n. 6023. p. 76-9. 1985.



- KARKI, R. et al. Defining “mutation” and “polymorphism” in the era of personal genomics. *BMC Med Genomics*. v. 8. p. 37. 2015.
- KRÜGER, J. et al. Zur Verwendung der sauren Erythrocytenphosphatase bei der Vaterschaftsbegutachtung. *Dtsch Z Gerichtl Med*. v. 64. n. 2. p. 127–46. 1968.
- 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. 2012.
- MARTINEZ, J. et al. Mutation rates for 20 STR loci in a population from São Paulo state, Southeast, Brazil. *Ann Hum Biol*. v. 44. n. 7. p. 659-62. 2017.
- MARTINS, J.A. et al. X-chromosome genetic variation in São Paulo State (Brazil) population. *Ann Hum Biol*. v. 37. n. 4. p. 598-603. 2010.
- 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.
- MEIRMANS, P.G.; HEDRICK, P.W. Assessing population structure: F(ST) and related measures. *Mol Ecol Resour*. v. 11. n. 1. p. 5-18. 2011.
- 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.
- MONTALI, L. et al. **Regiões metropolitanas**. v. 2. In: Por dentro do Estado de São Paulo. Campinas: Núcleo de Estudos de População/UNICAMP. 2010. 264p.
- MONTEIRO, J.M. **Negros da terra: índios e bandeirantes nas origens de São Paulo**. 2ª reimpressão. São Paulo: Companhia das Letras. 1994. 300p.
- MOTA, C.G. São Paulo: exercício de memória. *Estud. av*. v. 17, n. 48, p. 241-63. 2003.
- MOURA, R.R. et al. Meta-analysis of Brazilian genetic admixture and comparison with other Latin America countries. *Am J Hum Biol*. v. 27. n. 5. p. 674-80. 2015.
- NEI, M.; ROYCHOUDHURY, A.K. Sampling variances of heterozygosity and genetic distance. *Genetics*. v. 76. p. 379–90. 1974.
- NATIONAL RESEARCH COUNCIL. **The evaluation of forensic DNA evidence**. 2ª ed. Washington: National Academy Press, 1996.
- NOTHNAGEL, M. et al. Collaborative genetic mapping of 12 forensic short tandem repeat (STR) loci on the human X chromosome. *Forensic Sci Int Genet*. v. 6. n. 6. p. 778-84. 2012.
- OHNO, S. **Sex chromosomes and sex-linked genes**. Springer-Verlag Berlin. 1967. 192p.

PEPINSKI, W. et al. Population genetics of 30 Indels in populations of Poland and Taiwan. **Mol Biol Rep.** v. 40. n. 7. p. 4333–8. 2013.

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. **Bridging the gap between SNPs and STRs: Insertion deletion polymorphisms in forensic genetics; principles and applications.** Tese de Doutorado. Faculdade de Medicina e Odontologia, Universidade de Santiago de Compostela. 2011.

PEREIRA, V. et al. Genetic characterization of Somali and Iraqi populations using a set of 33 X-chromosome Indels. **Forensic Sci Int Genet.** 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 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. 2017.

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

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 X-chromosomal 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.

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

RESQUE, R.L. et al. Estimates of interethnic admixture in the Brazilian population using a panel of 24 X-Linked Insertion/Deletion markers. **Am J Hum Biol.** v. 22. n. 6. p. 849-52. 2010.

ROHLAND, N.; HOFREITER M. Comparison and optimization of ancient DNA extraction. **Biotechniques.** v. 42. n. 3. p. 343-52. 2007.

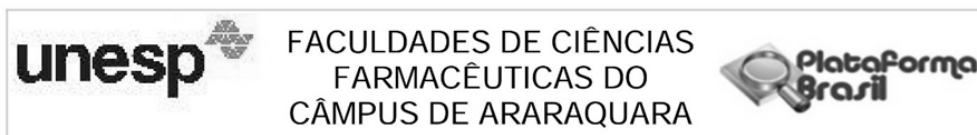
ROSS, M.T. et al. The DNA sequence of the human X chromosome. **Nature.** v. 434. n. 7031. p. 325-37. 2005.

SAIZ, M. et al. Allelic frequencies and statistical data from 30 Indel loci in Uruguayan population. **Forensic Sci Int Genet.** v. 9. p. e27-9. 2014.

- SANTOS-LOPES, S.S. et al. A Worldwide Phylogeography for the Human X Chromosome. **PLoS ONE**. v. 2. n. 6. p. e557. 2007.
- SÃO PAULO. Governo do Estado de São Paulo. **Conheça São Paulo: História**. Disponível em: <<http://www.saopaulo.sp.gov.br/conhecasp/historia>>. Acesso em: 19 out 2017.
- SCHAFFNER, S.F. The X chromosome in population genetics. **Nat Rev Genet**. v. 5. n. 1. p. 43-51. 2004.
- SCHRADER, C., et al. PCR inhibitors - occurrence, properties and removal. **J Appl Microbiol**. v. 113. n. 5 p. 1014-26. 2012.
- SEONG, K.M. et al. Population genetics of insertion-deletion polymorphisms in South Koreans using Investigator DIPplex kit. **Forensic Sci Int Genet**. v. 8. n. 1. p. 80-3. 2014.
- 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.
- SKALETSKY, H. et al. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. **Nature**. v. 423. n. 6942. p. 825-37. 2003.
- 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.
- THE 1000 GENOMES PROJECT CONSORTIUM. A global reference for human genetic variation. **Nature**. v. 526. n. 7571. p. 68-74. 2015.
- 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.
- TOMAS, C. et al. Thirty autosomal insertion-deletion polymorphisms analyzed using the Investigator® DIPplex Kit in populations from Iraq, Lithuania, Slovenia, and Turkey. **Forensic Sci Int Genet**. v. 25. p. 142-4. 2016.

- TORRES, S.R. et al. Population genetic data and forensic parameters of 30 autosomal Indel markers in Santa Catarina State population, Southern Brazil. **Mol Biol Rep.** v. 41; n. 8. p. 5429-33. 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.** v. 10. n. 4. p. 506-13. 1991.
- WANG, S. et al. Genetic variation and population structure in native Americans. **PLoS Genet.** v. 3. n. 11. p. e185. 2007.
- WANG, S. et al. Geographic patterns of genome admixture in Latin American Mestizos. **PLoS Genet.** v. 4. n. 3. p. e1000037. 2008.
- WANG, L. et al. A comparative study of insertion/deletion polymorphisms applied among Southwest, South and Northwest Chinese populations using Investigator® DIPplex. **Forensic Sci Int Genet.** v. 21. p. 10-4. 2016.
- WATSON, J.D.; CRICK, F.H. The structure of DNA. **Cold Spring Harb Symp Quant Biol.** v. 18. p. 123-31. 1953.
- WEBER, J.L. et al. Human diallelic insertion/deletion polymorphisms. **Am J Hum Genet.** v. 71. n. 4. p. 854-62. 2002.
- WIJCHERS, P.J.; FESTENSTEIN, R.J. Epigenetic regulation of autosomal gene expression by sex chromosomes. **Trends Genet.** v. 27. n. 4. p. 132-40. 2011.
- WRIGHT, S. The genetical structure of populations. **Ann Eugen.** v. 15. n. 4. p. 323-54. 1951.
- ZARRABEITIA, M.T. et al. Study of six X-linked tetranucleotide microsatellites: population data from five Spanish regions. **Int J Legal Med.** v. 120. n. 3. p. 147-50. 2006.
- ZHANG, S. et al. Developmental validation of an X-Insertion/Deletion polymorphism panel and application in HAN population of China. **Sci Rep.** v. 5. p. 18336. 2015.
- 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 1 – Aprovação do Comitê de Ética em Pesquisa com Seres Humanos



### PARECER CONSUBSTANCIADO DO CEP

#### DADOS DA EMENDA

**Título da Pesquisa:** Polimorfismos de inserção/deleção no cromossomo X: análise de 32 marcadores na população do Estado de São Paulo (Brasil)

**Pesquisador:** Juliana Martinez

**Área Temática:**

**Versão:** 5

**CAAE:** 36321514.0.0000.5426

**Instituição Proponente:** Faculdade de Ciências Farmacêuticas do Câmpus de Araraquara da UNESP

**Patrocinador Principal:** Financiamento Próprio

#### DADOS DO PARECER

**Número do Parecer:** 1.425.912

#### Apresentação do Projeto:

O relatório parcial está apresentado de forma condizente.

#### Objetivo da Pesquisa:

Relatório parcial, os objetivos continuam os mesmos.

#### Avaliação dos Riscos e Benefícios:

Não houve modificação.

#### Comentários e Considerações sobre a Pesquisa:

O relatório de pesquisa parcial apenas exclui o Hemocentro de Botucatu para coleta de amostras de sangue, que será realizada como coleta voluntária entre alunos e funcionários da UNESP Araraquara.

#### Considerações sobre os Termos de apresentação obrigatória:

Os termos não foram modificados.

#### Recomendações:

#### Conclusões ou Pendências e Lista de Inadequações:

O Comitê de Ética em Pesquisa da Faculdade de Ciências Farmacêuticas, reunido no dia 24 de fevereiro de 2016, analisou e aprovou a emenda interposta no projeto de pesquisa em questão. Os

**Endereço:** Rodovia Araraquara Jau, km 1

**Bairro:** Campus Universitário

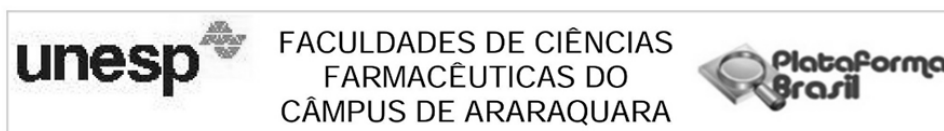
**CEP:** 14.801-902

**UF:** SP

**Município:** ARARAQUARA

**Telefone:** (16)3301-6897

**E-mail:** sta@fcfar.unesp.br



Continuação do Parecer: 1.425.912

próximos relatórios parciais deverão ser entregues em: 1) AGOSTO de 2016; 2) FEVEREIRO de 2017 e 3) AGOSTO de 2017, e o Relatório Final juntos aos Termos de Consentimento Livre e Esclarecido (originais e assinados em todas as folhas) deverão ser entregues em FEVEREIRO de 2018.

**Considerações Finais a critério do CEP:**

**Este parecer foi elaborado baseado nos documentos abaixo relacionados:**

| Tipo Documento  | Arquivo  | Postagem               | Autor            | Situação |
|---|--|------------------------|------------------|----------|
| Informações Básicas do Projeto                            | PB_INFORMAÇÕES_BÁSICAS_591515E2.pdf  | 21/09/2015<br>16:44:30 |                  | Aceito   |
| Outros  | Relatorio_Set_2015.pdf   | 21/09/2015<br>16:40:06 | Juliana Martinez | Aceito   |
| TCLE / Termos de Assentimento / Justificativa de Ausência | Termo de Consentimento Livre e Esclarecido - INDEL 16.01.15 modificado cabeçalho.doc | 08/06/2015<br>22:54:36 |                  | Aceito   |
| Outros  | Declaração de Colaboração - SJ Rio Preto.pdf   | 08/06/2015<br>19:59:40 |                  | Aceito   |
| TCLE / Termos de Assentimento / Justificativa de Ausência | Termo de Consentimento Livre e Esclarecido - INDEL 16.01.15.doc                      | 16/01/2015<br>15:27:46 |                  | Aceito   |
| Projeto Detalhado / Brochura Investigador                 | Projeto Doutorado Indel - UNESP 21.12.14.doc   | 21/12/2014<br>17:17:21 |                  | Aceito   |
| Outros  | NOTA DE ESCLARECIMENTO CEP.docx  | 11/12/2014<br>16:26:53 |                  | Aceito   |
| Outros  | Declaração CEP.pdf   | 16/09/2014<br>14:26:15 |                  | Aceito   |
| Projeto Detalhado / Brochura Investigador                 | Projeto Doutorado Indel - UNESP final.doc  | 16/09/2014<br>14:25:40 |                  | Aceito   |
| Folha de Rosto  | Folha de Rosto Indels - CEP.pdf  | 16/09/2014<br>14:23:47 |                  | Aceito   |

**Situação do Parecer:**

Aprovado

**Necessita Apreciação da CONEP:**

Não

**Endereço:** Rodovia Araraquara Jau, km 1  
**Bairro:** Campus Universitário **CEP:** 14.801-902  
**UF:** SP **Município:** ARARAQUARA  
**Telefone:** (16)3301-6897 **E-mail:** sta@fctar.unesp.br



Continuação do Parecer: 1.425.912

ARARAQUARA, 26 de Fevereiro de 2016

---

**Assinado por:**  
**Adriano Mondini**  
**(Coordenador)**

**Endereço:** Rodovia Araraquara Jau, km 1  
**Bairro:** Campus Universitário **CEP:** 14.801-902  
**UF:** SP **Município:** ARARAQUARA  
**Telefone:** (16)3301-6897 **E-mail:** sta@fcar.unesp.br

# CAPÍTULO II



# Genetic characterization of an X-STR decaplex system in the State of Mato Grosso, Brazil: distribution, forensic efficiency and population structure

Joyce A. Martins<sup>1</sup> · Denise P. Martins<sup>1</sup> · Camila I. F. Oliveira-Brancati<sup>1</sup> · Juliana Martinez<sup>2</sup> · Regina M. B. Cicarelli<sup>2</sup> · Dorotéia R. S. Souza<sup>1</sup>

Received: 7 October 2016 / Accepted: 17 March 2017  
© Springer-Verlag Berlin Heidelberg 2017

**Abstract** Studies with X-STR loci show population genetic substructure, which makes necessary the characterization of such markers in the different geographical and/or ethnic populations. Therefore, this study assessed the distribution and forensic efficiency of an X-STR decaplex system in the population of the State of Mato Grosso, as well as analysed the population structure of this State based on the aforementioned system. All X-STR markers were in Hardy–Weinberg equilibrium and linkage equilibrium, and the DXS6809 was the most informative marker. The power of discrimination value in females and males was 0.99999999995 and 0.9999994, respectively. Analysis of molecular variance indicated 1.10% ( $p < 0.00001$ ) of heterogeneity among Europeans, Africans, Brazilians and other Latin Americans, and in relation to such groups, the population of the State of Mato Grosso showed lower genetic variation when compared with the Brazilian group ( $-0.10\%$ ,  $p = 0.67327$ ). The genetic distance analysis showed lower values of  $F_{ST}$  ( $0.0004 \leq F_{ST} \leq 0.00331$ ), with non-significant  $p$  value ( $p > 0.00024$ ), between the populations of Mato Grosso and Mato Grosso do Sul, Paraná and the Southeast region of Brazil (except for one sample of Rio

de Janeiro).  $F_{ST}$  values with significant  $p$  values ( $p \leq 0.00024$ ) were obtained between the population of Mato Grosso and Iberian, African and some Latin American populations. The X-STR decaplex system proved to be extremely useful in the population of the State of Mato Grosso, and the data obtained does not show the need for a specific forensic database for this State in relation to the Brazilian populations compared in this study, except for population of Rio de Janeiro.

**Keywords** X-STR · Population structure · Mato Grosso · Brazil

## Introduction

The genotype analysis of X-chromosomal short tandem repeat (X-STR) markers is broadly used in some complex cases of biological relationships [1]. This is due to the inheritance pattern of the X chromosome, by which males can only inherit the mother's genetic material and transmit it only to their daughters. Thus, in some investigations, X-STRs have a higher mean exclusion chance (MEC), as well as a more efficient haplotype typing than autosomal markers [1]. If a mother–son relationship is being questioned, for example, X-STR analysis allows for a nearly twofold greater power of exclusion than that obtained for autosomal STRs (AS-STRs) with moderate heterozygosity [2]. In addition, X-STRs can differentiate pedigrees involving two non-inbred individuals (half-siblings, grandparent–grandchild and avuncular), as well as some situations of incest, which are indistinguishable by the analysis of unlinked autosomal markers [3]. Among other applications, there are tests in which the alleged father is absent or unavailable, which can be more easily solved with X-STRs than AS-STRs if the alleged paternal grandmother is analysed. In her absence, her profile


**Electronic supplementary material** The online version of this article (doi:10.1007/s00414-017-1578-6) contains supplementary material, which is available to authorized users.

✉ Joyce A. Martins  
joyce\_apa@hotmail.com

<sup>1</sup> Research Centre for Biochemistry and Molecular Biology at the Medical School of São José do Rio Preto—FAMERP, Av. Brigadeiro Faria Lima, 5416, São José do Rio Preto, SP 15090-000, Brazil

<sup>2</sup> School of Pharmaceutical Sciences, Laboratory of Paternity Investigation, Univ Estadual Paulista—UNESP, Rodovia Araraquara-Jaú, km 01, Araraquara, SP 14801-902, Brazil

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

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

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

### ABSTRACT

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

### ARTICLE HISTORY

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

### KEY WORDS

Short tandem repeats; mutation rates; paternity testing; Brazil

### Introduction

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

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

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

Figueiredo et al., 2015; Manta et al., 2013). Therefore, it is important to know the mutation rates in different Brazilian populations.

### Materials and methods

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

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

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

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



Contents lists available at ScienceDirect

Forensic Science International: Genetics Supplement Series

journal homepage: [www.elsevier.com/locate/fsigss](http://www.elsevier.com/locate/fsigss)

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

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

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

### ARTICLE INFO

**Keywords:**  
Ancestry  
Southeast region  
Brazil  
Genetic markers

### ABSTRACT

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

### 1. Introduction

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

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

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

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

In 1763 the capital of the colony is transferred to Rio de Janeiro. In

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

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

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

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

\* Corresponding author.

E-mail address: [danielobraganholi@hotmail.com](mailto:danielobraganholi@hotmail.com) (D.F. Braganholi).

<http://dx.doi.org/10.1016/j.fsigss.2017.09.174>

Received 22 August 2017; Accepted 19 September 2017

1875-1768/ © 2017 Elsevier B.V. All rights reserved.



Contents lists available at ScienceDirect

Forensic Science International: Genetics Supplement Series

journal homepage: [www.elsevier.com/locate/fsigss](http://www.elsevier.com/locate/fsigss)

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

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

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

### ARTICLE INFO

#### Keywords:

STRs  
D22S1045  
D2S441  
D10S1248  
São Paulo State  
Brazil

### ABSTRACT

São Paulo state is one of the Brazilian states with the highest cultural and ethnic miscegenation, mainly of Italian and Portuguese descendants, but also strongly influenced by Amerindians, Africans, Spanish, Germans, Arabs and Japanese, as well as the high number of immigrants from the northeast region. In this work we analyzed statistical data of three STRs markers (D22S1045, D2S441 and D10S1248) that are used in our laboratory's forensic routine for paternity investigations. The use of these data will be extremely important for the laboratory to perform statistical analyzes for the conclusion of the reports. We analyzed 221 samples obtained from unrelated individuals born in the São Paulo state, which the allele frequencies and statistical parameters were estimated with PowerStats version 12 (Promega Corp.). The power of discrimination (PD) and power of exclusion (PE) for the D10S1248 marker were 0.914 and 0.576 respectively and the allele 14 showed the highest frequency (0.294) and 8, 10, 18 and 19 alleles presented the lowest frequency (0.002). PD and PE of the D2S441 marker were 0.907 and 0.551 respectively; the highest frequency allele was 11 (0.312) and the lowest frequency alleles were 9 and 12.3 (0.002). The marker D22S1045 showed PD 0.890 and PE 0.489 with a higher frequency allele 16 (0.373) and lowest frequency allele 8 (0.002). Also, significant differences were found between the São Paulo population and the other populations for the three markers analyzed.

### 1. Introduction

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

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

### 2. Materials and methods

A total of 221 samples obtained from unrelated individuals born in the São Paulo state were collected on FTA® classic cards (GE Healthcare, Life Sciences). Samples were amplified with the PowerPlex® Fusion 6C System (Promega), following the User Guide [3]. Capillary electrophoresis and detection were performed on the 3500 Genetic Analyzer using POP-4® polymer (Applied Biosystems). Genotypes were assigned using the GeneMapper ID v1.2.1 software (Applied Biosystems). Allele frequencies and statistical parameters were estimated with PowerStats version 12 (Promega Corp.) [4]. Allelic frequencies for each locus in São Paulo (Southeast Brazil) population were compared with other populations: European, African [5], Colombian region [6] and the population of the Rio Grande do Sul' state (Southern Brazil) [7]. Pairwise genetic distances were calculated between populations using Arlequin v3.5 [8] software. Our laboratory successfully participated in proficiency testing provided by the GEP-ISFG Working Group.

\* Corresponding author.

E-mail address: [isabrunelli@hotmail.com](mailto:isabrunelli@hotmail.com) (I.B. Ambrosio).

<http://dx.doi.org/10.1016/j.fsigss.2017.09.135>

Received 2 September 2017; Accepted 19 September 2017

1875-1768/ © 2017 Elsevier B.V. All rights reserved.