

UNIVERSIDADE ESTADUAL PAULISTA “JÚLIO DE MESQUITA FILHO”
CAMPUS DE ARARAQUARA
INSTITUTO DE QUÍMICA

FERNANDA SILVA POLVERARI

**Caracterização genética da população do Estado do Mato Grosso
e do Distrito Federal (Brasília) pela análise de 32 polimorfismos
de inserção/deleção (InDels) no cromossomo X**

Araraquara
2018

FERNANDA SILVA POLVERARI

**Caracterização genética da população do Estado do Mato Grosso
e do Distrito Federal (Brasília) pela análise de 32 polimorfismos
de inserção/deleção (InDels) no cromossomo X**

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

Orientadora: Prof^a. Dr^a. Regina Maria Barretto
Cicarelli

Araraquara
2018

FICHA CATALOGRÁFICA

P762c Polverari, Fernanda Silva
Caracterização genética da população do Estado do Mato Grosso e do Distrito Federal (Brasília) pela análise de 32 polimorfismos de inserção/deleção (InDels) no cromossomo X / Fernanda Silva Polverari. – Araraquara : [s.n.], 2018
94 f. : il.

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

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

CERTIFICADO DE APROVAÇÃO

TÍTULO DA TESE: "Caracterização genética da população do Estado do Mato Grosso e do Distrito Federal (Brasília) pela análise de 32 polimorfismos de inserção/deleção (InDels) no cromossomo X"


AUTORA: FERNANDA SILVA POLVERARI

ORIENTADORA: REGINA MARIA BARRETTO CICARELLI

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


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


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


Prof. Dr. JOÃO ARISTEU DA ROSA
Departamento de Ciências Biológicas / Faculdade de Ciências Farmacêuticas - UNESP - Araraquara


Profa. Dra. LEONOR GUSMÃO
Departamento de Biologia / Universidade Estadual do Rio de Janeiro - UFRJ - Rio de Janeiro


Dr. RODRIGO RODENBUSCH
Departamento de Ações em Saúde / Secretaria Estadual de Saúde - Porto Alegre - RS

Araraquara, 26 de abril de 2018

DEDICATÓRIA

Ao meu filho Rafael, pelo amor mais puro que existe. Isto é por você meu amor!

A minha amada mãe Iara que sempre me incentivou a prosseguir, por maiores que fossem as dificuldades, fortalecendo-me a cada dia.

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

AGRADECIMENTOS

Agradeço a Deus, por estar presente em todos os dias da minha vida, iluminando e direcionando o meu caminho.

Ao meu filho Rafael, pelo amor e carinho incondicional, por compreender os momentos de ausência e aos queridos Sonia e Thomaz, família que a vida me deu de presente, obrigada pela dedicação, incentivo e por sempre acreditarem em tudo o que eu faço.

À professora Dra. Regina Cicarelli pelo incentivo e confiança de desenvolver este trabalho, que foram fundamentais para o meu crescimento científico, profissional e pessoal. O meu muito obrigado por ter me recebido em seu laboratório e compartilhado seus conhecimentos.

À professora Dra. Leonor Gusmão pela disponibilidade em transmitir o seu conhecimento, essencial para a finalização deste trabalho.

À Juliana Martinez, pelos ensinamentos e discussões, pela paciência e constante disponibilidade e principalmente pela amizade.

Ao perito Dr. Jorge Marcelo de Freitas (Polícia Federal) por ter disponibilizado as amostras de Brasília.

Às pesquisadoras Dra. Joyce Aparecida Martins e Me. Denise Poltronieri Martins pela disponibilização das amostras do Mato Grosso e por compartilhar conhecimento científico, mesmo a distância, elucidando diversas dúvidas.

À Isabela Brunelli, pela grande amizade, pelas conversas sérias e também descontraídas, pelo apoio em todos os momentos e por toda a ajuda e conhecimento compartilhado nestes últimos anos de convívio e muito trabalho.

À Bianca Belon pela amizade, pelos muitos momentos de alegria, pelas conversas, conselhos, puxões de orelha, por se preocupar comigo e me amparar nos momentos difíceis.

Ao Danilo Braganholi, pela amizade, pelo apoio técnico no desenvolvimento desse projeto e pelas valiosas informações fornecidas.

Aos amigos do LIP (Laboratório de Investigação de Paternidade): Flávia, Simone, Isadora e Natália, pela amizade, companhia no dia a dia, pelos momentos de conversa durante os cafés e pelas inúmeras risadas. É sempre bom trabalhar com vocês!

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

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

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

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

“Se você procurar bem, você acaba encontrando, não a explicação (duvidosa) da vida, mas a poesia (inexplicável) da vida”.

Carlos Drummond de Andrade

RESUMO

A análise dos polimorfismos do DNA é a melhor ferramenta encontrada para resolução de casos de identificação humana, sendo os marcadores STRs (*short tandem repeat*) localizados em regiões autossômicas os principais e mais utilizados para esta finalidade. Apesar da indiscutível reprodutibilidade destes marcadores, quando são analisados em amostras degradadas podem não apresentar bons resultados, o que dificulta a resolução dos casos. Assim, há a necessidade de uma alternativa e, atualmente, a mais indicada é a utilização dos polimorfismos bialélicos. A análise do cromossomo sexual X também vem ganhando importância significativa no contexto forense, devido ao seu padrão de transmissão entre os genitores. Neste trabalho, caracterizamos as populações brasileiras do estado do Mato Grosso e do Distrito Federal pela análise de 32 polimorfismos de inserção/deleção no cromossomo X (X-InDels), tendo em vista que os dados destes polimorfismos nestas populações são ainda inéditos, e para que esses marcadores também possam no futuro auxiliar a resolução de casos forenses em nosso país. Para identificar a diversidade genética foram analisados os perfis genotípicos de 303 indivíduos não aparentados nascidos no estado do Mato Grosso e 179 indivíduos não aparentados e residentes em Brasília. Os resultados indicam que o painel dos 32 X-InDels é bastante eficiente para a sua finalidade, sendo que praticamente todos os marcadores se mostraram altamente informativos para as populações estudadas. O teste de equilíbrio de Hardy-Weinberg foi realizado nas amostras femininas de ambas as populações e não foram verificados desvios significativos. O painel demonstrou alta eficiência forense, confirmado pelo alto poder de discriminação para Mato Grosso ($PD_F=0,999999999997$ e $PD_M=0,99999997$) e para Brasília ($PD_F=0,999999999998$ e $PD_M=0,99999998$), e pelo elevado poder de exclusão observado em trios: suposto pai/mãe/filha (0,999998/Mato Grosso e 0,999997/Brasília) e duos: suposto pai/filha (0,9996/Mato Grosso e 0,9995/Brasília). No estudo comparativo com outras populações, Mato Grosso e Brasília estão mais próximos ao estado de São Paulo, à três departamentos colombianos e às populações europeias. A proporção de ancestralidade confirmou a miscigenação das populações estudadas, sendo identificado uma contribuição praticamente equilibrada para europeus, africanos e nativo-americanos. 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 complementar para a análise de amostras forenses, pois foram identificados elevados valores de poder de discriminação e exclusão.

Palavras-chave: Identificação humana. Genética populacional. Polimorfismos de inserção/deleção. InDels. Cromossomo X. X-InDels. Mato Grosso. Brasília. Brasil.

ABSTRACT

The analysis of DNA polymorphisms is the best tool found for solving cases of human identification, and the Short Tandem Repeat (STR) markers used in the autosomal regions are the main and most marker used for this purpose. Despite the undeniable reproducibility of these markers, when analyzed in degraded samples they may not present good results, which makes it difficult to solve the cases. Because of this, there is a need for an alternative tool and currently the most indicated is the use of the biallelic polymorphisms. In this way, the analysis of the sex chromosome X has gained significant importance in the forensic context, due to its pattern of transmission between the parents. In this work, we characterize the Brazilian populations of the state of Mato Grosso and the Federal District by the analysis of 32 insertion/deletion polymorphisms on the X chromosome (X-InDels), considering that the data of this polymorphism in these populations are still unpublished, and for that these markers may also in the future help the resolution of forensic cases in our country. To identify the genetic diversity, the genotypic profiles of 303 unrelated individuals born in the state of Mato Grosso and 179 unrelated individuals living in Brasília were analyzed. The results shows that the 32 X-InDels panel is very efficient to its purpose, being almost all markers highly informative for the studied populations. The Hardy-Weinberg equilibrium test was performed on the female samples from both populations and no significant deviations were observed. The panel demonstrated high forensic efficiency, confirmed by the high discriminatory power for Mato Grosso (PDF= 0.9999999999997 and PDM=0.999999997) and Brasília (PDF=0.9999999999998 and PDM=0.999999998), and by the high exclusion power observed in trios: supposed father/mother/daughter (0.999998/Mato Grosso and 0,999997/Brasília) and duos: supposed father/daughter (0.9996/Mato Grosso and 0,9995/Brasília). In the comparative study with other populations, Mato Grosso and Brasília are closer to the state of São Paulo, to three departments in Colombia and to European populations. The proportion of ancestry confirmed the admixture of the populations studied, with a practically balanced contribution being identified for Europeans, Africans and American natives. We could concluded that knowledge about insertion/deletion markers on the X chromosome can be amplified, since in the literature there is still fewer available material on the subject, however the data of this work already demonstrates its potential as a complementary method for the analysis of forensic samples, since were identified high values of power of discrimination and exclusion.

Keywords: Human identification. Population genetics. Insertion/deletion polymorphism. InDels. X chromosome. X-InDels. Mato Grosso. Brasília. Brazil.

LISTA DE FIGURAS

Figura 1.	Esquema ilustrando o tamanho médio de fragmentos amplificados na PCR e o processo de amplificação para marcadores STRs, SNPs e InDels.....	21
Figura 2.	Representação esquemática dos cromossomos sexuais X e Y.....	25
Figura 3.	Distribuição de 30 marcadores X-STR de uso forense.....	29
Figura 4.	Ideograma do cromossomo X indicando a localização dos 32 marcadores X-InDel selecionados.....	32
Figura 5.	Eletroferograma com perfil alélico de amostra masculina obtida com o multiplex 32 X-InDels.....	48
Figura 6.	Eletroferograma com perfil alélico de amostra feminina obtida com o multiplex 32 X-InDels.....	49
Figura 7.	Gráfico MDS bidimensional elaborado a partir da distância genética (FST) entre África, Europa, Sudeste Asiático, Nativo-americano, diversas populações da Colômbia e Brasil (São Paulo, Mato Grosso e Brasília).....	63
Figura 8.	A - Representação esquemática da mistura estimada para a população de Mato Grosso e Brasília (DF) utilizando 32 marcadores X-InDels obtido através da análise do STRUCTURE baseado na contribuição da população africana (AFR), europeia (EUR) e nativo-americana (NAM) (K=3); B - Dispersão dos indivíduos analisados entre os três grupos populacionais de referência.....	66

LISTA DE TABELAS

Tabela 1.	Marcadores X-InDels analisados: localização de cada marcador no genoma, alelos possíveis e sequência dos primers.....	41
Tabela 2.	Frequências alélicas e diversidade gênica dos 32 marcadores X-InDel nas populações do Mato Grosso e de Brasília. Número de cromossomos analisados: 406 (Mato Grosso) e 346 (Brasília).....	50
Tabela 3.	Parâmetros estatísticos da diversidade genética para os 32 marcadores X-InDels nas amostras femininas das populações estudadas. Tamanho da amostra: 103 (Mato Grosso) e 167 (Brasília).....	52
Tabela 4.	Ideograma do cromossomo X indicando a localização dos 32 marcadores X-InDels selecionados.....	53
Tabela 5.	Desequilíbrio de ligação dos pares de marcadores analisados na população masculina de Brasília que obtiveram valor de $p \leq 0,05$. Tamanho da amostra: 12.....	54
Tabela 6.	Frequência haplotípica observada na população masculina do Mato Grosso para os marcadores em desequilíbrio de ligação. Tamanho da amostra: 200.....	57
Tabela 7.	Parâmetros estatísticos de interesse forense para os marcadores 32 X-Indels na população do Mato Grosso. Tamanho da amostra: 406.....	58
Tabela 8.	Parâmetros estatísticos de interesse forense para os marcadores 32 X-Indels na população de Brasília. Tamanho da amostra: 346.....	59
Tabela 9.	Parâmetros estatísticos indicativos de interesse forense obtido para o conjunto dos 32 X-Indels.....	61
Tabela 10.	Amostras usadas na comparação entre as populações, número de cromossomos analisados (N) e respectivas referências.....	62
Tabela 11.	Proporção de ancestralidade (%) para as amostras testadas de populações africanas, europeias e nativo-americanas (PEREIRA et al., 2012) e as do Mato Grosso e de Brasília.....	67
Tabela 12.	Distâncias genéticas (FST) entre populações (abaixo da diagonal) e valor de p correspondente (acima da diagonal).....	69

LISTA DE ABREVIATURAS E SIGLAS

DNA	Ácido desoxirribonucleico
dbSNP	<i>Single Nucleotide Polymorphism database</i>
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 _T	Chance média de exclusão em trios envolvendo filha
MEC _D	Chance média de exclusão em duo pai/filha
MEC _{PGM}	Chance média de exclusão em casos 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 pseudoautosômica
PD	Poder de discriminação
PD _F	Poder de discriminação em mulheres
PD _M	Poder de discriminação em homens
SNP	<i>Single Nucleotide Polymorphism</i>
STR	<i>Short Tandem Repeat</i>
X-InDel	Inserção/deleção no cromossomo X
X-STR	<i>Short Tandem Repeat</i> no cromossomo X

SUMÁRIO

1.	Introdução.....	13
2.	Revisão da Literatura.....	15
	2.1 - Marcadores biológicos para identificação humana.....	16
	2.2 - Polimorfismos genéticos e análise do DNA.....	17
	2.2.1 - <i>Short Tandem Repeats</i> (STRs).....	18
	2.2.2 - <i>Single Nucleotide Polymorphisms</i> (SNPs).....	19
	2.2.3 - <i>Insertion/Deletion Polymorphisms</i> (InDels).....	20
	2.3 - Cromossomo X.....	23
	2.3.1 - Aplicações dos marcadores genéticos do cromossomo X.....	26
	2.3.2 - Marcadores genéticos do cromossomo X.....	28
	2.4 - Aspectos Históricos da Formação da Região Central do Brasil, do Estado do Mato Grosso e de Brasília (DF).....	33
3.	Objetivos.....	36
4.	Materiais e Métodos.....	38
	4.1 - Aspectos éticos.....	39
	4.2 - Casuística.....	39
	4.3 - Coleta das amostras biológicas e extração do DNA.....	39
	4.4 - Reação de amplificação dos X-InDels.....	40
	4.4.1 - Preparação para eletroforese capilar.....	44
	4.5 - Determinação dos perfis alélicos e análise estatística.....	44
5.	Resultados e Discussão.....	46
	5.1 - Avaliação da metodologia.....	47
	5.2 - Variabilidade genética.....	50
	5.3 - Desequilíbrio de ligação entre os marcadores X-InDels.....	53
	5.4 - Parâmetros estatísticos de eficiência forense.....	58
	5.5 - Comparação genética entre as populações e análise de ancestralidade....	61
6.	Conclusão.....	70
7.	Referências.....	72
	Anexo I - Aprovação do Comitê de Ética em Pesquisa com Seres Humanos.....	78
	Anexo II - Trabalhos publicados.....	82

1. INTRODUÇÃO

A introdução das técnicas de biologia molecular na análise para identificação humana (*Human IDentification*) é um avanço relativamente recente e a determinação de identidade genética pela análise do DNA constitui um dos produtos mais revolucionários da Genética Moderna. Em menos de 20 anos, tornou-se uma ferramenta indispensável na área forense, sendo aceita rotineiramente em processos judiciais em todo o mundo (FRANÇA, 2001). Isto se deve ao alto poder de discriminação do DNA, o qual permite obter a identificação de um indivíduo e/ou determinar o vínculo biológico entre indivíduos com uma menor probabilidade de erro.

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

Os marcadores genéticos mais analisados na rotina forense são principalmente os marcadores nucleares autossômicos, pois apresentam alto grau de diversidade e alta resolubilidade dos casos (ALVAREZ-CUBERO et al., 2012).

Entretanto, há situações em que o material biológico para a análise pode estar degradado e/ou em baixa quantidade, ou ainda em casos deficientes de vínculo biológico, como exames de DNA realizados na ausência do suposto pai, nos quais análise apenas dos marcadores autossômicos pode não ser suficiente para solucionar o caso. Nestas situações, a análise de marcadores em outras regiões polimórficas, como no cromossomo X, podem complementar a análise tradicional de forma eficaz (ALVAREZ-CUBERO et al, 2012).

2. REVISÃO DA LITERATURA

2.1 Marcadores biológicos para identificação humana

A identificação humana é o processo científico pelo qual se determina a identidade de uma pessoa, por meio do estabelecimento de um conjunto de caracteres que a individualize, fazendo-a igual apenas a si mesma (com exceção em gêmeos monozigóticos) (FRANÇA, 2001). Dentre os processos de identificação humana, tem-se o campo da ciência que envolve a análise de marcadores biológicos proteicos e genéticos.

Os estudos modernos sobre a variação genética humana tiveram seu início por volta do princípio do século XX, com a descoberta dos grupos sanguíneos ABO (JORDE; WATKINS; BAMSHAD, 2001), sendo tal sistema aplicado por várias décadas nos exames forenses de amostras biológicas em evidências relacionadas a crimes ou à identificação de pessoas. Posteriormente, outros sistemas de marcadores proteicos como os componentes grupo-específicos (transferrina, albumina, ceruloplasmina, haptoglobina, fosfoglicomutase-1, fosfatase ácida, esterase D, etc) mostraram-se variáveis entre os grupos populacionais e passaram, também, a ser utilizados.

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

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

A tipagem molecular de material genético foi utilizada oficialmente pela primeira vez, em 1985, por Jeffreys, na Inglaterra para a resolução de um problema de imigração (JEFFREYS et al, 1985b). Um ano após, o mesmo autor empregou

esta técnica para identificar o verdadeiro estuprador e assassino de duas vítimas. A partir deste caso, que ficou conhecido com *Enderby (Queen x Pitchfork)*, a análise do DNA passou a ser potencialmente utilizada na rotina forense (TANDE, 1989).

2.2 Polimorfismos genéticos e análise do DNA

A análise de DNA tem como objetivo diferenciar um indivíduo de outro, por meio de um grande número de características genéticas, dando-lhe uma identidade absoluta como pessoa, podendo assim ser diferenciado dentre bilhões de outras pessoas (BONACCORSO, 2004).

A maioria das informações genéticas é idêntica para os diferentes indivíduos da espécie humana – é isso que explica o fato de todas as pessoas terem as mesmas características gerais: a quantidade de dedos, a forma geral do corpo e todas as demais peculiaridades que nos distinguem de outras espécies. Entretanto, para um grande número de características genéticas, as informações são semelhantes, mas não idênticas. As informações genéticas que variam na população chamam-se polimorfismos. É por meio da análise dessas informações polimórficas que se torna possível distinguir os indivíduos entre si e estabelecer uma relação de vínculo biológico em testes de parentesco (JOBILING et al, 2004).

A palavra “polimorfismo” tem origem grega e significa “muitas formas”, assim, regiões no genoma que apresentam variabilidade superior a 1% são denominados polimorfismos genéticos. Essas variações são constituídas por polimorfismos de sequência ou de comprimento, a depender das alterações ocorridas na sequência do DNA (BEIGUELMAN, 1996).

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

Os polimorfismos de comprimento são originados por inserção ou deleção de um ou mais nucleotídeos 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).

2.2.1 Short Tandem Repeats (STRs)

Os marcadores STR, também conhecidos como microssatélites, são unidades de repetição do DNA constituídas de 2 a 6 pb que se repetem em tandem por inúmeras vezes (BUTLER, 2007). As sequências de repetição dos STRs são nomeadas pelo comprimento da unidade de repetição, logo existem, mono-, di-, tri-, tetra-, penta- e hexanucleotídeos. Devido às características inerentes aos microssatélites com bloco de repetição menor que quatro nucleotídeos, a Sociedade Internacional de Genética Forense (ISFG) recomenda a utilização de tetranucleotídeos para identificação humana, pois estes possuem um grau de polimorfismo razoável, baixa taxa de mutação e a formação reduzida de stutter (comparada aos dinucleotídeos) (MONTEIRO, 2007).

Os STRs tornaram-se marcadores de DNA muito utilizados na identificação humana porque são facilmente amplificados pela reação em cadeia da polimerase (PCR), o número de repetições pode ser altamente variável entre os indivíduos (BUTLER, 2011) e, permitem a utilização de reações em *multiplex*, que é a amplificação simultânea de múltiplas regiões do DNA, adicionando mais de um par de iniciadores na reação de PCR.

Além disso, a detecção pode ser realizada em sistemas fluorescentes que permitem a automação dos sistemas de eletroforese e a interpretação dos perfis de DNA (BUTLER, 2005).

Os STRs são considerados marcadores genéticos essenciais para a aplicação forense. Na maioria dos casos, a análise de 20 STRs autossômicos é suficiente para a identificação humana ou para testes de paternidade (YUAN et al, 2014). Entretanto, em situações em que a amostra biológica se encontra degradada ou em casos complexos de reconstrução para a identificação de paternidade, o resultado final estatístico pode ser inconclusivo (PINTO et al, 2013).

Nos casos em que a análise de STRs não é suficiente para a conclusão, a prática comum consiste em ampliar o número dos marcadores analisados. Nesse

contexto, há autores que sugerem a escolha de marcadores bialélicos, tais como os polimorfismos de base única (SNPs) ou de inserção/deleção (InDels), uma vez que esses são muito menos susceptíveis a mutações quando comparados aos STRs e geram fragmentos de tamanho reduzido na amplificação por PCR, sendo extremamente importante na amplificação em amostra de DNA degradado (PINTO et al, 2013).

2.2.2 *Single Nucleotide Polymorphisms (SNPs)*

Os SNPs consistem na substituição de uma única base na sequência do DNA, e é a forma mais simples e mais comum de variação genética (DAWSON et al., 2001). Um SNP caracteriza-se, por exemplo, pela substituição de uma base “A” (adenina) por “T” (timina) num determinado ponto no genoma.

Normalmente apresentam-se como marcadores bialélicos (referência ou alternativo) e podem ocorrer em regiões codificadoras ou com função regulatória, porém, na maior parte das vezes são encontrados em espaços intergênicos, sem função determinada. Apesar da baixa taxa de mutação, os SNPs são distribuídos abundantemente no genoma ocorrendo em média entre 1-2 Kb (CLIFFORD et al., 2000).

Uma vez que os SNPs são mais estáveis que os STRs, eles são úteis para o estudo da evolução história das populações humanas em escalas de tempo mais profundas (JAKOBSSON et al., 2008). Na área forense, apesar de apresentar uma diversidade genética menor que os STRs, os SNPs têm várias características desejáveis como marcadores genéticos: podem ser analisados em fragmentos muito curtos, o que é importante para melhorar o sucesso de amplificação em DNA altamente degradado e apresentam baixa taxa de mutação, o que é interessante em testes de parentesco (NACHMAN; CROWELL, 2000).

Os SNPs são utilizados na identificação de haplótipos do DNA mitocondrial (DNA mt) e cromossomo Y e, mais recentemente, na análise de amostras altamente degradadas como na identificação de vítimas de desastres, por exemplo, bem como para inferir ascendência biogeográfica de indivíduos e desvendar características fenotípicas (JOBLLING; GILL, 2004; BUDOWLE; van DAAL, 2008; KAYSER; de KNIJFF, 2011).

2.2.3 Insertion/Deletion Polymorphisms (InDels)

Os polimorfismos de inserção/deleção (InDels) são polimorfismos de comprimento, caracterizados pela inserção ou deleção de um ou mais nucleotídeos em uma determinada região do genoma. Weber e colaboradores (2002) foram pioneiros na identificação e caracterização de InDels e, devido à abundância no genoma humano e a facilidade de análise, destacaram a sua utilidade para estudos genéticos.

Além disso, este trabalho representou o início de um banco de dados online de polimorfismos de inserção/deleção criado pela *Marshfield Clinic* (<http://www.marshfieldclinic.org/mgs>), incluindo várias informações sobre os polimorfismos: alelos identificados, diferenças de comprimento, *primers* para amplificação, posição no genoma, bem como as frequências alélicas nos principais grupos populacionais do mundo (africanos, europeus, asiáticos e nativo-americanos) (PEREIRA; GUSMÃO, 2012). Informações sobre os InDels também foram incluídos no *Single Nucleotide Polymorphism databe (dbSNP)* (<https://www.ncbi.nlm.nih.gov/snp>).

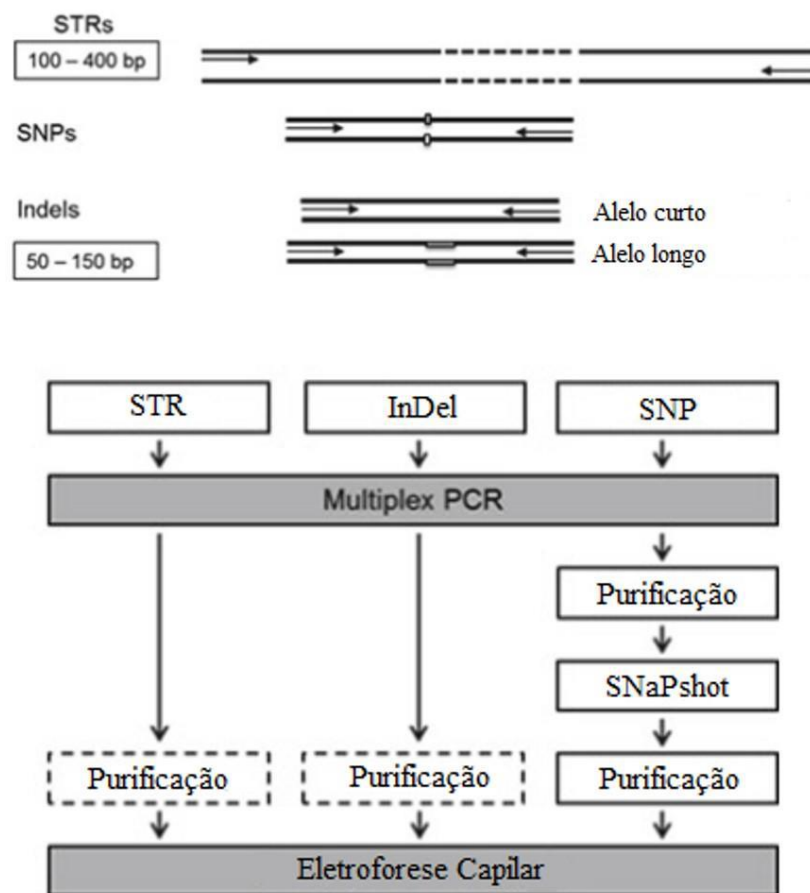
Em 2006, foram identificados novos InDels e relatado um mapa inicial contendo mais de 415.000 polimorfismos únicos no genoma humano (MILLS et al., 2006).

Estes polimorfismos possuem características interessantes para uso na identificação humana: baixa taxa de mutação em comparação com os STRs; são amplamente distribuídos no genoma, sendo encontrados em média a cada 7,2 kpb (MILLS, 2006); representam 20% de todos os polimorfismos do genoma humano (WEBER, 2002); podem ser detectados por diferentes técnicas (MULLANEY, 2010); o tamanho do fragmento a ser amplificado na PCR é pequeno (50 a 150 pb) o que facilita a análise em amostras degradadas (ZIDKOVA et al., 2011); as frequências alélicas podem apresentar diferenças significativas entre grupos de populações geograficamente diferentes, sendo potenciais marcadores para estudo de ancestralidade (YANG et al., 2005). Além disso, as sequências podem ser analisadas em multiplex e as reações são tão simples quanto aquelas dos STRs (TORRES et al, 2014).

Para a análise dos cromossomos autossômicos existe atualmente no mercado um kit comercial que analisa os polimorfismos de inserção/deleção,

chamado DIPplex Kit[®] (Qiagen), sendo que sua eficácia como ferramenta complementar em identificação humana vem sendo comprovada em diversas partes do mundo (PEPINSKI et al, 2013; SAIZ et al, 2014; SEONG et al, 2014; WANG et al, 2014), inclusive no Brasil (TORRES et al, 2014).

Figura 1: Esquema ilustrando o tamanho médio de fragmentos amplificados na PCR e o processo de amplificação para marcadores STRs, SNPs e InDels.



Fonte: adaptado de Pereira e Gusmão, 2012.

Assim, estudos que buscam a padronização de novos marcadores vêm sendo elaborados, a fim de aumentar a possibilidade de utilização de InDels na prática forense. Nessa perspectiva, Pereira e colaboradores (2009) descreveram um novo multiplex para a identificação humana na qual foi possível analisar 38 InDels autossômicos em uma única reação. Os amplicons gerados na PCR dessa reação

possuem no máximo 160 pb e podem ser analisados diretamente em analisadores 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) avaliaram sua aplicabilidade em casos forenses no Brasil. A estatística mostrou diferenças entre as frequências alélicas do Rio de Janeiro e de Europeus, Africanos e Asiáticos, apontadas por Pereira e colaboradores (2009). Diante dos casos forenses, os autores analisaram restos humanos em testes de paternidade e identificaram uma eficiente amplificação de DNA degradado, o que contribuiu com a resolução de processos em que apenas o uso de STRs foi insuficiente.

Complementando a análise forense, além do DNA nuclear dos cromossomos autossômicos, também é possível a análise de marcadores nos cromossomos sexuais (X e Y) e no DNA mitocondrial, sendo que a seleção de um ou outro irá depender das particularidades do caso.

Marcadores do cromossomo Y são ditos de linhagem paterna, pois são transferidos de pai para filho; enquanto o DNA mitocondrial é de linhagem materna, sendo passado da genitora para seus filhos. No caso do cromossomo X, a mulher herda um cromossomo de cada um dos seus progenitores e o homem herda apenas o cromossomo materno, sendo seu único cromossomo X transmitido à todas as filhas isento de recombinação. Por outro lado, nos marcadores autossômicos, homens e mulheres herdam dois cromossomos: um de herança materna, e outro, de herança paterna. Devido as suas características, os marcadores STRs autossômicos são amplamente utilizados na rotina dos laboratórios forenses.

No entanto, outros marcadores podem auxiliar a análise dos tradicionais STRs autossômicos. Como mencionado anteriormente, os marcadores InDels vêm atraindo cada vez mais a atenção dos cientistas da área de identificação humana como uma nova ferramenta a ser utilizada para acrescentar robustez na análise de amostras degradadas ou em casos com resultado inconclusivo. Em complementação, os marcadores InDels no cromossomo X, constitui-se em uma importante ferramenta em casos específicos de investigação de relação biológica, sendo de grande utilidade na prática forense.

2.3 Cromossomo X

O cromossomo X pode ser usado como ferramenta para gerar informações tanto sobre a herança materna quanto sobre a herança paterna de uma determinada população (Santos-Lopes et al., 2007). Nas mulheres, metade da informação genética existente no cromossomo X é proveniente da linhagem materna e a outra metade da linhagem paterna. Nos homens o cromossomo X (linhagem materna) está presente em apenas uma cópia, o que favorece a identificação de haplótipos.

Este cromossomo é formado por aproximadamente 155 Mb e representa cerca de 5% do genoma das células femininas e 2,5% do das células masculinas. Possui uma baixa densidade gênica, foram identificados 1.098 genes, representando 4% do total de genes humanos. A quantidade de regiões repetitivas é elevada, correspondendo a 56% do DNA deste cromossomo, sendo que a média do genoma é de 45%. Dentro dessas regiões repetitivas foram identificados os elementos intercalares curtos (SINEs – *Short Interspersed Nuclear Elements*) em quantidades abaixo da média dos cromossomos e os elementos intercalares longos (LINEs – *Long Interspersed Nuclear Elements*) em quantidades acima da média (MARTINS, 2011). Além de terem sido registrados 153.146 polimorfismos de nucleotídeos únicos (SNPs) (ROSS et al. 2005).

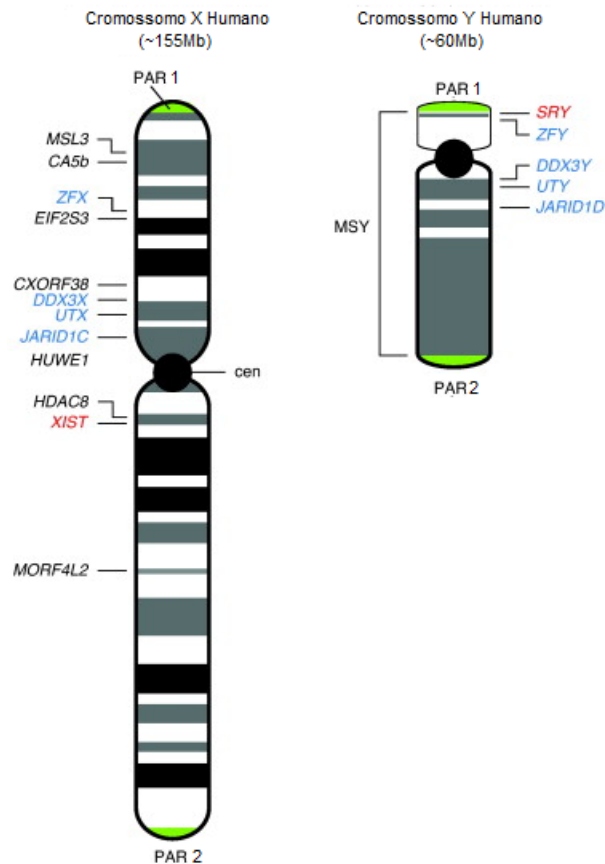
Apesar de apenas 4% dos genes do genoma humano estarem presentes no cromossomo X, estes são de grande interesse na genética médica, uma vez que cerca de 10% das doenças com herança mendeliana têm sido atribuídas a este cromossomo. Isto ocorre pelo fato do homem herdar apenas um cromossomo X, revelando de forma direta o fenótipo decorrente de uma mutação recessiva. Muitas desordens genéticas, incluindo às ligadas ao cromossomo X, podem ser acessadas pela base de dados OMIM (*Online Mendelian Inheritance in Man*) (HAMOSH, 2005).

Indivíduos do sexo feminino têm o dobro de cópias dos genes ligados ao X em relação ao sexo masculino. Com a finalidade de tornar proporcional os produtos gênicos do cromossomo X entre homens e mulheres, um dos cromossomos X da mulher é inativado no início do desenvolvimento embrionário (CHOW et al., 2005) e sofre uma reativação nas células germinativas para recombinar com o outro cromossomo X feminino. No homem, o cromossomo X não se recombina ao longo de todo o seu comprimento, sendo tal processo restrito às regiões

pseudoautossômicas (PAR, *pseudoautosomal region*) 1 e 2, localizadas nas extremidades dos braços dos cromossomos sexuais.

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 específica do sexo masculino, corresponde a 95% do cromossomo e recombina 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 pseudoautossômicas (PAR); MSY (*Male-Specific region of the Y*).

Fonte: adaptado de Wijchers; Festenstein, 2011.

Os marcadores STR do cromossomo X, utilizados na identificação humana, estão localizados em regiões não recombinantes entre os cromossomos X e Y. Entretanto, ocorre recombinação entre os cromossomos X na mulher, garantido a variabilidade desses marcadores. Já nos homens isso é inviável devido a hemizigose do cromossomo X nestas regiões (SZIBOR et al., 2007).

Assim, da mesma forma que os autossomos, o cromossomo X se recombina (mulheres) e, similarmente ao cromossomo Y e DNA mitocondrial, tem um modelo que permite determinar diretamente o haplótipo (homens) (SCHAFFNER, 2004).

O cromossomo X apresenta características genéticas distintas em relação aos autossomos, principalmente: (1) menor tamanho populacional efetivo (SCHAFFNER, 2004); (2) taxa de mutação menor. Segundo Schaffner (2004), todos os cromossomos masculinos possuem maior taxa de mutação que os femininos devido ao elevado número de mitoses nas células masculinas (exceto nas células germinativas), como dois cromossomos X estão presentes nas mulheres e apenas um nos homens a taxa de mutação global deste cromossomo é inferior a dos autossomos; e (3) maior desequilíbrio de ligação, uma vez que o cromossomo X recombina apenas na meiose feminina (PEREIRA et al., 2007; ZARRABEITIA et al., 2009).

Devido ao modo de transmissão 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.

2.3.1 Aplicações dos marcadores genéticos do cromossomo X

As propriedades de transmissão do cromossomo X o torna altamente interessante para estudos de identificação individual e de diversidade genética nas populações humanas. Dentro dos estudos da genética forense, o cromossomo X vem ganhando interesse nos últimos anos especialmente na análise das seguintes situações:

a) Análise de relações biológicas

Testes de investigação de paternidade que envolve trios comuns (suposto pai, mãe e filha) podem facilmente ser solucionados apenas com marcadores autossômicos e não necessitam de marcadores adicionais (SZIBOR et al., 2003). Se o teste de paternidade envolve uma dupla suposto pai e filho, os X-STRs também não são utilizados, já que não há herança paterna do cromossomo X.

Entretanto, se o teste de paternidade envolve um duo (suposto pai e filha), o uso de marcadores no cromossomo X é capaz de complementar a análise de marcadores autossômicos de forma eficiente em situações onde são observadas

uma ou duas incompatibilidades, pois é necessária a confirmação de exclusão ou ocorrência de mutação. Como o pai transfere 100% de seu material genético do cromossomo X para a filha, os marcadores nesse cromossomo apresentam elevada chance média de exclusão (MEC, do inglês *Mean Exclusion Chance*), o que significa que possuem alta capacidade de exclusão nas investigações de parentesco (SZIBOR et al., 2003; SZIBOR, 2007).

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

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

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

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

Em testes de investigação de paternidade ou identificação humana com dificuldade para se analisar o material biológico, tais como em exumação, embora o material a ser genotipado é do suposto pai, o número de STRs amplificados com sucesso pode ser menor do que geralmente se obtém com amostras de sangue ou saliva e o laudo pode permanecer inconclusivo. Como os X-STRs possuem maiores valores de MEC que os autossômicos, é possível se obter um suficiente poder estatístico para a resolução do caso quando um pequeno número destes marcadores são analisados (SZIBOR et al., 2003).

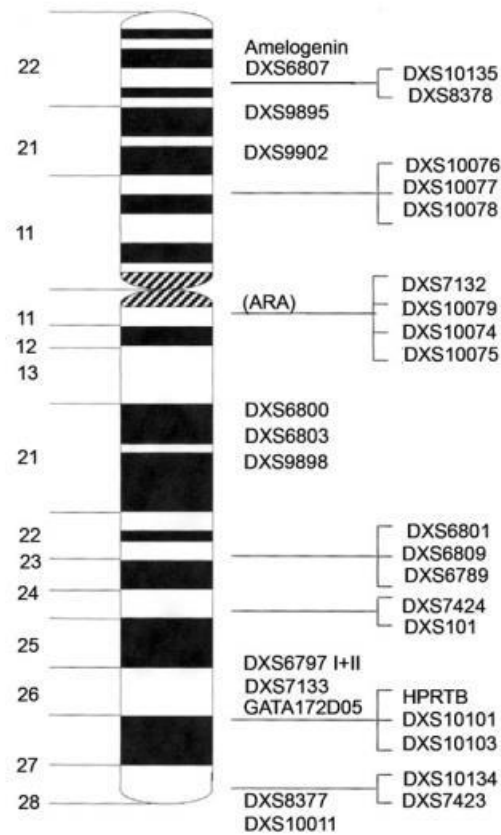
Por outro lado, nas situações em que se pretende identificar um perfil feminino em meio a uma mistura contendo material genético masculino, os marcadores de X tendem a ser mais eficientes, pois os alelos femininos somente poderão coincidir completamente com os masculinos se a mulher for homozigótica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

2.3.2 Marcadores genéticos do cromossomo X

Na identificação humana e na rotina forense, a análise de marcadores STRs destacam-se como uma ferramenta eficiente nas investigações de vínculo genético. 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).

Neste contexto, cada vez mais cresce a investigação em tal cromossomo e diversos marcadores têm sido estudados por diferentes grupos. Mais de 50 X-STRs foram identificados e Szibor e colaboradores (2007) em sua revisão descreveram os 30 marcadores de uso forense mais conhecidos até aquela época (Figura 3).

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



Fonte: Szibor, et al., 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 microsatélites. 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).

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 restrita 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 Diagnostic GmbH 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 (Biotype; Qiagen), que amplifica simultaneamente 12 X-STRs mais amelogenina (<http://www.chrx-str.org/>, acesso em 15 de março de 2018).

Em 2006, o Grupo de Línguas Espanhola e Portuguesa da Sociedade Internacional de Genética Forense (GHEP-ISFG) elaborou o primeiro trabalho colaborativo com X-STRs, no qual desenvolveu um sistema *multiplex in house* com a análise de 10 X-STRs (*decaplex*), nos quais foram caracterizados em 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. Recentemente foi desenvolvido um sistema *multiplex* com a análise de 17 X-STRs, incluindo os marcadores do sistema *decaplex*, demonstrando uma confiabilidade desta ferramenta para fins de genotipagem forense (Prieto-Fernández, et al, 2016).

Para a análise de marcadores X-STR estudos vêm sendo produzidos em diversas partes do mundo (DONG et al, 2014; ISRAR et al, 2014; MARTINS et al, 2010; NOTHNAGEL et al, 2012). Um banco de marcadores X-STR no Brasil foi estabelecido pela equipe da orientadora Profa. Dra. Regina Maria Barretto Cicarelli (MARTINS et al, 2014).

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

Buscando a padronização de marcadores X-InDel para uso forense, um grupo brasileiro (FREITAS et al, 2010) desenvolveu uma metodologia utilizando 33 X-InDels em uma única reação de PCR em *multiplex*. Os autores demonstraram a

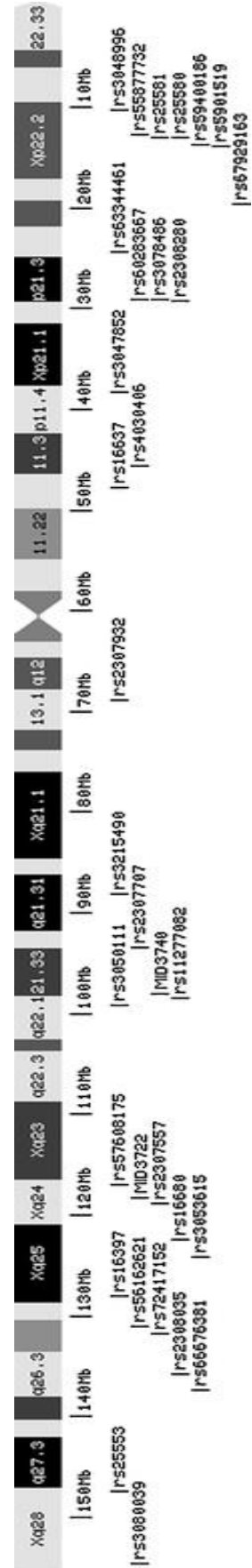
possibilidade de aplicação da metodologia em casos de identificação humana e testes de parentesco em uma população miscigenada da Amazônia e concluíram que se tratam de marcadores robustos e eficientes, que podem facilmente ser adotados na rotina forense.

A eficiência do painel 33 X-InDel foi recentemente analisada na população argentina (CAPUTO et al., 2017) e os autores identificaram que 5 marcadores e 2 blocos haplotípicos apresentaram heterozigosidade 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.

Partindo deste mesmo contexto, Pereira e colaboradores (2012) desenvolveram uma PCR em multiplex com 32 marcadores X-InDels 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 aquelas em que o DNA encontra-se degradado. A distribuição dos 32 marcadores ao longo do cromossomo X está disposta na figura 4.

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

Fonte: Pereira et al., 2012.



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.

2.4 Aspectos Históricos da Formação da Região Central do Brasil, do Estado do Mato Grosso e de Brasília (DF)

A colonização do centro-oeste brasileiro sofreu influência cultural e racial dos grupos que contribuíram com a formação da região. Além dos povos indígenas que ocupavam a região desde o período pré-colonial, há também influências espanholas e, principalmente, portuguesas devido ao processo de colonização.

Um tratado de limites, intitulado de Tordesilhas, firmado dois anos depois do descobrimento da América em 1494, dividiu o continente em duas partes, cabendo a Portugal as terras situadas ao leste e a Espanha as situadas ao oeste, o que tornava a região que hoje se localiza o Mato Grosso pertencente à população espanhola (SIQUEIRA, 2009).

As primeiras incursões no território do Mato Grosso datam de 1525, quando o espanhol Pedro Aleixo Garcia, seguindo as águas dos rios Paraná e Paraguai, direcionou-se ao sertão do continente. Muitas outras excursões seguiram o mesmo caminho, entretanto, devido às riquezas presentes nas minas do Peru, os espanhóis passavam pela região do Mato Grosso, mas não fixavam-se nela (FERREIRA, 2009; SILVA, 2011; SIQUEIRA, 2009).

Em busca de mão-de-obra para trabalhar no plantio da cana-de-açúcar, no século XVII, os bandeirantes paulistas direcionaram-se ao sertão, sem respeitar o Tratado de Tordesilhas, devido aos rumores de que este era habitado por índios. Ao encontrar os índios às margens do rio Coxipó-Mirim, os bandeirantes também descobriram enormes jazidas de ouro, dando início à corrida do ouro, fato que ajudou a povoar a região. A partir de então, outros auríferos foram descobertos e uma Ata de Fundação do descobrimento das minas foi elaborada para a região, que

passou a ser posse da população portuguesa e, posteriormente, se tornou a capital de Mato Grosso, Cuiabá (FERREIRA, 2009; SILVA, 2011; SIQUEIRA, 2009).

Os motivos pelos quais ocorreram as expedições para oeste do Brasil são diversos. A coroa portuguesa precisava ocupar as terras a oeste para se defender da ocupação espanhola de oeste para leste e preservar o Tratado de Tordesilhas. As expedições feitas pelos paulistas foram de caráter principal econômico como a procura por indígenas que era uma mão-de-obra mais barata que a escrava, ocorridas em 1718 e 1719, a mineração em 1719 com o propósito de exploração de ouro e pedras preciosas (MENDES, 2012).

Passado o período da mineração, a partir da década de 30, a política de colonização do Mato Grosso entra em uma nova fase econômica com a criação do programa “Marcha para o Oeste”. A proposta do governo incluía a construção de estradas, aeroportos e núcleos de colonização, direcionando o estado para a produção agrícola. Somente partir dos anos 60, a agricultura mato-grossense começa a despontar no cenário brasileiro, com a entrada de gaúchos e paranaenses que investiram na cultura do trigo e da soja (CUNHA, 2006).

A partir das décadas de 70 e 80, ainda impulsionados pelo auxílio estatal com programas de incentivo à modernização agropecuária e interação da região aos outros mercados (por exemplo, Prodoeste - Programa de Desenvolvimento do Centro-Oeste), há uma grande expansão do agronegócio no estado, marcado pela instalação de importantes empresas agro-industriais nacionais e internacionais (CUNHA, 2000; CUNHA, 2006).

Diante desse novo contexto da “marcha modernizadora do oeste”, a mão-de-obra presente não era suficiente para atender às necessidades distritais, o que provocou intenso direcionamento dos fluxos migratórios para essas áreas que pareciam promissoras. Em consequência, a população do estado do Mato Grosso, que na década de 70 era 579.360 habitantes, passa para 1.039.350 na década de 80 e dobra novamente em 1996 (CUNHA, 2000; CUNHA, 2006).

O Distrito Federal e Brasília foram fundados em 21 de abril de 1960 baseado no desejo de transferir a capital federal do Rio de Janeiro para o centro do país. Ambos foram construídos em apenas 4 anos em um território anteriormente não povoado na região Centro-Oeste, como resultado dos esforços do presidente Juscelino Kubitschek, em conjunto com o arquiteto Oscar Niemeyer, o urbanista

Lúcio Costa, e milhares de trabalhadores civis, conhecidos como "candangos" (HABERMALZ, 2006).

De acordo com o Censo 2010, a população do Distrito Federal é composta por 2.556.121 habitantes, 54% das pessoas nasceram localmente, e os 46% restantes eram imigrantes das outras macrorregiões brasileiras: Norte (2%), Nordeste (23,5%), Centro-Oeste (6,5%), Sudeste (12,5%) e Sul (1,5%) (IBGE, 2010). Segundo estas frequências e outras descrições da literatura, as principais migrações aconteceram da região Nordeste e Sudeste (CODEPLAN, 2013). Além disso, o Brasil foi colonizado por Portugal, e o processo de mistura entre o nativo americano, africano e europeu ocorreu em diferentes proporções nas cinco macrorregiões, o que torna a população brasileira heterogênea (IBGE, 2010; GODINHO et al., 2008; CORDEIRO et al., 2009; PENA et al., 2009). Assim, a população do Distrito Federal é nova, mista formada a partir de populações heterogêneas, de rápido crescimento e instalada em um território artificial.

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 nas populações do Estado de Mato Grosso e do Distrito Federal permitem compreender a diversidade desses marcadores em populações altamente miscigenadas auxiliando na resolução de casos forenses em nosso país.

3. OBJETIVOS

Objetivo Geral

Avaliar a eficiência de 32 polimorfismos de inserção/deleção do cromossomo X nas populações do estado do Mato Grosso e do Brasília.

Objetivos Específicos

- Caracterizar as populações do estado do Mato Grosso e Brasília pela análise de 32 polimorfismos de inserção/deleção no cromossomo X, verificando diversos parâmetros estatísticos: equilíbrio de Hardy-Weinberg, desequilíbrio de ligação entre os pares de marcadores, heterozigosidade esperada e observada, frequência alélica dos marcadores;
- Avaliar a eficiência forense desses marcadores nas populações estudadas;
- Verificar a distância genética entre as populações estudadas e comparar com os dados de outras populações relatados na literatura;
- Identificar a ancestralidade das populações do Estado do Mato Grosso e da cidade de Brasília (DF).

4. MATERIAIS E MÉTODOS

4.1 - Aspectos éticos

Este trabalho se enquadra dentro de um projeto mais amplo intitulado “Caracterização de Marcadores Genéticos Forense na população do Estado de Mato Grosso, Brasil”, coordenado pela Dra. Joyce Aparecida Martins Lopes Ferraz da Faculdade de Medicina de São José do Rio Preto (FAMERP). Portanto, foi aprovado pelo Comitê de Ética em Pesquisa dessa faculdade (Parecer nº 1.566.458, 31 de maio de 2016).

4.2 - Casuística

Para a caracterização genética dos 32 polimorfismos de inserção/deleção, foram coletadas amostras sanguíneas de 303 indivíduos (200 homens e 103 mulheres) não aparentados, nascidos e residentes no Estado do Mato Grosso. A seleção desses indivíduos e a coleta das amostras biológicas foram realizadas entre doadores de sangue do Hemocentro de Cuiabá-MT.

Com relação a Brasília, amostras de sangue de 179 indivíduos (12 homens e 167 mulheres) não aparentados e residentes no Distrito Federal, foram coletadas na Fundação Hemocentro de Brasília. Estas amostras foram doadas pela Polícia Científica Federal

4.3 - Coleta das amostras biológicas e extração do DNA

As amostras de sangue foram coletadas em papel FTA[®] (*Whatman*[®] GE) e armazenadas em temperatura ambiente até momento da extração do DNA, que foi realizada conforme adaptação das metodologias propostas por Singer-Sam, Tanguay e Riggs (1989) e Walsh, Metzger, e Higuchi (1991).

Para a extração do DNA, 2 discos de 1.2mm de papel de filtro FTA[®] (*Whatman*[®] GE), contendo a amostra sanguínea do indivíduo, foram cortados com o auxílio da caneta Harris Micro Punch[®] (*Whatman*[®] GE) e depositados em tubo de 0,2ml. Os discos foram agitados por 5 segundos em 50µL de água ultrapura (Milli-Q[®] - Milipore Corporation). O sobrenadante foi descartado e os discos incubados com 50µL de água ultrapura (Milli-Q[®] - Milipore Corporation) por 30 minutos em temperatura ambiente. O sobrenadante foi descartado e os discos foram incubados em 100µL de Chelex 100 (Bio-Rad Laboratories) na concentração de 5% por pelo

menos 1 hora à 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 e transferiu-se o sobrenadante para novo microtubo, o qual foi armazenado à 4°C.

4.4 – Reação de amplificação dos X-InDels

Neste trabalho foi utilizada a metodologia descrita por Pereira e colaboradores (2012) para a amplificação de 32 marcadores de inserção/deleção no cromossomo X. Para isto foi solicitada a síntese dos 32 pares de *primers* marcados com as fluorescências 6-FAM[®], VIC[®], PET[®], NED[®] da Applied Biosystems[®] (Thermo Fischer Scientific) conforme apresentados na tabela 1.

Tabela 1: Marcadores X-InDels analisados: localização da cada marcador no genoma, alelos possíveis e sequência dos *primers*. **Fonte:** Pereira e colaboradores (2012).

Identificação	Marcador	Posição (pb)*	Alelos	Forward Primers (5'-3')	Reverse Primers (5'-3')	
1	MID2612	rs3048996	10234839	- /ATC	gACCCACGGTGTGGAATTCAG	NED-CACAGCACCAGGAAAATAGC
2	MID3712	rs55877732	12572196	- /GAA	gtttAGTCTTGCTGCAATGTACCC	VIC-TTCAAGGGCAATGATGTTTG
3	MID357	rs25581	12912861	- /TGAGA	gTTTTATAGACTGTGGCCCCC	PET-GTTAGTGGTTGGATTGCTCG
4	MID356	rs25580	12918048	- /CTT	gtttCCAACCTCCACGTGAGAAATG	PET-AGTCTGATGCAGTGGCAAAC
5	MID3703	rs59400186	13711300	- /GTTA	VIC-AGCTTCCAAGTAGTTCTGCC	gTTTGGCTTACTTCCTCCTCC
6	MID3774	rs5901519	13809000	- /ACC	gAAGACGGGAATTGAGTCACC	NED-TTTTTGTGCACAGGCACTCC
7	MID3692	rs67929163	19516252	- /CATAT	6FAM-ACATAAAAGCAAGCTTTGGC	gtttctCCCAGGTGTGTGAACTTTTTTC
8	MID3716	rs63344461	24235114	- /GAG	6FAM-AAAGGGAGCATCTACTCCAG	gtttctAGGGCAATCCAGAATTGGAC
9	MID3690	rs60283667	28984076	- /TCAC	GGGCACCATATTAGGCATGT	VIC-CCCACCATCTAACCCATTTC
10	MID3719	rs3078486	29040938	- /TTAACT	gTTCTTTCTCATCTGGCACCC	VIC-CTATGAAGCCTATAGATTGG
11	MID2089	rs2308280	29157973	- /TTA	VIC-AATCCATTCTGGAATAAGATGTCA	gtttctTCCACTCTCAGGGATTCCTTT
12	MID2692	rs3047852	38262701	- /ATT	gtttctCAAGTTCATATGGTGTCTTGG	PET-TGCATTACACAGAGCAACTC
13	MID3701	rs4030406	45539201	- /ATTA	gtttctAGTTGGAGATGCAATGAAGC	NED-AGAGACAGGTGAATTGAGGC
14	MID198	rs16637	47680386	- /CAACCAAT	6FAM-CAGGCACAGGAGAGGAAGAG	gTCCACCCCTAGTTAAACAGC
15	MID1736	rs2307932	68733480	- /ATA	VIC-GTGAAAGGTGAGCTTGTCTG	gtttctAGGCCTTTTTGGTTAACTGG
16	MID3730	rs3215490	88009689	- /GACA	6FAM-AGGATCCTGACTAAGATAGC	GAATCTCTGGAAACACTTGG
17	MID1511	rs2307707	93392006	- /GTCT	gCTGCCTGGGATTTTTCCCTT	PET-CAGGGGAGAACACCCACTAA 6FAM-
18	MID3740	-	97906546	- /GT	gtttctACTTGCTTTGCTTTTCCCTC	GTACAACTGCAAGGAACRAG***
19	MID3732	rs11277082	98331815	- /ACCTCACTCA	VIC-CAGAGTCATCTATTCCCCAG	gtttctCACCCATGTGGTTTCATTTC
20	MID3727	rs3050111	99165489	- /TT	gtttctGGTGGAATTCTTTTGCATGTG	PET-TTTTGGGAAGCACACTCACC
21	MID3754	rs57608175	116901987	/GGTCATCACGAG	NED-TTTCACCAAGGACTTGAAAGG	gttCAGCTCACACTAGGGCCTTC
22	MID3722	-	118156157	/AAAGTGTACACAT	VIC-TGGCCCTTCTGAGTTCAAAC	gCAGTGTAAATAAGGTGGGAGC
23	MID1361	rs2307557	118748515	- /ACA	6FAM-TCAGTCCTTAACAAGGGAGC	gtttGTCATTGTGAAGGCTACCTG
24	MID243	rs16680	122370414	- /TGT	PET-TACAGTTGGCTGCTTTTCCC	gtttctATACGAAGATCTGTGGGAAC
25	MID2637	rs3053615	124135529	- /CT	gtttctTATGTGTCAAATGGGAGGC	6FAM-AATCCTCAAATCACAGTGGC

Identificação	Marcador	Posição (pb)*	Alelos	Forward Primers (5'-3')	Reverse Primers (5'-3')	
26	MID111	rs16397	127958384	- /GTG	GAGGCAGGGAAATCAGTTAG	NED-TTGATTCCAGCTTTCCCTTT
27	MID3736	rs56162621	130975547	- /CT	6FAM-GGGTTAGGAGCCCCTGCT	gtttcttGGATGTATGACACACAACGC
28	MID3753	rs72417152	131760172	- /GTATAT	GCTACACCAATGGACAGATG	PET-TGTGGTGTGCATGATTTGC
29	MID1839	rs2308035	135695920	- /CA	gttGATATCCCATAACGCCCATTT	NED-TCCTTTTGCTACGCAGACCT
30	MID3760	rs66676381	137369795	- /TTAAA	gCAAGGTTTCGTAATCATTAG	NED-AACCTAGTTCACAACCCCTG
31	MID329	rs25553	147393784	- /TACTCT	gTCTCAAACCTTCCCATGGC	6FAM-AGAAGTTAGAGGGTGTCTGG
32	MID2652	rs3080039	154561961	- /TAA	PET-GCTGCTCTTTGCTTTAATTC	gTATGGTAGGCACTGTGCTAA

* Localização no cromossomo X de acordo com o navegador UCSC Genome Browser, tabela 131.

Fonte: Pereira e colaboradores (2012).

As reações de amplificação por PCR dos 32 marcadores X-InDels foram realizadas em todas as amostras de DNA através de um sistema *multiplex*. O protocolo proposto por Pereira e colaboradores (2012) sugere realizar a amplificação dos 32 X-InDels em uma reação com volume final de 10 μ L. Neste trabalho as reações foram preparadas com a redução do volume final da reação para 5 μ L, porém manteve-se a concentração estipulada do *Multiplex* PCR Master Mix 2x (Qiagen) e primers, como segue:

Reação em Multiplex	Volume
<i>Multiplex</i> PCR Master Mix 2x (Qiagen)	2,5 μ L
Mistura de <i>primers</i> 10x	0,5 μ L
Água ultrapura (mili-Q)	q.s.p. 5,0 μ L
Solução de DNA extraído por Chelex	1,0 μ L (0,3 – 5ng/ μ L)

Obs: A mistura de *primers* contém todos os *primers* na concentração de 1 μ M exceto para os marcadores MID2089, MID3774, MID3727 (2 μ M) e MID3719 (3 μ M).

A ciclagem foi realizada em termociclador Veriti Thermal Cycler (Applied Biosystems[®] – Thermo Fischer Scientific) nas seguintes condições:

95°C - 15 min
 94°C - 30 seg
 60°C - 90 seg
 72°C - 45 seg
 72°C - 60 min

} 29 ciclos

Obs: Esta ciclagem apresenta uma modificação em comparação com a publicada pelo autor. O passo de ciclagens foi diminuído de 30 para 29 ciclos para diminuir o sinal excessivo obtido no eletroferograma (PEREIRA et al. 2012).

As amostras padrão 9947A (Promega) e 9948 (Promega) foram utilizadas como controles na reação de amplificação.

4.4.1 – Preparação para eletroforese capilar

O protocolo original de Pereira e colaboradores (2012) utilizou como padrão de corrida o GeneScan[®] 500 LIZ[®] (Thermo Fischer Scientific). Para as amostras desse estudo foi realizada a adequação dos eletroferogramas para o padrão de corrida GeneScan[®] 600 LIZ[®] (Thermo Fischer Scientific). Os produtos amplificados de todas as amostras foram submetidos à corrida eletroforética em analisador genético ABI3500 (Applied Biosystems[®] – Thermo Fischer Scientific), conforme protocolo abaixo:

Hi-Di [®] <i>Formamide</i> (Thermo Fisher Scientific)	8,7µL
Padrão de tamanho GeneScan [®] 600 LIZ [®] (Thermo Fisher Scientific)	0,3µL
Produto de amplificação	1,0µL
Volume final	10,0µL

Na eletroforese no Analisador Genético ABI3500, utilizou-se POP-4 *Polymer* e capilar de 36 cm (todos da Applied Biosystems[®] - Thermo Fisher Scientific).

4.5 – Determinação dos perfis alélicos e análise estatística

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

Com o auxílio do *software* Arlequin v3.5 (EXCOFFIER; LISCHER, 2010), foram determinadas as frequências alélicas das populações estudadas. O cálculo dos valores de heterozigosidade esperada e observada e o teste exato ao equilíbrio de Hardy-Weinberg foram realizados somente para as amostras femininas, com o auxílio do mesmo *software*. A análise do desequilíbrio de ligação, feita apenas usando os haplótipos masculinos, foi realizada também com o auxílio *software* Arlequin.

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

as chances de se obter um resultado falso-positivo quando muitas comparações são realizadas em um único conjunto de dados. Esse método consiste na divisão do nível de significância estipulado (em nosso caso, 0,05) pelo número de testes que foram executadas.

Diversos parâmetros estatísticos de interesse forense foram calculados com o auxílio da ferramenta de cálculo do banco de dados internacional de X-STRs (ChX-STR – <http://www.chrx-str.org/>), tais como: Poder de Discriminação em mulheres (PD_F), Poder de Discriminação em homens (PD_M) e Chance Média de Exclusão (MEC) para as seguintes situações:

- MEC_T : Chance média de exclusão em trios envolvendo filha, como descrito por Desmarais et al. (1998);
- MEC_D : Chance média de exclusão na ausência da mãe (duo – pai/filha), como descrito por Desmarais et al. (1998);
- MEC_{PGM} : Chance média de exclusão em casos deficiente envolvendo mãe, filha e avó paterna, segundo Krüeger et al. (1968).

Além disso, para avaliar a existência ou não de diferença significativa entre as populações estudadas e compará-las com dados de outras relatadas na literatura, foi realizado o cálculo de distância genética (F_{ST}), 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 Trial⁵ (StatSoft).

Com o auxílio do *software* Structure v2.3.4 (PRITCHARD; STEPHENS; DONNELLY, 2000) foi realizada a comparação para a proporção de ancestralidade das populações do Mato Grosso e de Brasília. 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*.

5. RESULTADOS E DISCUSSÃO

5.1 - Avaliação da metodologia

A reação de amplificação descrita por PEREIRA e colaboradores (2012) foi eficiente para análise dos 32 X-InDels. Os eletroferogramas dos perfis alélicos obtidos em analisador genético ABI3500 de amostras masculina e feminina estão apresentados nas figuras 5 e 6, respectivamente. Pode-se notar que a metodologia foi eficiente, pois todos os marcadores foram bem amplificados, obtendo-se uma precisa determinação dos alelos.

Figura 5: Eletroferograma do perfil alélico de amostra masculina obtido com o multiplex 32 X-InDels. **Fonte:** Dados da pesquisa.

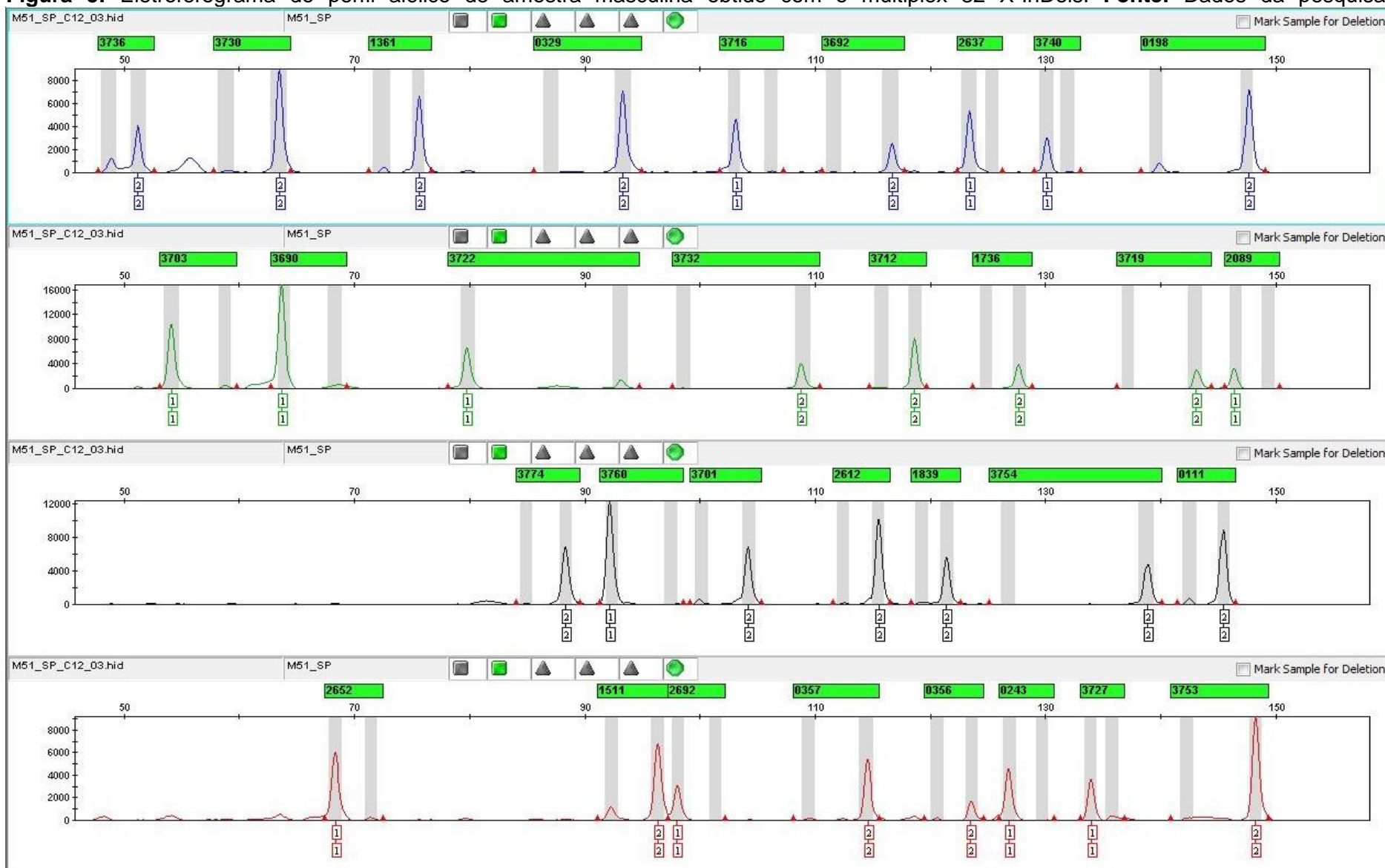
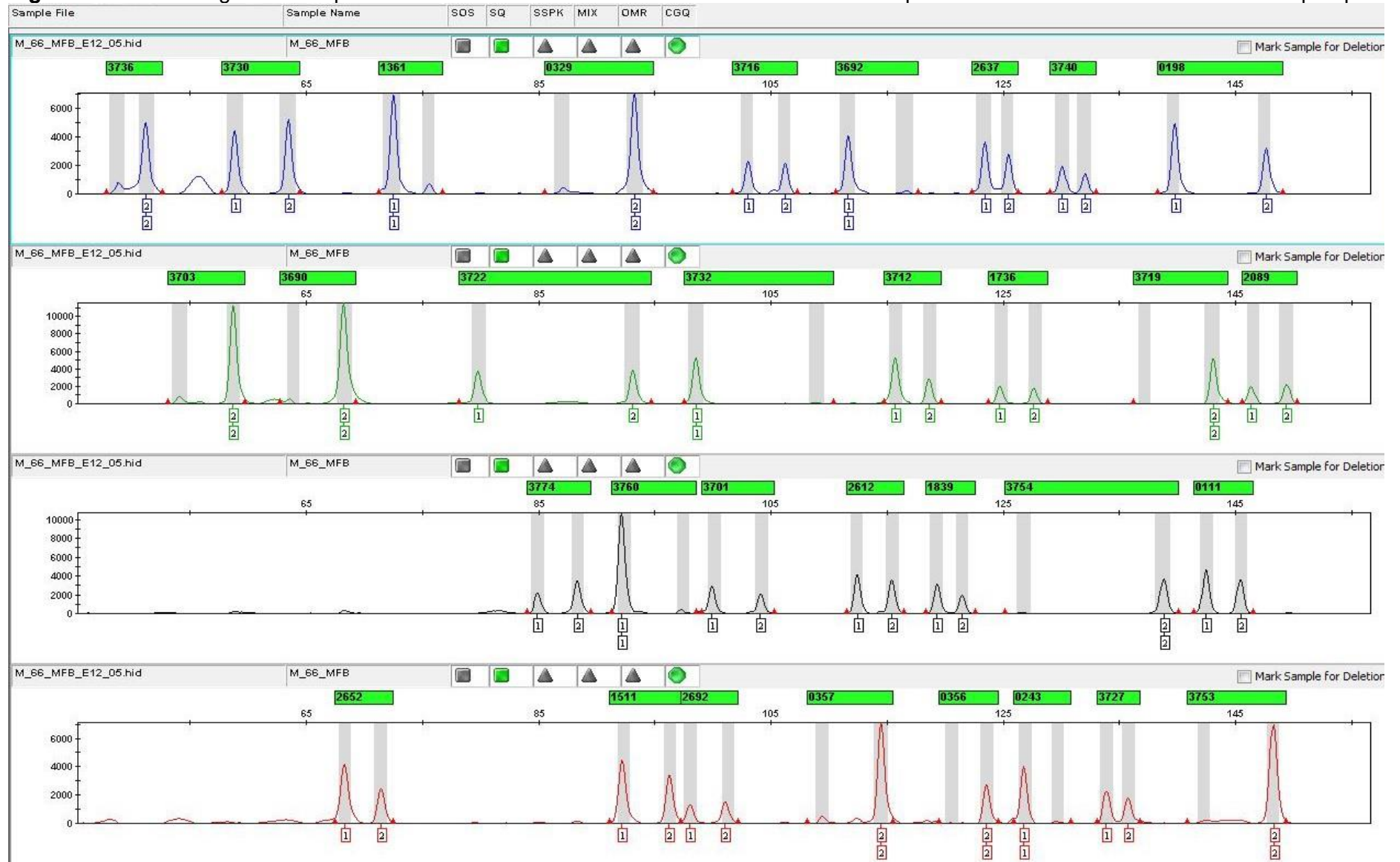


Figura 6: Eletroferograma do perfil alélico de amostra feminina obtido com o multiplex 32 X-InDels. **Fonte:** Dados da pesquisa.



5.2 – Variabilidade genética

A frequência alélica para os 32 X-InDels foi avaliada separadamente por gênero para da cada população estudada. Tanto na população do Estado do Mato Grosso ($F_{ST} \leq 0,0003$; $p = 0,36501 \pm 0,0046$) quanto na população de Brasília ($F_{ST} \leq 0,003$; $p = 0,34937 \pm 0,0052$) não foram observadas diferenças significativas entre as amostras masculinas e femininas. Assim os resultados obtidos para homens e mulheres foram agrupados em uma única amostra para cada população e usados no cálculo final das frequências alélicas e parâmetros forenses, e para comparações populacionais. A Tabela 2 ilustra as distribuições das frequências alélicas das populações estudadas, bem como os valores da diversidade gênica.

Tabela 2: Frequência alélica e diversidade gênica dos 32 marcadores X-InDels nas populações do Mato Grosso e de Brasília. Número de cromossomos analisados: 406 (Mato Grosso) e 346 (Brasília).

Locus	Número rs	Mato Grosso		Brasília		Diversidade Gênica	
		Alelo Curto	Alelo Longo	Alelo Curto	Alelo Longo	Mato Grosso	Brasília
MID3727	rs3050111	0,2414	0,7586	0,2515	0,7486	0,3671	0,3775
MID3753	rs72417152	0,1305	0,8695	0,0896	0,9104	0,2276	0,1636
MID3736	rs56162621	0,5468	0,4532	0,5723	0,4278	0,4968	0,4910
MID3730	rs3215490	0,2562	0,7438	0,2457	0,7543	0,3820	0,3717
MID1361	rs2307557	0,2660	0,7340	0,2370	0,7630	0,3915	0,3627
MID329	rs25553	0,2020	0,7980	0,2457	0,7543	0,3232	0,3717
MID3716	rs63344461	0,5320	0,4680	0,6098	0,3902	0,4992	0,4773
MID3692	rs67929163	0,1823	0,8177	0,2457	0,7543	0,2988	0,3717
MID2637	rs3053615	0,9015	0,0985	0,8786	0,1214	0,1781	0,2139
MID3740	–	0,3030	0,6970	0,3728	0,6272	0,4234	0,4690
MID198	rs16637	0,6232	0,3769	0,6098	0,3902	0,4708	0,4773
MID3703	rs59400186	0,2537	0,7463	0,3121	0,6879	0,3796	0,4307
MID3690	rs60283667	0,4187	0,5813	0,3960	0,6041	0,4880	0,4797
MID3722	–	0,3301	0,6700	0,3035	0,6965	0,4433	0,4240
MID3732	rs11277082	0,2586	0,7414	0,3035	0,6965	0,3844	0,4240
MID3712	rs55877732	0,3350	0,6650	0,3353	0,6647	0,4466	0,4470
MID1736	rs2307932	0,2956	0,7044	0,2717	0,7283	0,4174	0,3969
MID3719	rs3078486	0,2266	0,7734	0,1705	0,8295	0,3514	0,2837
MID2089	rs2308280	0,3325	0,6675	0,2861	0,7139	0,4450	0,4097
MID3774	rs5901519	0,4606	0,5394	0,3526	0,6474	0,4981	0,4579
MID3760	rs66676381	0,7931	0,2069	0,7890	0,2110	0,3290	0,3339
MID3701	rs4030406	0,5296	0,4704	0,4942	0,5058	0,4995	0,5000

MID2612	rs3048996	0,4360	0,5640	0,4798	0,5202	0,4930	0,5000
MID1839	rs2308035	0,5123	0,4877	0,5665	0,4335	0,5000	0,4926
MID3754	rs57608175	0,4335	0,5665	0,3960	0,6041	0,4924	0,4797
MID111	rs16397	0,5936	0,4064	0,5607	0,4393	0,4837	0,4941
MID2652	rs3080039	0,6650	0,3350	0,6156	0,3844	0,4466	0,4746
MID1511	rs2307707	0,6232	0,3769	0,5636	0,4364	0,4708	0,4933
MID2692	rs3047852	0,5271	0,4729	0,5116	0,4884	0,4998	0,5000
MID357	rs25581	0,3424	0,6576	0,3064	0,6936	0,4514	0,4262
MID356	rs25580	0,3645	0,6355	0,3410	0,6590	0,4644	0,4508
MID243	rs16680	0,8325	0,1675	0,8093	0,1908	0,2796	0,3096
Média						0,4164	0,4175

Fonte: Dados da pesquisa.

Como esperado, praticamente todos os marcadores se mostraram altamente polimórficos nas populações estudadas, sendo que a média da diversidade gênica foi de 0,416 ($\pm 0,086$) para a população do Mato Grosso e de 0,417 ($\pm 0,085$) para a população de Brasília. O valor médio da diversidade gênica das duas populações foi maior do que a média identificada no estudo feito por Pereira e colaboradores em populações da Europa (0,388), Ásia (0,347) e Angola e Moçambique (0,408), inclusive maior que a média da diversidade observada na população de algumas regiões da Colômbia, variando de 0,319 a 0,395 (IBARRA et al., 2014). Nessas populações, somente os marcadores MID2637 (Mato Grosso) e MID3753 (Brasília) se mostraram pouco informativos, com menos de 20% de diversidade, conforme demonstrado em outros trabalhos que utilizaram o mesmo painel de marcadores X-InDels (PEREIRA et al., 2012).

Os dados gerados das amostras femininas foram utilizados no teste exato para o equilíbrio de Hardy-Weinberg. Para um nível de significância (p) de 0,0016, obtido usando a correção de Bonferroni para testes múltiplos, não foram detectados desvios significativos nas populações estudadas (Tabela 3).

Segundo Szibor e colaboradores (2003), para avaliação estatística de marcadores no cromossomo X, para sua posterior aplicação forense, há necessidade desses marcadores se apresentarem em equilíbrio de Hardy-Weinberg. Se uma população está em equilíbrio de Hardy-Weinberg, presume-se que esta é infinita, que não há eventos de mutação e seleção, que os cruzamentos são aleatórios e que o fluxo gênico existente não é capaz de alterar a composição alélica desta população. Assim, as frequências de seus alelos permanecem inalteradas ao

longo das gerações e as proporções genóticas atingem um equilíbrio estável (ANDREWS, 2010).

A maioria dos estudos realizados demonstra uma distribuição genótica de acordo com o equilíbrio de Hardy-Weinberg, o que também foi observado neste estudo. Somente para os marcadores MID1736 e MID3719 em Mato Grosso e para os marcadores MID243 e MID3690 em Brasília, foram observados desequilíbrios ($p \leq 0,05$). Entretanto, após a aplicação da correção de Bonferroni (nível de significância (p) de 0,0016) nenhum desequilíbrio foi observado. Diante disso, as frequências alélicas encontradas na amostragem podem ser consideradas representativas para a população estudada.

Tabela 3: Parâmetros estatísticos da diversidade genética para os 32 marcadores X-InDels nas populações femininas estudadas. Tamanho da amostra: 103 (Mato Grosso) e 167 (Brasília).

Locus	Mato Grosso			Brasília		
	Ho	He	P-HWE	Ho	He	P-HWE
MID3727	0,3107	0,3262	0,7609	0,3234	0,3776	0,0676
MID3753	0,2330	0,2360	1,0000	0,1437	0,1640	0,1288
MID3736	0,4660	0,5020	0,5534	0,4790	0,4903	0,8743
MID3730	0,3689	0,3694	1,0000	0,3773	0,3685	0,8343
MID1361	0,3689	0,3978	0,4645	0,3952	0,3654	0,3944
MID329	0,3010	0,3319	0,3732	0,4192	0,3716	0,1416
MID3716	0,4757	0,5017	0,6924	0,4910	0,4742	0,7439
MID3692	0,3107	0,3262	0,7597	0,3353	0,3716	0,2144
MID2637	0,1456	0,1357	1,0000	0,1976	0,2160	0,2785
MID3740	0,4369	0,4267	1,0000	0,4611	0,4727	0,7478
MID198	0,3981	0,4835	0,0980	0,4491	0,4782	0,5179
MID3703	0,3884	0,3978	0,8071	0,4371	0,4279	0,8571
MID3690	0,5243	0,4888	0,5448	0,3892	0,4807	0,0157
MID3722	0,4757	0,4598	0,8303	0,4132	0,4279	0,7162
MID3732	0,3107	0,3376	0,3940	0,4431	0,4301	0,7215
MID3712	0,4854	0,4444	0,3791	0,3892	0,4451	0,1193
MID1736	0,3301	0,4304	0,0223	0,4072	0,4004	1,0000
MID3719	0,2718	0,3694	0,0134	0,2575	0,2759	0,3985
MID2089	0,4660	0,4569	1,0000	0,4192	0,4057	0,7045
MID3774	0,5340	0,5023	0,5566	0,4252	0,4601	0,3992
MID3760	0,3010	0,3085	0,7553	0,3234	0,3323	0,8147
MID3701	0,5049	0,4944	0,8425	0,4910	0,5013	0,8779
MID2612	0,5243	0,4986	0,6907	0,5150	0,5011	0,7558
MID1839	0,4272	0,5013	0,1647	0,4790	0,4920	0,7552
MID3754	0,5340	0,4932	0,4269	0,4311	0,4819	0,1990
MID111	0,5340	0,4977	0,5505	0,4491	0,4911	0,2748

MID2652	0,4854	0,4509	0,5133	0,4611	0,4727	0,7460
MID1511	0,4952	0,4704	0,6746	0,4371	0,4911	0,1590
MID2692	0,4854	0,5024	0,8433	0,4850	0,5014	0,7572
MID357	0,4563	0,4267	0,6421	0,4192	0,4208	1,0000
MID356	0,4563	0,4411	0,8235	0,4431	0,4471	1,0000
MID243	0,2816	0,2704	1,0000	0,2455	0,3070	0,0190

Ho, heterozigose observada; He, heterozigose esperada; P-HWE, valor de p para o equilíbrio de Hardy-Weinberg, Nível de significância de 0,0016 (após correção de Bonferroni para múltiplos testes).

Fonte: Dados da pesquisa.

5.3 – Desequilíbrio de ligação entre os marcadores X-Indels

Com o intuito de definir como construir base de dados genéticos e avaliar os parâmetros forenses nas populações de referência, ao estudarmos marcadores localizados no mesmo cromossomo, é essencial a análise do desequilíbrio de ligação (LD) ou *linkage disequilibrium*.

O desequilíbrio de ligação é um parâmetro estatístico que refere-se à associação não randômica de alelos de diferentes loci, não necessariamente no mesmo cromossomo, e estima o desvio da frequência haplotípica em uma dada população em relação à frequência alélica esperada. Para marcadores com elevado desequilíbrio de ligação, não se pode fazer inferência às frequências haplotípicas baseada na frequência alélica, sendo que a frequência haplotípica observada deve ser calculada diretamente na população.

O teste exato para a avaliação do LD foi realizado para todos os marcadores, utilizando-se os dados das amostras masculinas das populações estudadas. A análise resultou em 496 pareamentos para cada população e os dados estão apresentados nas tabelas 4 e 5, sendo demonstrados apenas os marcadores que apresentaram valor de $p \leq 0,05$.

Tabela 4: Desequilíbrio de ligação dos pares de marcadores analisados na população masculina de Mato Grosso que obtiveram valor de $p \leq 0,05$. Tamanho da amostra: 200.

Par de Indels	Distância em pb	p -LD
MID3736-MID329	16.418.237	0,0198
MID1361-3692	99.232.263	0,0285
MID1361-MID198	71.068.129	0,0060
MID3740-MID3690	68.922.470	0,0197
MID329-MID3722	29.237.627	0,0289
MID2637-MID3732	25.803.714	0,0021

MID1361-MID1736	50.015.035	0,0068
MID329-MID1736	78.660.304	0,0267
MID3690-MID3719	56.862	0,0000*
MID2637-MID2089	94.977.556	0,0099
MID3719-MID2089	117.035	0,0000*
MID3703-MID3774	97.700	0,0000*
MID3736-MID1839	4.720.373	0,0006
MID1736-1839	66.962.440	0,0183
MID3727-MID3754	17.736.498	0,0180
MID3740-MID3754	18.995.441	0,0034
MID1839-MID3754	18.793.933	0,0200
MID3736-MID111	3.017.163	0,0072
MID3760-MID111	9.411.411	0,0233
MID3703-MID2652	140.850.661	0,0073
MID3727-MID1511	5.773.483	0,0044
MID3732-MID1511	56.230.146	0,0030
MID3701-MID1511	47.852.805	0,0339
MID3753-MID2692	93.497.471	0,0314
MID3740-MID2692	59.643.845	0,0412
MID3719-MID2692	9.221.763	0,0146
MID1839-MID2692	97.433.219	0,0147
MID3692-MID357	6.603.391	0,0487
MID3692-MID356	6.598.204	0,0097
MID3701-MID356	32.621.153	0,0467
MID357-MID356	5.187	0,0000*
MID3740-MID243	24.463.868	0,0398
MID2089-MID243	93.212.441	0,0165
MID1511-MID243	28.978.408	0,0193

p -LD, valor de p para o teste de desequilíbrio de ligação; *, valor de p estatisticamente significativo para o teste de desequilíbrio de ligação, Nível de significância de 0,0001 (após correção de Bonferroni para múltiplos testes); ^a, distância em pares de bases (pb) entre os marcadores.

Fonte: Dados da pesquisa.

Tabela 5: Desequilíbrio de ligação dos pares de marcadores analisados na população masculina de Brasília que obtiveram valor de $p \leq 0,05$. Tamanho da amostra: 12.

Par de Indels	Distância em pb	p -LD
MID3727-MID3740	1.258.943	0,0453
MID3690-MID3701	16.555.125	0,0107
MID3712-MID3754	104.329.791	0,0472
MID3719-MID2089	117.035	0,0104
MID2089-MID2652	125.403.988	0,0150
MID3760-MID3754	20.467.808	0,0040
MID3760-MID111	9.411.411	0,0046
MID357-MID356	5.187	0,0023

MID3730-MID329	59.384.095	0,0188
MID329-MID1511	54.001.778	0,0186
MID3692-3760	117.853.543	0,0051
MID3692-MID3754	97.385.735	0,0044

p -LD, valor de p para o teste de desequilíbrio de ligação; ^a, distância em pares de bases (pb) entre os marcadores.

Fonte: Dados da pesquisa.

Neste trabalho, a análise do desequilíbrio de ligação para todos os pares de marcadores, revelou associação significativa entre os marcadores MID356-MID357, mesmo após correção de Bonferroni para múltiplas análises ($p \leq 0,0001$). Tal observação já era esperada, pois estes marcadores apresentam-se muito próximos no cromossomo X (~5,2Kb de distância; Tabela 4) e está em concordância com outros estudos em diferentes populações, como em descendentes de judeus (FERRAGUT et al., 2017), na população da Argentina (CAPUTO et al., 2017), Colômbia (IBARRA et al., 2014), Somália e Iraque (PEREIRA et al., 2011), Portugal, Miranda do Ouro, Espanha (Zamora e Maiorca), Angola, Moçambique, Macau (PEREIRA et al., 2012; PINTO et al., 2015; FERRAGUT et al., 2017) e inclusive na população da Amazônia brasileira, de Belém, PA (FREITAS et al., 2010). Portanto, os marcadores MID356-MID357, devem ser tratados como haplótipos quando forem utilizados juntos.

O par de marcadores, separados por aproximadamente 97,7Kb, MID3703-MID3774 também apresentaram desequilíbrio de ligação, sendo tal fato relatado na população colombiana (IBARRA et al., 2014), em Africanos, Europeus e em populações de diferentes estados brasileiros, inclusive nas tribos nativo-americanas da Amazônia brasileira (RESQUE et al., 2010).

Valores significativos de p foram também observados para os marcadores MID3690-MID3719 e MID3719-MID2089, sendo que para o primeiro par de marcador, o desequilíbrio foi também relatado na população de Angola, Moçambique (PEREIRA et al., 2012), Colômbia (IBARRA et al., 2014), Alemanha e países bálticos (EDELMANN et al., 2016). Para o par de marcador MID3719-MID2089, também foi identificado o desequilíbrio em descendentes de judeus, Espanha (Maiorca) (FERRAGUT et al., 2017), Somália e Portugal (PEREIRA et al., 2012), Iraque (PEREIRA et al., 2011), Alemanha, países bálticos (EDELMANN et al., 2016) e Colômbia (IBARRA et al., 2014), o que sugere que esses marcadores também devem ser analisados em bloco.

No trabalho de Pereira e colaboradores (2012) é sugerido que os marcadores MID3690-MID3719-MID2089 sejam analisados em um único bloco haplotípico, pois na verdade, esses três marcadores estão localizados a uma distância de ~170kb um dos outros, formando um *cluster* no cromossomo. É importante ressaltar que o LD não é decorrente apenas da distância genética entre os marcadores X-InDels, podendo ser ocasionado também por vários outros fatores como: seleção, mutação, deriva aleatória, efeito fundador, miscigenação ou subdivisão populacional (TILLMAR, 2010; ZARRABEITIA et al., 2006).

Para a população de Brasília não foi verificado nenhuma associação entre os pares de marcadores ao contrário das populações mencionadas anteriormente. Este fato pode ser devido a efeitos de amostragem uma vez que, para o cálculo de LD entre pares de marcadores do cromossomo X são utilizados apenas dados de amostras masculinas, e na amostragem dessa população foram analisados somente 12 perfis masculinos, número que pode ter sido insuficiente para identificar uma possível associação.

Para os marcadores em desequilíbrio, as frequências haplotípicas deve ser utilizada no lugar das frequências alélicas nas análises estatísticas forenses. A tabela 6 inclui as frequências observadas para os três blocos haplotípicos na população do Mato Grosso.

Tabela 6: Frequência haplotípica observada na população masculina do Mato Grosso para os marcadores em desequilíbrio de ligação. Tamanho da amostra: 200.

BLOCO I			
357	356	Frequência	
1	1	0,3800	
1	2	0,0000	
2	1	0,0250	
2	2	0,5950	
BLOCO II			
3703	3774	Frequência	
1	1	0,0400	
1	2	0,1950	
2	1	0,3900	
2	2	0,3750	
BLOCO III			
3690	3719	2089	Frequência
1	1	1	0,0100
1	1	2	0,0150
1	2	1	0,1350
1	2	2	0,2600
2	1	1	0,1500
2	1	2	0,0350
2	2	1	0,0200
2	2	2	0,3750

1. Alelo curto; 2. Alelo longo

Fonte: Dados da pesquisa.

5.4 – Parâmetros estatísticos de eficiência forense

Os parâmetros estatísticos indicativos de eficiência forense em relação ao painel dos 32 X-InDels estão detalhados nas tabelas 7 e 8 para cada marcador e haplótipo.

Tabela 7: Parâmetros estatísticos de interesse forense para os marcadores 32 X-Indels na população do Mato Grosso. Tamanho da amostra: 406.

Locus	PD_F	PD_M	MEC_T	MEC_D	MEC_{PGM}
BLOCO I	0,648084	0,500950	0,398085	0,601915	0,398085
BLOCO II	0,823445	0,667650	0,601552	0,398448	0,601552
BLOCO III	0,899284	0,749100	0,711334	0,288666	0,711334
MID3760	0,494812	0,328185	0,274332	0,725668	0,274332
MID3754	0,620460	0,491156	0,370539	0,629461	0,370539
MID3753	0,376627	0,226940	0,201189	0,798811	0,201189
MID3740	0,577131	0,422350	0,333161	0,666839	0,333161
MID3736	0,622781	0,495620	0,372800	0,627200	0,372800
MID3732	0,546367	0,383471	0,309946	0,690054	0,309946
MID3730	0,544331	0,381084	0,308472	0,691528	0,308472
MID3727	0,531275	0,366231	0,299169	0,700831	0,299169
MID3722	0,591112	0,442234	0,344449	0,655551	0,344449
MID3716	0,623968	0,497949	0,373973	0,626027	0,373973
MID3712	0,593319	0,445537	0,346285	0,653715	0,346285
MID3701	0,624122	0,498252	0,374125	0,625875	0,374125
MID3692	0,462899	0,298095	0,253665	0,746335	0,253665
MID2692	0,624263	0,498532	0,374265	0,625735	0,374265
MID2652	0,593319	0,445537	0,346285	0,653715	0,346285
MID2637	0,307881	0,177596	0,161825	0,838175	0,161825
MID2612	0,620798	0,491798	0,370865	0,629135	0,370865
MID1839	0,624848	0,499696	0,374848	0,625152	0,374848
MID1736	0,572729	0,416417	0,329715	0,670285	0,329715
MID1511	0,608454	0,469668	0,359374	0,640626	0,359374
MID1361	0,552262	0,390497	0,314253	0,685747	0,314253
MID329	0,488842	0,322356	0,270399	0,729601	0,270399
MID243	0,441092	0,278874	0,239989	0,760011	0,239989
MID198	0,608454	0,469668	0,359374	0,640626	0,359374
MID111	0,615779	0,482478	0,366086	0,633914	0,366086
Total	0,99999999997	0,99999997	0,999998	0,9996	0,998

PD_F, poder de discriminação em mulheres; PD_M, poder de discriminação em homens; MEC_T, chance média de exclusão em trios envolvendo filha, mãe e suposto pai; MEC_D, chance média de exclusão em duo pai/filha; MEC_{PGM}, chance média de exclusão em casos envolvendo mãe, filha e avó paterna; Bloco I: MID357-MID356; Bloco II: MID3703-MID3774; Bloco III: MID3690-MID3719-MID2089.

Fonte: Dados da pesquisa.

Tabela 8: Parâmetros estatísticos de interesse forense para os marcadores 32 X-Indels na população de Brasília. Tamanho da amostra: 346.

Locus	PD_F	PD_M	MEC_T	MEC_D	MEC_{PGM}
MID3727	0,540324	0,376446	0,305590	0,188223	0,152795
MID3753	0,286367	0,163145	0,149838	0,081574	0,074915
MID3736	0,619616	0,489560	0,369725	0,244780	0,184863
MID3730	0,535203	0,370622	0,301942	0,185311	0,150971
MID1361	0,527115	0,361651	0,296256	0,180826	0,148128
MID329	0,535203	0,370622	0,301942	0,185311	0,150971
MID3716	0,612064	0,475875	0,362646	0,237937	0,181323
MID3692	0,535203	0,370622	0,301942	0,185311	0,150971
MID2637	0,358387	0,213324	0,190570	0,106662	0,095285
MID3740	0,607259	0,467656	0,358305	0,233828	0,179152
MID198	0,612064	0,475875	0,362646	0,237937	0,181323
MID3703	0,582236	0,429417	0,337218	0,214709	0,168609
MID3690	0,613470	0,478347	0,363939	0,239174	0,181970
MID3722	0,577425	0,422752	0,333392	0,211376	0,166696
MID3732	0,577425	0,422752	0,333392	0,211376	0,166696
MID3712	0,593441	0,445721	0,346388	0,222861	0,173194
MID1736	0,556565	0,395740	0,317435	0,197870	0,158717
MID3719	0,445735	0,282886	0,242874	0,141443	0,121437
MID2089	0,566707	0,408519	0,325075	0,204260	0,162538
MID3774	0,600441	0,456546	0,352329	0,228273	0,176165
MID3760	0,499601	0,332935	0,277512	0,166467	0,138756
MID3701	0,624967	0,499933	0,374967	0,249967	0,187483
MID2612	0,624590	0,499181	0,374590	0,249591	0,187295
MID1839	0,620465	0,491163	0,370543	0,245582	0,185271
MID3754	0,613470	0,478347	0,363939	0,239174	0,181970
MID111	0,621235	0,492633	0,371290	0,246317	0,185645
MID2652	0,610562	0,473269	0,361277	0,236634	0,180639
MID1511	0,620860	0,491915	0,370925	0,245958	0,185462
MID2692	0,624866	0,499733	0,374866	0,249866	0,187433
MID357	0,579068	0,425007	0,334692	0,212504	0,167346
MID356	0,595901	0,449463	0,348455	0,224732	0,174227
MID243	0,474487	0,308729	0,261072	0,154364	0,130536
Total	0,999999999998	0,99999998	0,999997	0,9995	0,997

PD_F, poder de discriminação em mulheres; PD_M, poder de discriminação em homens; MEC_T, chance média de exclusão em trios envolvendo filha, mãe e suposto pai; MEC_D, chance média de exclusão em duo pai/filha; MEC_{PGM}, chance média de exclusão em casos envolvendo mãe, filha e avó paterna.

Fonte: Dados da pesquisa.

Dentre os diversos parâmetros estatísticos calculados, o Poder de Discriminação (PD) consiste na probabilidade de dois indivíduos escolhidos ao acaso na população apresentarem genótipos diferentes. Neste trabalho, ao avaliar a eficiência do sistema como um todo, o poder de discriminação obtido foi de 0,999999999997 e 0,99999997 em mulheres e homens, respectivamente, para a população do Mato Grosso. Em relação à população de Brasília, para mulheres o valor combinado de PD_F foi de 0,999999999998 e o valor combinado para homens foi de PD_M 0,99999998.

Com relação ao Poder de Exclusão (PE, ou chance média de exclusão, MEC) trata-se da probabilidade média de um marcador genético excluir a paternidade de um indivíduo qualquer da população. MEC foi introduzida para marcadores autossômicos em trios envolvendo mãe, filho e suposto pai (CHAKRABORTY et al., 1999); no entanto, este parâmetro não se aplica quando marcadores do cromossomo X são testados. Para este caso, foram descritas expressões para o cálculo da MEC em situações envolvendo o trio: mãe, filha e suposto pai (MEC_T) ou quando apenas o duo, filha e suposto pai é testado (MEC_D) (DESMARAIIS et al., 1998). Para o sistema de 32 marcadores X-Indels estudado, o valor combinado de MEC_T para a população do Mato Grosso foi de 0,999998 e o valor de MEC_D foi 0,9996. Nos casos deficientes, ou seja, onde o suposto pai está ausente e os genótipos disponíveis são da mãe, filha e avó paterna, o valor combinado de MEC_{PGM} foi 0,998. Já para a população de Brasília, os valores estimados foram bem semelhantes, sendo que os valores combinados foram de 0,999997, 0,9995 e 0,9997, para MEC_T , MEC_D e MEC_{PGM} , respectivamente.

Ainda em relação aos parâmetros determinados, nas populações em estudo, os valores obtidos permitiram um nível satisfatório de discriminação em casos forenses como já observado em outras populações. Quando comparados com dados de outras populações, os 32 marcadores X-Indels foram mais discriminativos na população do Mato Grosso e de Brasília do que em populações de Portugal e Macau e tão eficientes quanto foram para a população de Somália, Angola e Moçambique, conforme tabela 9 (PEREIRA et al., 2012).

Tabela 9: Parâmetros estatísticos indicativos de interesse forense obtido para o conjunto dos 32 X-Indels.

	Somália	Ang+Moç	Portugal	Macau	Mato Grosso	Brasília
PD _{Mulheres}	0,999999999998	0,999999999993	0,99999999997	0,9999999995	0,99999999997	0,99999999998
PD _{Homens}	0,99999998	0,99999995	0,9999997	0,9999990	0,9999997	0,99999998
MEC _{Trios}	0,999998	0,999996	0,999990	0,99997	0,999998	0,999997
MEC _{Duos}	0,9996	0,9995	0,9990	0,998	0,9996	0,9995

Ang, Angola; Moç, Moçambique.

Fonte: Dados da pesquisa.

5.5 – Comparação genética entre as populações e análise de ancestralidade

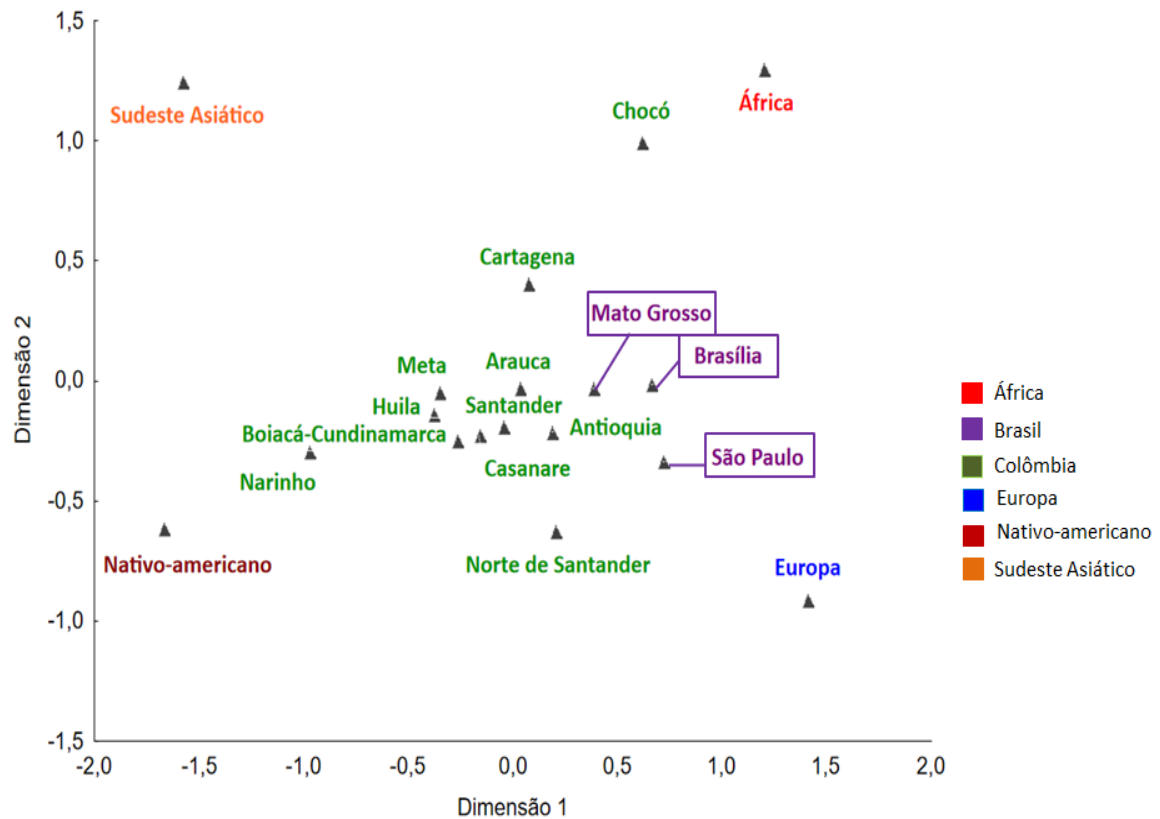
A população do Brasil é altamente miscigenada. No presente estudo, com os dados obtidos foi realizada a análise da distância genética (F_{ST}) entre as populações do estado do Mato Grosso e de Brasília. Além disso, as mesmas foram comparadas com a população do estado de São Paulo e populações de outros países com dados disponíveis na literatura que analisaram o mesmo conjunto de marcadores X-InDels (Tabela 10).

Tabela 10: Amostras usadas na comparação entre as populações, número de cromossomos analisados (N) e respectivas referências.

População	N	Referência
EUROPA		
Portugal	324	Pereira et al. (2012)
Espanha (Mairoca)	112	Ferragut et al. (2017)
ÁFRICA		
Angola e Moçambique	116	Pereira et al. (2012)
AMÉRICA LATINA		
Brasil		
Mato Grosso	406	Este estudo
Brasília	346	Este estudo
São Paulo	782	Martinez (2017)
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	
NATIVO-AMERICANO		
Pasto (Colômbia)	83	Ibarra et al. (2014)
SUDESTE ASIÁTICO		
Macau	100	Pereira et al. (2012)
Taiwan	20	Pereira et al. (2015)

Os valores de distância genética (F_{ST}) e os respectivos valores de p estão apresentados na tabela 12 no final desse subitem. A fim de melhorar a visualização dos resultados, com os valores de distância genética (F_{ST}) foi realizada uma análise em escalonamento multidimensional (MDS) em 2 dimensões (Figura 7). A adição de uma terceira dimensão ao gráfico MDS não revelou diferenças importantes (dados não mostrados) e o valor de stress (0,097) foi apenas um pouco menor do que o obtido para duas dimensões (0,032).

Figura 7: Gráfico MDS bidimensional elaborado a partir da distância genética (F_{ST}) entre África, Europa, Sudeste Asiático, Nativo-americano, diversas populações da Colômbia e Brasil (São Paulo, Mato Grosso e Brasília).



Fonte: Dados da pesquisa.

Os dados indicam uma proximidade genética entre as populações do estado do Mato Grosso e a cidade Brasília (DF), ou seja, as populações do centro-oeste brasileiro são semelhantes geneticamente (valor obtido de $F_{ST} = 0,0020$; $p = 0,0344$), como observado no gráfico MDS. Tal fato também foi observado para STRs no cromossomo X em populações do centro-oeste (MARTINS et al., 2017). Com relação a São Paulo, ambas as populações se mostraram próximas, entretanto é importante ressaltar que a região centro-oeste é constituída por indivíduos de todas as macrorregiões brasileiras, ressaltando a importância de caracterizarmos geneticamente toda a população brasileira para este conjunto de marcadores X-Indels. Ainda em relação às populações deste estudo, os maiores valores obtidos de

F_{ST} (indicando maior distância genética) foram entre as populações do Mato Grosso e Brasília e o sudeste asiático.

As três populações brasileiras se apresentam próximas da população da Europa, demonstrando a forte influência genética europeia nas populações brasileiras, gerada no período da colonização do Brasil por Portugal. Em comparação com as populações da Colômbia, a população do Mato Grosso e de Brasília são mais similares a três departamentos colombianos (Antioquia, Santander e Arauca), que são caracterizados por possuir maior componente europeu.

Sabe-se que a população indígena é a mais antiga a habitar o Brasil, entretanto ao compararmos as populações deste estudo com a população nativo-americana observou-se uma maior distância genética. Tal fato pode ser explicado, pois não há na literatura dados dos marcadores estudados para as populações nativas do Brasil, por esse motivo utilizou-se os dados obtidos para os nativos de Pasto na Colômbia (IBARRA et al., 2014).

A formação da população brasileira é caracterizada por grande diversidade étnica e intensa miscigenação. A elevada miscigenação ocorreu no período colonial com a chegada dos imigrantes 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.

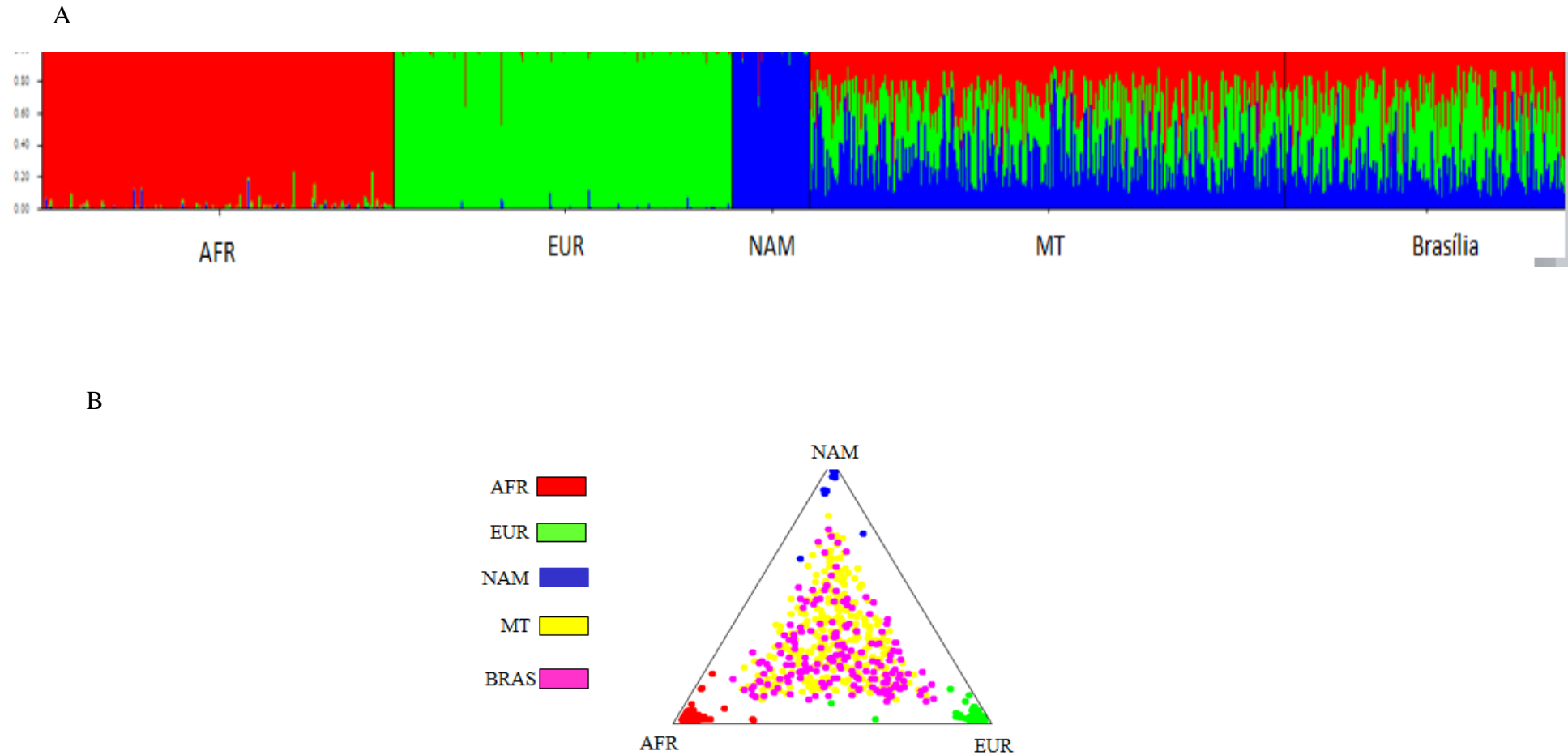
Segundo Moura e colaboradores (2015) a população do centro-oeste e sudeste do Brasil possui semelhante média de contribuição ancestral para marcadores autossômicos (Centro-Oeste: 64% de europeus, 24% de africanos e 12% de nativo-americanos; Sudeste: 67% de europeus, 23% de africanos e 10% de nativo-americanos). Em outro estudo, foi verificada a contribuição ancestral baseado em 24 marcadores X-Indels em quatro regiões brasileiras e confirmaram semelhanças entre o Sudeste e o Nordeste (Sudeste: 38% de europeus, 35% de africanos e 27% de nativo-americanos; Nordeste: 37% de europeus, 36% de africanos e 27% de nativo-americanos) (RESQUE et al., 2010). Embora o centro-oeste não tenha sido analisado, acredita-se que a composição de ancestralidade no cromossomo X é semelhante para essas regiões, como observado em autossômicos (MOURA et al., 2015).

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 do Mato Grosso e de Brasília.

Para essa análise foram utilizados os dados de africanos e europeus publicados por Pereira e colaboradores (2012).

As proporções de ancestralidade identificadas com base em marcadores X-InDel para as populações estudadas estão apresentadas na figura 8 e confirmam o sugerido por Resque e colaboradores (2010).

Figura 8: A - Representação esquemática da mistura estimada para cada indivíduo das populações de Mato Grosso (MT) e Brasília utilizando 32 marcadores X-InDels obtido através da análise do STRUCTURE baseado na contribuição africana (AFR), europeia (EUR) e nativo-americana (NAM) (K=3); B - Dispersão dos indivíduos analisados entre os três grupos populacionais de referência.



Fonte: Dados da pesquisa.

Como pode ser observado na tabela 11, as proporções de ancestralidade das populações estudadas são muito semelhantes, apresentando uma contribuição praticamente equilibrada de europeu, africano e nativo-americano.

Tabela 11: Proporção de ancestralidade (%) para as amostras testadas de populações de africanos, europeus e nativo-americanos (PEREIRA et al., 2012) e as do Mato Grosso e de Brasília.

	Africano	Europeu	Nativo-americano
Africanos	98%	1%	1%
Europeus	1%	98%	1%
Nativo-americanos	1%	1%	98%
Mato Grosso	34%	34%	32%
Brasília	35%	38%	26%

Fonte: Dados da pesquisa.

A proporção de ancestralidade identificada para a população do Mato Grosso foi 34% de europeu, 34% para africano e 32% para nativo-americano e para a população de Brasília foi 38% de europeu, 35% para africano e 26% para nativo-americano, confirmando as semelhanças em as populações das regiões Centro-Oeste e Sudeste, como observados em marcadores autossômicos. As diferenças das proporções de ancestralidade baseadas em marcadores autossômicos e marcadores no cromossomo X são esperadas em populações formadas pela miscigenação de diferentes grupos étnicos com o relacionamento entre homens (apenas um cromossomo X) de um grupo colonizador e mulheres (dois cromossomos X) de dois grupos de origens geográficas distintas (nativo-americanos e africanos), como observado na formação da população brasileira.

A população da região Centro-Oeste do país é miscigenada, principalmente por causa do fluxo migratório ocorrido ao longo do tempo. Dos séculos XVI ao XVIII os primeiros europeus chegaram à região centro-oeste, atrás de ouro e pedras preciosas, eram os bandeirantes que desbravaram sertões em todo o território brasileiro. No final do século XIX, foi iniciada a ocupação dos cursos superior e médio dos rios Tocantins e Araguaia, envolvendo portugueses, indígenas, negros e mestiços (HANDELMANN, 1931).

Já no século XX, ocorreram três principais fluxos de migração para o Centro-Oeste, o primeiro, quando migrantes gaúchos lideraram o desenvolvimento de

estradas de ferro e produção de mate, além da agropecuária. O segundo, no governo de Getúlio Vargas na primeira dita "Marcha para o Oeste", com a intenção de povoar a região central do Brasil. E a terceira e mais efetiva, no governo de Juscelino Kubitschek com a construção de Brasília, uma nova "Marcha para o Oeste" foi instaurada, e deslocou principalmente migrantes nordestinos para a região (CASSIANO, 2002).

Os dados de informação de ancestralidade obtidos neste trabalho, referentes aos estados do Mato Grosso e de Brasília, pertencentes à região centro-oeste do Brasil, confirmam a alta miscigenação da população brasileira.

Tabela 12: Distância genética (F_{ST}) entre as populações (abaixo da diagonal) e os valores de p correspondentes (acima da diagonal).

	Antioquia ¹	Boiacá-Cundinamarca ¹	Huila ¹	Santander ¹	Norte de Santander ¹	Arauca ¹	Casanare ¹	Meta ¹	Cartagena ¹	Chocó ¹	Narinho ¹	Nativo de Pasto ¹	África	Europa	Sudeste Asiático	São Paulo	Brasília	Mato Grosso
Antioquia ¹	*	0,0366	0,0074	0,4532	0,0252	0,0349	0,3296	0,0408	0,0003	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0003	0,0080	0,0890
Boiacá-Cundinamarca ¹	0,0050	*	0,1601	0,4255	0,0369	0,0177	0,5257	0,5728	0,0004	0,0000	0,0012	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Huila ¹	0,0081	0,0021	*	0,1398	0,0340	0,0047	0,3246	0,2938	0,0012	0,0000	0,0139	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Santander ¹	0,0001	0,0000	0,0020	*	0,0451	0,0536	0,5032	0,2947	0,0005	0,0000	0,0001	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0042
Norte de Santander	0,0136	0,0124	0,0125	0,0108	*	0,0196	0,1412	0,0361	0,0071	0,0000	0,0008	0,0000	0,0000	0,0000	0,0000	0,0434	0,0182	0,0168
Arauca ¹	0,0051	0,0066	0,0089	0,0036	0,0159	*	0,7702	0,0455	0,0008	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0022
Casanare ¹	0,0024	-0,0019	0,0025	-0,0013	0,0117	-0,0067	*	0,1464	0,1031	0,0013	0,1992	0,0013	0,0000	0,0000	0,0001	0,0077	0,0403	0,1232
Meta ¹	0,0084	-0,0014	0,0016	0,0014	0,0160	0,0086	0,0103	*	0,0131	0,0000	0,0056	0,0000	0,0000	0,0000	0,0000	0,0001	0,0006	0,0031
Cartagena ¹	0,0138	0,0132	0,0109	0,0101	0,0208	0,0114	0,0105	0,0122	*	0,0063	0,0002	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0004
Chocó ¹	0,0404	0,0488	0,0442	0,0429	0,0539	0,0374	0,0426	0,0451	0,0094	*	0,0000	0,0000	0,0101	0,0000	0,0000	0,0000	0,0000	0,0000
Narinho ¹	0,0346	0,0177	0,0092	0,0202	0,0396	0,0268	0,0070	0,0192	0,0245	0,0605	*	0,0940	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Nativo de Pasto ¹	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
África	0,0640	0,0838	0,0766	0,0714	0,0845	0,0619	0,0652	0,0813	0,0307	0,0071	0,0970	0,1433	*	0,0000	0,0000	0,0000	0,0000	0,0000
Europa	0,0381	0,0602	0,0673	0,0485	0,0356	0,0608	0,0731	0,0629	0,0627	0,0843	0,1245	0,1809	0,1035	*	0,0000	0,0000	0,0000	0,0000
Sudeste Asiático	0,0929	0,0673	0,0594	0,0798	0,0678	0,0873	0,0919	0,0676	0,0734	0,1143	0,0709	0,0937	0,1609	0,1349	*	0,0000	0,0000	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	0,0620	0,0132	0,0858	*	0,0007	0,0002
Brasília	0,0064	0,0227	0,0225	0,0125	0,0160	0,0157	0,0180	0,0227	0,0118	0,0248	0,0544	0,0979	0,0392	0,0245	0,1019	0,0046	*	0,0344
Mato Grosso	0,0020	0,0120	0,0132	0,0052	0,0123	0,0067	0,0080	0,0122	0,0071	0,0252	0,0392	0,0785	0,0448	0,0337	0,0872	0,0052	0,0020	*

¹, departamentos da Colômbia.

Fonte: Dados da pesquisa.

6. CONCLUSÃO

- Foram estabelecidas as tabelas de frequências alélicas para os 32 marcadores X-InDels na população brasileira do estado do Mato Grosso-MT e Brasília (DF);
- O método para análise dos 32 X-InDels foi eficiente, e a avaliação da eficiência forense mostrou que o sistema foi altamente discriminativo para as populações estudadas;
- Os cálculos estatísticos indicaram que para a população de Brasília, os marcadores InDels analisados não estão em desequilíbrio de ligação, entretanto tal fato pode ser devido a efeitos de amostragem, o que sugere a análise de mais amostras masculinas para esta população;
- Na maioria das comparações, valores significativos de distância genética foram obtidos entre as populações estudadas e as estrangeiras com dados publicados na literatura. Observou-se maior proximidade dessas populações com o estado de São Paulo e com três departamentos colombianos (Antioquia, Santander e Arauca), seguido dos europeus (Portugal e Espanha);
- As proporções de ancestralidade identificada nas populações estudadas foram similares ao descrito na literatura, o que reafirma a importante contribuição do cromossomo X de nativo-americanos para a formação da população do centro-oeste brasileiro;
- As populações estudadas do estado do Mato Grosso e de Brasília (DF) apresentam diversidade para os polimorfismos analisados e confirmam a miscigenação brasileira.

7. REFERÊNCIAS

- ALVAREZ-CUBERO, M.J. et al. Genetic identification of missing persons: DNA analysis of human remains and compromised samples. **Pathobiology**. v. 79. n. p. 228–38. 2012.
- ANDREWS, C. The Hardy-Weinberg Principle. *Nature Education Knowledge*. v. 3. n. 10. p. 65. 2010.
- BEIGUELMAN, B. **Dinâmica dos genes nas famílias e nas populações**. 2. ed. Ribeirão Preto: Sociedade Brasileira de Genética, 1996.
- BONACCORSO, N. Análise forense do DNA. Monografia apresentada em 2000 no Concurso de Ingresso para Professor da ACADEPOL e atualizada em 2004.
- BONFERRONI, C.E. Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze. v. 8. p. 3-62. 1936.
- BRASIL. Ministério do Planejamento. Orçamento e Gestão. Instituto Brasileiro de Geografia e Estatística. **Pesquisa Nacional por Amostra de Domicílios: Estado de Mato Grosso**. 2013. Disponível em: http://www.ibge.gov.br/estadosat/temas.php?sigla=mt&tema=pnad_2013 Acesso em: 03 set 2015.
- BUDOWLE, B. et al. CODIS and PCR based short tandem repeat loci: law enforcement tools. In: EUROPEAN SYMPOSIUM ON HUMAN IDENTIFICATION, 2nd, 1998, Madison. **Proceedings**. Madison: Promega Corporation, 1988, p. 73-88.
- BUENO, E. Brasil: uma história. 2. ed. São Paulo: Ática, 2003. 447 p.
- BUTLER, J.M. **Forensic DNA Typing: Biology, Technology, and Genetics of STR Markers**. 2 ed. Nova Iorque: Elsevier Academic Press. 2005. 688p.
- BUTLER, J.M. Short tandem repeat typing technologies used in human identity testing. **BioTechniques**. v. 43. n. 4. p. ii-v. 2007.
- BUTLER, J.M. **Advanced Topics in Forensic DNA Typing: Methodology**. 1 ed. Nova Iorque: Elsevier Academic Press. 2011. 704p.
- CALABREZ, M.C.T. **Influência do calor na análise de DNA extraído de sangue e tecidos humanos**: importância para a identificação de corpos carbonizados. 1999. 92f. Tese (Doutorado) – Faculdade de Medicina da Universidade de São Paulo, 1999.
- CASSIANO, L.C. *Marcha para o Oeste: um itinerário para o Estado Novo (1937-1945)*. 2002. Dissertação (Mestrado em História). Universidade de Brasília, 2002, p.69.
- 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. et al. The utility of short tandem repeat loci beyond human identification: implications for development of new DNA typing systems. **Electrophoresis**, v. 20, n. 8, p. 1682–96, jun. 1999.
- CHOW, J. C. et al. Silencing of the mammalian X chromosome. **Annu. Rev. Genomics Hum. Genet.**, v. 6, p. 69-92, 2005.
- CLIFFORD, R.; EDMONSON, M.; HU, Y.; NGUYEN, C.; SCHERPBIER, T.; BUETOW, K. H. Expression-based genetic physical maps of single nucleotide polymorphisms identified by the cancer genome anatomy project. **Genome Research**, v. 10, p. 1259-1265, 2000.
- CORDEIRO, Q. et al. A review of psychiatric genetics research in the Brazilian population. **Rev Bras Psiquiatr.**, v. 31. n. 2. p. 154-162. 2009.

- CUNHA, J. M. P. da. Dinâmica migratória e o processo de ocupação do Centro-Oeste brasileiro: o caso de Mato Grosso. **R. bras. Est. Pop.** v. 23. n. 1. 87-107. 2006.
- CUNHA, J. M. P. da . Migrações no Centro-Oeste Brasileiro: as tendências e características do período de 1986- 1996. Cadernos de Demografia, SEDUH - Brasília - DF, v. 13, 2000.
- DAWSON, E. et al. A SNP resource for human chromosome 22: extracting dense clusters of SNPs from the genomic sequence. **Genome Research**, v. 11, p. 170-178, 2001.
- DESMARAIS, D.; ZHONG, Y.; CHAKRABORTY, R.; PERREAULT, C.; BUSQUE, L. Development of a e Esa(HUMARA). **J. Forensic Sci.**, v.43, p.1046–1049, 1998.
- DONG, C. et al. Development of three X-linked tetrameric microsatellite markers for forensic purposes. **Mol Biol Rep.** 2014. In Press.
- EDELMANN, J. et al. X-chromosomal 21-indel marker panel in German and Baltic populations. **Int J Legal Med.** v.130 n. 2. p.357-360. 2016.
- 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.
- FERREIRA, A. C. Conquista colonial, resistência indígena e formação do Estado-Nacional: Os índios Guaicuru e Guana no Mato Grosso dos séculos XVIII-XIX. Revista de Antropologia-USP, São Paulo, v. 52, n. 1. 2009.
- FRANÇA, G.V. Medicina Legal. 6. ed., Rio de Janeiro, **Guanabara-Koogan**, 2001. cap. 3, p.32-63.
- 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.
- GODINHO, N. M. O. et al. Regional patterns of genetic admixture in South America. **Forensic Science International: Genetics.** v.1. n. 1. p. 329–30. 2009.
- 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.
- HABERMALZ U. Brasília - eine Stadt aus dem Nichts: auf den Spuren Oscar Niemeyers. Berner Fachhochschule, 2006.
- 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.
- HANDELMANN, G. H. **História do Brasil.** Instituto Histórico e Geográfico Brasileiro. v.1, 1931.
- 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**, v9, n.1. p. e87202. 2014.
- ISRAR, M. et al. Development and characterization of a new 12-plex ChrX miniSTR system. **International Journal of Legal Medicine.** v. 128. n. 4. p. 595-8. 2014
- IBGE. Instituto Brasileira de Geografia e Estatística. **Censo Demográfico de 2010.** Rio de Janeiro: IBGE; 2010.
- JEFFREYS, A. J.; WILSON, V; THEIN, S. L. Hipervariable “minisatellite” regions in human DNA. **Nature**, v.314, p.67-73, 1985a.

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

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

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

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

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. **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

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.

MARTINS, J.A. et al. Genetic characterization of an X-STR decaplex system in the State of Mato Grosso, Brazil: distribution, forensic efficiency and population structure. **Int J Legal Med**. v. 131. n. 6. p. 1523-1530. 2017.

MENDES, M. A. *História e geografia de Mato Grosso*. Cuiabá: Cafarnaum. 2012.

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

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

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.

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

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

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.

- PANETO, G. G. **Análise de polimorfismos do DNA mitocondrial e indivíduos residentes na grande São Paulo para aplicação na identificação humana**. 2010. 118 f. Tese (Doutorado em Biociências e Biotecnologia Aplicadas á Farmácia) - Faculdade de Ciências Farmacêuticas, Universidade Estadual Paulista, Araraquara, 2010.
- PENA, S. D. J. Segurança Pública: determinação de identidade genética pelo DNA. **Parcerias Estratégicas**, n. 20, p.447-460, 2005.
- PENA S. D. J. et al. DNA tests probe the genomic ancestry of Brazilians. **Braz J Med Biol Res.** v. 42(10), p. 870-876. 2009.
- PEPINSKI, W. Population genetics of 30 InDels in populations of Poland and Taiwan. **Mol Biol Rep.** v. 40. n. 7. 2013.
- PEREIRA, R.; GOMES, I.; AMORIM, A.; GUSMÃO, L. Genetic diversity of 10 X chromosome STRs in northern Portugal. **International Journal of Legal Medicine**, v. 121, p. 192-197, 2007.
- 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.
- PEREIRA, R. et al. A new multiplex for human identification using insertion/deletion polymorphisms. **Electrophoresis.** v. 30. n. 21. p. 3682-90. 2009.
- 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. e137-138. 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.
- PINHEIRO, Maria de Fátima. Genética e biologia forense, e criminalística. In: Faculdade de Medicina da Universidade do Porto. In: NOÇÕES GERAIS SOBRE OUTRAS CIÊNCIAS FORENSES MEDICINA LEGAL, 2004. 57p.
- PINTO, N. et al. Assessing paternities with inconclusive STR results: The suitability of bi-allelic markers. **Forensic Sci Int Genet.** v. 7. n. 1. p. 16-21. 2013.
- 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.
- PHILLIPS, C.; FONDEVILA, M.; GARCIA-MAGARINOS, M.; RODRIGUEZ, A.; SALAS, A.; CARRACEDO, A.; LAREU, M. V. Resolving relationship tests that show ambiguous STR results using autosomal SNPs as supplementary markers. **Forensic Science International Genetics.** V. 2 , p. 198-204. 2008.
- PRIETO-FERNÁNDEZ, E. et al. Development of a new highly efficient 17 X-STR multiplex for forensic purposes. **Electrophoresis.** v. 37. p. 1651-1658. 2016.
- 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.849-852. 2010.
- 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. 2014.

- SANTOS-LOPES, S.S. et al. A Worldwide Phylogeography for the Human X Chromosome. **PLoS ONE**. v. 2. n. 6. p. e557. 2007.
- SCHAFFNER, S.F. The X chromosome in population genetics. **Nat Rev Genet**. v. 5. n. 1. p. 43-51. 2004.
- 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. 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.
- SILVA, V. W. Controle e domínio territorial no sul do estado de Mato Grosso: uma análise da atuação da Cia Matte Larangeira no período de 1883 a 1937. *Agrária*. 2011; (15). 102-125.
- SIQUEIRA, E. M. História de Mato Grosso. Cuiabá: Entrelinhas. 2009.
- SUN, K. et al. Development of an 18 X-InDel multiplex PCR system. **Fa Yi Xue Za Zhi**. v. 30. n. 2. 2014. p. 101-9.
- 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. 93-99. 2007.
- TANDE, C.M. DNA Typing: A New Investigatory Tool. **Duke Law Journal**, v.1989, n.2, p.474-494, 1989.
- TILLMAR, A. **Populations and Statistics in Forensic Genetics**. PhD-thesis. Linköping University Medical Dissertations N°. 1175. 2010.
- TILLMAR, A. et al. DNA Commission of the International Society for Forensic Genetics (ISFG): Guidelines on the use of X-STRs in kinship analysis. **Forensic Sci Int Genet**. v. 29. p. 269-75. 2017.
- TORRES, S.R.R. et al. Population genetic data and forensic parameters of 30 autosomal InDel markers in Santa Catarina State population, Southern Brazil. **Mol Biol Rep**. 2014.
- YANG, N.; Li, H.; CRISWELL, L. A.; GREGERSEN, P. K.; ALARCON-RIQUELME, M. E.; KITTLES, R.; SHIGETA, R. et al., **Hum. Genet.**,118, p. 382-392. 2005.
- YUAN, L. et al. Population genetics analysis of 38 STR loci in the she population from Fujian province of China. **Legal Medicine**. 2014. In Press.
- 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, Z. et al. Population genetics of 30 insertion-deletion polymorphisms in two Chinese populations using Qiagen Investigator® DIPplex kit. **Forensic Sci Int Genet**. v. 11. 2014.
- 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.
- WEBER, J.L. et al. Human diallelic insertion/deletion polymorphisms. **Am J Hum Genet**. v. 71. n. 4. p. 854-62. 2002.
- ZIDKOVA, A.; HORINEK, A.; KEBRDLOVA, V.; KORABECNA, M. Application of the new insertion-deletion polymorphism kit for forensic identification and parentage testing on the Czech population. **Int J Legal Méd**. 2011.

ANEXO I

Anexo 1 – Aprovação do Comitê de Ética em Pesquisa com Seres Humanos



FACULDADE DE MEDICINA DE
SÃO JOSÉ DO RIO PRETO-
FAMERP - SP



PARECER CONSUBSTANCIADO DO CEP

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: CARACTERIZAÇÃO DE MARCADORES STR DO CROMOSSOMO X NA POPULAÇÃO DO ESTADO DE MATO GROSSO, BRASIL

Pesquisador: Joyce Aparecida Martins Lopes Ferraz

Área Temática: Genética Humana:

(Trata-se de pesquisa envolvendo Genética Humana que não necessita de análise ética por parte da CONEP.);

Versão: 1

CAAE: 39088514.8.0000.5415

Instituição Proponente: Faculdade de Medicina de São José do Rio Preto- FAMERP - SP

Patrocinador Principal: Financiamento Próprio

DADOS DA NOTIFICAÇÃO

Tipo de Notificação: Outros

Detalhe: notificação

Justificativa: Encaminhamento de alteração no projeto de pesquisa.

Data do Envio: 30/05/2016

Situação da Notificação: Parecer Consubstanciado Emitido

DADOS DO PARECER

Número do Parecer: 1.566.458

Apresentação da Notificação:

Trata-se de uma notificação do estudo CAAE: 39088514.8.0000.5415.

Objetivo da Notificação:

Notificar alteração do estudo.

Avaliação dos Riscos e Benefícios:

-

Comentários e Considerações sobre a Notificação:

Foi encaminhada notificação em 30/05/16, solicitando alterações conforme consta:

Endereço: BRIGADEIRO FARIA LIMA, 5416

Bairro: VILA SAO PEDRO

CEP: 15.090-000

UF: SP

Município: SAO JOSE DO RIO PRETO

Telefone: (17)3201-5813

Fax: (17)3201-5813

E-mail: cepfamerp@famerp.br



FACULDADE DE MEDICINA DE
SÃO JOSÉ DO RIO PRETO-
FAMERP - SP



Continuação do Parecer: 1.586.458

Diante das dificuldades com a aquisição de auxílios financeiros para o desenvolvimento deste projeto de pesquisa, venho por meio desta notificação solicitar a alteração da análise de marcadores AS-STR para a análise de marcadores de inserção e deleção do cromossomo X (X-InDels), em virtude de haver reagentes remanescentes de outro projeto de pesquisa para esta última análise. A utilização dos X-InDels tem sido uma alternativa em exames de DNA com amostras degradadas, não havendo dados deste polimorfismo para a população do Estado de Mato Grosso. Assim, serão analisados 32 marcadores X-InDel em 500 indivíduos (100 mulheres e 400 homens) da população do referido Estado. Os procedimentos de coleta das amostras sanguíneas e extração do DNA para análise do sistema X-InDels serão os mesmos do projeto original submetido ao CEP. Após extração do DNA, os marcadores X-InDels serão amplificados conforme protocolo descrito por Pereira e colaboradores (2012) e, posteriormente, a eletroforese capilar e genotipagem serão efetuadas como descrito para o sistema X-STR decaplex no projeto original.

Serão calculados a frequência alélica, a heterozigosidade observada e esperada, o equilíbrio de Hardy-Weinberg, o desequilíbrio de ligação e a distância genética (FST) entre populações pelo uso do programa Arlequin v3.0. Parâmetros de interesse forense serão identificados através da análise do poder de discriminação, do poder de exclusão e da probabilidade de coincidência de perfis genéticos (matching 2 probabilities), calculados pelo programa PowerStats v1.2 (Promega Corporation).

Considerações sobre os Termos de apresentação obrigatória:

De acordo.

Recomendações:

Não há.

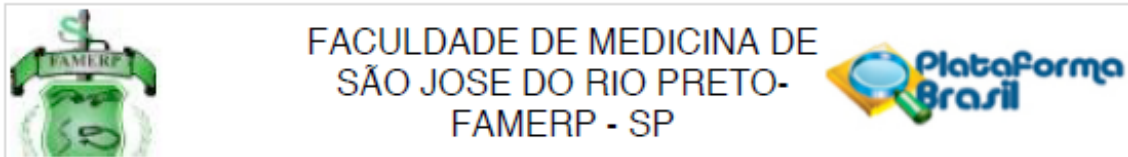
Conclusões ou Pendências e Lista de Inadequações:

Sem pendências ou inadequações.

Considerações Finais a critério do CEP:

O Comitê de Ética em Pesquisa em Seres Humanos da FAMERP aprova a alteração da análise de marcadores AS-STR para a análise de marcadores de inserção e deleção do cromossomo X (X-InDels), tamanho amostral e demais alterações contempladas nesta notificação; referente ao estudo CAAE: 39088514.8.0000.5415 sob a responsabilidade de Joyce Aparecida Martins Lopes Ferraz com o título: "CARACTERIZAÇÃO DE MARCADORES STR DO CROMOSSOMO X NA POPULAÇÃO DO ESTADO DE MATO GROSSO, BRASIL"

Endereço: BRIGADEIRO FARIA LIMA, 5416
Bairro: VILA SAO PEDRO CEP: 15.090-000
UF: SP Município: SAO JOSE DO RIO PRETO
Telefone: (17)3201-5813 Fax: (17)3201-5813 E-mail: cepfamerp@famerp.br



Continuação do Parecer: 1.566.458

Este parecer foi elaborado baseado nos documentos abaixo relacionados:

Tipo Documento	Arquivo	Postagem	Autor	Situação
Outros	NotificacaoCEP.pdf	30/05/2016 19:25:35	Joyce Aparecida Martins Lopes Ferraz	Aceito

Situação do Parecer:

Aprovado

Necessita Apreciação da CONEP:

Não

SAO JOSE DO RIO PRETO, 31 de Maio de 2016

Assinado por:
LUCIANO GARCIA LOURENCAO
(Coordenador)

Endereço: BRIGADEIRO FARIA LIMA, 5416
 Bairro: VILA SAO PEDRO CEP: 15.090-000
 UF: SP Município: SAO JOSE DO RIO PRETO
 Telefone: (17)3201-5813 Fax: (17)3201-5813 E-mail: cspfamerp@famerp.br

ANEXO II



Contents lists available at ScienceDirect

Forensic Science International: Genetics Supplement Series

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

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



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

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

ARTICLE INFO

Keywords:

STRs
D22S1045
D2S441
D10S1248
São Paulo State
Brazil

ABSTRACT

São Paulo state is one of the Brazilian states with the highest cultural and ethnic miscegenation, mainly of Italian and Portuguese descendants, but also strongly influenced by Amerindians, Africans, Spanish, Germans, Arabs and Japanese, as well as the high number of immigrants from the northeast region. In this work we analyzed statistical data of three STRs markers (D22S1045, D2S441 and D10S1248) that are used in our laboratory's forensic routine for paternity investigations. The use of these data will be extremely important for the laboratory to perform statistical analyzes for the conclusion of the reports. We analyzed 221 samples obtained from unrelated individuals born in the São Paulo state, which the allele frequencies and statistical parameters were estimated with PowerStats version 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-ISPG Working Group.

* Corresponding author.

Email address: isabrunelli@hotmail.com (I.B. Ambrosio).

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

Received 2 September 2017; Accepted 19 September 2017

Available online 23 September 2017

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

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

Allele	D2S1045	D2S441	D10S1248
8	0,0023		0,0023
9		0,0023	
10	0,0113	0,1742	0,0023
11	0,1176	0,3122	0,0113
11.3		0,0566	
12	0,0158	0,0814	0,0588
12.3		0,0023	
13	0,0068	0,0362	0,2579
14	0,0362	0,3009	0,2941
15	0,2986	0,0294	0,1968
16	0,3733	0,0045	0,1471
17	0,1267		0,0249
18	0,0113		0,0023
19			0,0023
N	442	442	442
HO	0,262	0,226	0,213
HE	0,738	0,774	0,787
PIC	0,700	0,735	0,749
PD	0,890	0,907	0,914
PE	0,489	0,551	0,576

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

Table 2
Comparison of the allele frequencies for D2S441, D10S1248 and D2S1045 loci between the São Paulo State population and other populations.

Microsatellite markers	Southern Brazil × São Paulo	Europe × São Paulo	África × São Paulo	Colombian × São Paulo
D2S441	0.19992	0.17631	0.17925	–
D10S1248	0.00982	0.13288	0.10796	0.12997
D2S1045	0.13122	0.12334	0.10917	0.15853

– no data available.

3. Results and discussion

The power of discrimination (PD) and power of exclusion (PE) for the marker D10S1248 were 0.914 and 0.576 respectively; the highest frequency allele was 14 (0.294) and the lowest frequency were 8, 10, 18 and 19 (0.002) (Table 1). PD and PE of the marker D2S441 were 0.907 and 0.551, respectively; the allele with the highest frequency was 11 (0.312) and the lowest frequency were 9 and 12.3 (0.002). The D2S1045 marker showed PD 0.890 and PE 0.489, highest frequency in 16 allele (0.373) and lowest allele 8 (0.002).

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

4. Conclusion

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

Conflict of interest statement

None.

Acknowledgements

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

References

- [1] V. Cortellini, N. Corri, A. Venzietti, Genetic variation at 5 new autosomal short tandem repeat markers (D10S1248, D22S1045, D2S441, D1S1656, D12S391) in a population-based sample from Maghreb region, *Popul. Genet.* 52 (2011) 368–371.
- [2] IBGE, Instituto Brasileiro de Geografia e Estatística, Censo Demográfico de 2010, IBGE, Rio de Janeiro, 2010.
- [3] M.G. Enzenberger, K.A. Lenz, L.K. Mathies, et al., Developmental validation of the PowerPlex® Fusion 6C System, *Forensic Sci. Int. Genet.* 21 (2016) 134–144.
- [4] PowerStats version 1.2, Promega Corporation. Available on: <http://www.promega.com/geneticidtools/powerstats/>.
- [5] C. Phillips, L. Fernandez-Formoso, M. Garcia-Magariños, et al., Analysis of global variability in 15 established and 5 new European Standard Set (ESS) STRs using the CEPH human genome diversity panel, *Forensic Sci. Int. Genet.* 5 (2011) 155–169.
- [6] G. Bugosa, L. Fernandez-Formoso, M. Garcia-Magariños, et al., Allelic frequencies and forensic parameters for miniSTRs D10S1248, D14S1434 and D22S1045 (NC01) in a sample from Central Andean Colombian region, *Forensic Sci. Int. Genet. Suppl. Ser.* 5 (2015) e81–e82.
- [7] R. Rodenbusch, S. Schumacher, A.C. Mardini, et al., Allele frequencies of the five new generation forensic STR (D1S1656, D2S441, D10S1248, D12S391 and D22S1045) in the population from Rio Grande do Sul, Southern Brazil, *Forensic Sci. Int. Genet.* 6 (2012) e55–e57.
- [8] I. Excoffier, H. Lischer, Arlequin: an integrated software package for population genetics data analysis (2015).



Contents lists available at ScienceDirect

Forensic Science International: Genetics Supplement Series

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

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

D.F. Braganholi^a, I.B. Ambrosio, F.S. Polverari, J. Martinez, R.M.B. Cicarelli^aLaboratory of Paternity Investigation, NAC/PCF – 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 *tupí* 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

^a Corresponding author.

E-mail address: daniobraganholi@hotmail.com (D.F. Braganholi).

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

Received 22 August 2017; Accepted 19 September 2017

Available online 21 September 2017

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

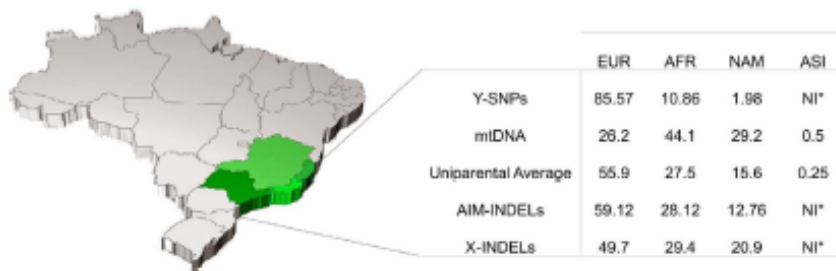


Fig. 1. In green is illustrated the location of the southeast region in the map of Brazil. The table shows the data (%) of the ancestry analysis obtained with the different markers. EUR: European; AFR: African; NAM: Native American; ASI: Asiatic *NI: Not identified. (For interpretation of the references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

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

2. Methodology

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

3. Results and discussion

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

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

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

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

All this historical context means that the Southeastern region, as

well as the whole of Brazil, presents a great genetic miscegenation as shown in Fig. 1.

4. Conclusion

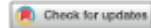
The genetic data observed are in agreement with the historical knowledge about the formation of the southeastern region of Brazil, where there is great miscegenation. The ancestry of European origin is predominant in the paternal pathway, whereas, Native American and African origins in the maternal pathway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

Conflict of interest statement

No one.


References

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



SHORT REPORT

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

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

CONTACT Juliana Martinez  jumirtz@hotmail.com  Laboratory of Paternity Investigation, São Paulo State University (UNESP), School of Pharmaceutical Sciences, São Paulo, Brazil

*These authors contributed equally to this work.

© 2017 Informa UK Limited, trading as Taylor & Francis Group

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

Locus	Genotype			Mutation
	Mother	Child	Father	
D13S317	11-14	12-14	11-14	?*
D13S317	9-11	11-12	11-13	?*
D13S317	11-13	10-13	11	Paternal
Penta E	7-15	15-16	7-17	Paternal
D18S51	15-21	16-20	16-20	Maternal
D18S51	14-19	15-19	14-16	Paternal
D18S51	14-21	21-28	15-27	Paternal
CSF1PO	11-12	12.3		Maternal
CSF1PO	10-12	10-11	10-12	?*
Penta D	9-12	11-12	10-13	Paternal
vWA	16	16-18	16-17	Paternal
vWA	15-17	17-18	16-19	Paternal
vWA	17	17-21	17-20	Paternal
vWA	14-15	14-16	17-19	Paternal
D21S11	24.3-30	29-30	28-30	Paternal
D7S820	10-11	11-13	12	Paternal
D5S818	12-13	13-14	11-13	?*
D5S818	11	12-14	12-14	Maternal
D8S1179	13-14	13-15	14-14	Paternal
D8S1179	13-14	13-15	13-14	?*
D12S391	16-18.3	18-18.3	19-20	Paternal
D12S391	17-23	17-21	18-22	Paternal
D12S391	19-21	19-21	20-23	Paternal
D19S433	14	14-14.2	13-13.2	Paternal
D19S433	12.2-16	15-16.2	15.2-16.2	Maternal
FGA	21-23	21-26	21-25	Paternal
FGA	23-30	20-31		Maternal
FGA	22-25	21-22	22	?*
FGA	20-21	20-28	23-27	Paternal
D3S1358	14-17	14-17	15-20	Paternal
D3S1358	15	15-18	15-19	?*
D151656	14-17.3	17-17.3	16-18	Paternal
D151656	17.3-18.3	12-17.3	11-17.3	Paternal
D151656	12-14	14-17	14-16	Paternal
D6S1043	12-22.3	13-22.3	12-20	Paternal

*Indicates that it is not possible to identify the origin of the mutation.

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

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

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

Results and discussion

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

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

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

Locus	Paternal			Maternal			Total		
	No. of meioses ^a	No. of mutations	Mutation rate	No. of meioses ^b	No. of mutations	Mutation rate	No. of meioses ^c	No. of mutations	Mutation rate
D13S317	650	1	1.5E-03	873	0	0.0E+00	1523	3	1.97E-03
D16S539	650	0	0.0E+00	873	0	0.0E+00	1523	0	0.0E+00
Penta E	650	1	1.5E-03	873	0	0.0E+00	1523	1	6.57E-04
D18S51	650	2	3.1E-03	873	1	1.1E-03	1523	3	1.97E-03
CSF1PO	650	0	0.0E+00	873	1	1.1E-03	1523	2	1.31E-03
TH01	650	0	0.0E+00	873	0	0.0E+00	1523	0	0.0E+00
Penta D	650	1	1.5E-03	873	0	0.0E+00	1523	1	6.57E-04
vWA	650	4	6.2E-03	873	0	0.0E+00	1523	4	2.63E-03
D21S11	650	1	1.5E-03	873	0	0.0E+00	1523	1	6.57E-04
TPOX	650	0	0.0E+00	873	0	0.0E+00	1523	0	0.0E+00
D7S820	650	1	1.5E-03	873	0	0.0E+00	1523	1	6.57E-04
D5S818	650	0	0.0E+00	873	1	1.1E-03	1523	2	1.31E-03
D8S1179	650	1	1.5E-03	873	0	0.0E+00	1523	2	1.31E-03
D12S391	580	3	5.2E-03	776	0	0.0E+00	1356	3	2.21E-03
D19S433	622	1	1.6E-03	837	1	1.2E-03	1459	2	1.37E-03
D251338	622	0	0.0E+00	837	0	0.0E+00	1459	0	0.0E+00
FGA	650	2	3.1E-03	873	1	1.1E-03	1523	4	2.63E-03
D3S1358	650	1	1.5E-03	873	0	0.0E+00	1523	2	1.31E-03
D151656	580	3	5.2E-03	776	0	0.0E+00	1356	3	2.21E-03
D6S1043	580	1	1.7E-03	776	0	0.0E+00	1356	1	7.37E-04

^aPaternal allelic transmission.

^bMaternal allelic transmission.

^cTotal paternal and maternal allelic transmission.

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

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

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

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

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

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

Conclusions

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

Acknowledgements

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


Disclosure statement

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

Funding

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

ORCID

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

References

- Ambrosio IB, Braganholi DF, Polverari FS, Cicarelli RMB. 2015. Mitochondrial DNA 30-SNP data confirm high prevalence of African lineages in the population of Espírito Santo, Brazil. *Forensic Sci Int* 5:e346–e347.
- Andrade ES, Gomes AV, Raposo G, Mauricio-da-Silva L, Silva Rdos S. 2009. Mutation rates at 14 STR loci in the population from Pernambuco, Northeast Brazil. *Forensic Sci Int Genet* 3:e141–e143.
- Brinkmann B, Klitsch M, Neuhuber F, Hühne J, Rolf B. 1998. Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. *Am J Hum Genet* 62:1408–1415.
- Butler JM, Reeder DJ. 2017. STR DNA Internet database. Available online at: <http://www.cstlnist.gov/strbase/> (accessed April 2017).
- Chula FG, Rodenbusch R, Schumacher S, Grandi T, Michelon CT, Gastaldo AZ, Costi C, et al. 2009. 15 STR loci frequencies with mutation rates in the population from Rio Grande do Sul, Southern Brazil. *Forensic Sci Int Genet* 3:e35–e38.
- Egeland T, Mostad P, Bente M, Margrethe S. 2000. Beyond traditional paternity and identification cases. Selecting the most probable pedigree. *Forensic Sci Int Genet* 110:47–59.
- Fan H, Chu JY. 2007. A brief review of short tandem repeat mutation. *Genomics Proteomics Bioinformatics* 5:7–14.
- Figueiredo RF, Ambrosio IB, Braganholi DF, Chemale G, Martins JA, Gomes V, Gusmão L, Cicarelli RMB. 2015. Male-specific contributions to the Brazilian population of Espírito Santo. *Int J Legal Med* 130:679.
- Francioli LC, Polak PP, Koren A, Menelaou A, Chun S, Renkens I, Genome of the Netherlands Consortium, et al. 2015. Genome-wide patterns and properties of de novo mutations in humans. *Nature Genetics* 47:822–826.
- Ge J, Eisenberg A, Budowle B. 2012. Developing criteria and data to determine best options for expanding the core CODIS loci. *Investig Genet* 3:1.
- Geada H, Viriato L, Vieira-Silva C, Cruz C, Lucas J, Ribeiro T, Espinheira R. 2003. STR mutations in paternity investigations: a study of 1-year consecutive cases. *Int Congr Ser* 1239:657–660.
- Gjertson DW, Biener CH, Baur MP, Carracedo A, Guidet F, Luque JA, Lessig R, et al. 2007. ISFG: recommendations on biostatistics in paternity testing. *Forensic Sci Int Genet* 1:223–231.
- Henke L, Henke J. 2006. Supplemented data on mutation rates in 33 autosomal short tandem repeat polymorphisms. *J Forensic Sci* 51:446–447.
- Hohoff C, Schürenkamp M, Brinkmann B. 2009. Meiosis study in a population sample from Nigeria: allele frequencies and mutation rates of 16 STR loci. *Int J Legal Med* 123:259–261.
- Keightley P. 2007. Rates and fitness consequences of new mutations in humans. *Genetics* 190:295–304.
- Kong A, Frigge ML, Masson G, Besenbacher S, Sulem P, Magnusson G, Gudjonsson SA, et al. 2012. Rate of de novo mutations, father's age, and disease risk. *Nature* 488:471–475.
- Manta FSN, Pereira R, Vianna R, Rodolfo Beutenmüller de Araújo A, Leite Góes Gitaí D, Aparecida da Silva D, de Vargas Wolgramm E, et al. 2013. Revisiting the genetic ancestry of Brazilians using autosomal AIM-Indels. *PLoS One* 8:e75145.
- Mardini AC, Rodenbusch R, Schumacher S, Chula FG, Michelon CT, Gastaldo AZ, Maciel LP, et al. 2013. Mutation rate estimates for 13 STR

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



Contents lists available at ScienceDirect

Forensic Science International: Genetics Supplement Series

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

Insertion of amelogenin marker in the 38 Hid-Indels multiplex reaction



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

^a Laboratory of Paternity Investigation, NAQ/FCF–UNESP, Araraquara, SP, Brazil

ARTICLE INFO

Article history

Received 4 August 2015

Accepted 16 September 2015

Available online 25 September 2015

Keywords

Amelogenin marker

INDELS

Degraded DNA

ABSTRACT

Pereira et al. [3] reported a method for multiplex analysis of 38 INDELS autosomes known polymorphic in different populations of the world which can be used in the human identification, but this methodology not originally introduced marker for the analysis of amelogenin. In laboratory of paternity of UNESP/Araraquara, Brazil, kinship analysis are performed using samples with degraded DNA, where the amplification of 38HID-INDELS was efficient compared to STRs in these samples. For paternity tests also analyze STRs, and the verification of the sex of samples is done by the amelogenin marker analysis present in commercial kits, but the bad amplification of STRs may interfere in the correct reading of amelogenin in the electropherogram. Trying to solve this problem we decided to add the amelogenin marker in multiplex reaction with 38HID-INDELS.

© 2015 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Ancient DNA may be to present difficulties or fails in STR amplification related to preservation and contamination, restricting the correct typing of the sample [1]. INDEL markers can overcome this problem because presents small fragment to be amplified in the PCR (about 200bp) [2]. Therefore, we use the analysis method for 38 autosomal INDELS [3] and insert the amelogenin marker in the multiplex reaction, to complement the STR results when necessary.

2. Objectives

Insert the amelogenin marker in the multiplex reaction with 38HID-INDELS for use in the analysis of the degraded samples.

3. Methodology

Through literature search and *in silico* analysis we select the pair of primers described by Krenke et al. [4], and we made a small alteration in the sequence of the reverse primer, for the amplification of a fragment with 6pb additional for correct

reading of the amelogenin marker and 38HID-INDELS in the electropherogram. The sequences of the primers are (5'–3'): "NED-CCCTGGGCTCTGTAAAGAA" (forward) and "CCAACCATCAGAGCT-TAAACTGGG" (reverse) and were used in the final concentration of 0.2 μM in the PCR, with reagents and cycling by the protocol [3].

4. Results and discussion

Until this moment, the multiplex reaction 38HID-INDELS with amelogenin marker (Fig. 1) was used in the analysis of fifteen cases of biological kinship with degraded DNA samples helping to solve them.

5. Conclusion

The alteration of the reverse primer allowed the multiplex analysis of the amelogenin marker with 38HID-INDELS without interaction between the primers.

Conflict of interest

None.

* Corresponding author. Fax: +55 16 33016940.

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

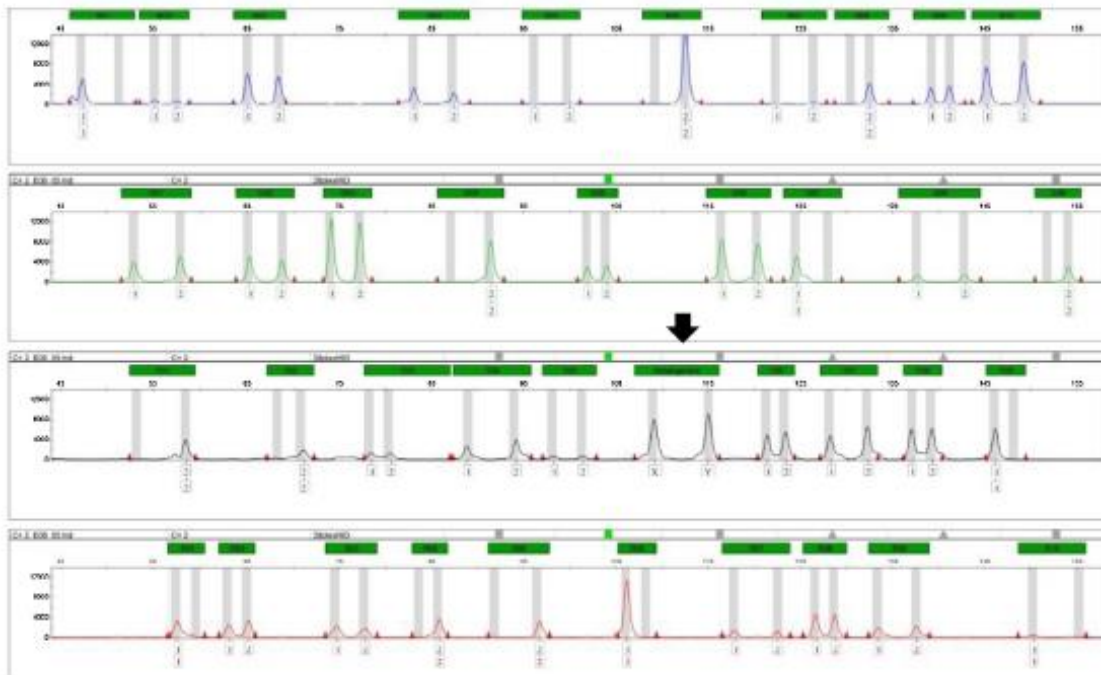


Fig. 1. Electropherogram of 38 HID-INDELs with amelogenin marker (indicated by arrow).

Acknowledgments

The authors would like to thank Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) by financial support and Rui Pereira by scientific support.

References

- [1] C. Romanini, M. Romero, M.S. Puerto, L. Catelli, C. Phillips, R. Pereira, L. Gusmão, C. Vullo, Ancestry informative markers: inference of ancestry in aged bone samples using an autosomal AIM-Indel multiplex, *Forensic Sci. Int. Genet.* 16 (2015) 58–63.
- [2] M. Pondevila, C. Phillips, N. Naverán, L. Fernandez, M. Cerezo, A. Salas, A. Carracedo, M.V. Lareu, Case report: identification of skeletal remains using short-amplicon marker analysis of severely degraded DNA extracted from a decomposed and charred femur, *Forensic Sci. Int. Genet.* 2 (2008) 212–218.
- [3] R. Pereira, C. Phillips, C. Alves, A. Amorim, A. Carracedo, L. Gusmão, A new multiplex for human identification using insertion/deletion polymorphisms, *Electrophoresis* 30 (2009) 3682–3690.
- [4] B.E. Krenke, A. Tereba, S.J. Anderson, E. Buel, S. Culhane, C.J. Finis, C.S. Tomsey, J. M. Zachetti, A. Masihay, D.R. Rabbach, E.A. Amiot, C.J. Sprecher, Validation of a 16-locus fluorescent multiplex system, *J. Forensic Sci.* 47 (4) (2002) 773–785.



Contents lists available at ScienceDirect

Forensic Science International: Genetics Supplement Series

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

Mitochondrial DNA 30-SNP data confirm high prevalence of African lineages in the population of Espírito Santo, Brazil



Isabela Brunelli Ambrosio, Danilo Faustino Braganholi, Fernanda Silva Polverari, Regina Maria Barretto Cicarelli*

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

ARTICLE INFO

Article history:

Received 22 August 2015

Accepted 14 September 2015

Available online 25 September 2015

Keywords:

DNA mitochondrial

SNPs

SNaPshot

Haplogroups

Espírito Santo-Brazil

ABSTRACT

Mitochondrial DNA (mtDNA) analysis is usually a last resort in routine forensic DNA casework especially when nuclear DNA analysis cannot be applied. Because of the Brazilian population is highly admixed it is necessary to select highly discriminatory SNPs. The goal of this article is to characterize the mtDNA haplogroups in 100 unrelated individuals from the state of Espírito Santo, Brazil, by using 30-SNPs of mtDNA in two multiplex SNaPshot. The population was classified according to its origin in 43% of African, 26% of Native American, 30% of European and 1% Asian in 19 different haplogroups. These data confirmed therefore that this population is also highly admixed because of the history of country colonization.

© 2015 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Mitochondrial DNA (mtDNA) sequence analysis has become a powerful tool in the field of forensic sciences for missing persons, victims of mass disasters comparing the mother's genotype or a relative who has maternal relationship with the individual to be identified [1,2]. The Brazilian population is highly admixed, resulting mostly from interethnic crosses between Native Americans, Europeans and Africans [2]. Classification in haplogroups has been an important tool used in population genetics in forensic cases and genetic studies of disease [3]. Paneto et al. [4] analysed 42-SNPs (single nucleotide polymorphisms) of mtDNA to classify individuals in the population of state of São Paulo by SNaPshot. In this article we used 30-SNPs of mtDNA in two multiplex SNaPshot to characterize the haplogroups in individuals from the state of Espírito Santo, Brazil, a population which has no data about it.

2. Material and methods

Blood samples from 100 unrelated individuals who are residents in the state of Espírito Santo were used. All the

participants gave their written informed consent. 30 of the 42-SNPs described in Paneto et al. [4] were used in this study (Fig. 1). We used blood spotted on filter paper and DNA was extracted from two discs of 1.2 mm using Chelex (BioRad). PCR and SNaPshot reactions were made in two separate reactions, Multiplex I (21 SNPs) and Multiplex II (9 SNPs). All steps were performed according to the protocol described by Paneto et al. [8]. The capillary electrophoresis was performed on ABI 3500 Genetic Analyzer (Applied Biosystems by Life Technologies). The samples were classified into haplogroups using the phylogenetic tree described by Paneto et al. [4], and confirmed by van Oven and Kayser [5].

3. Results and discussion

It was found 19 different haplogroups (Fig. 2) classified according to the origin of the population: 43% of Africa, 26% of Native American, 30% of European and 1% Asian. Our results are in agreement with the history of Brazil colonization: it was populated by Native Americans; after discovery (1500), European immigration, mainly Portuguese, who arrived between 1500 and 1808 followed by Africans reflecting the period of slavery. Asian immigration also occurred, but to a lesser extent [2]. These data confirm, therefore, that the Brazilian population is highly racially mixed, result of ethnic crossings.

* Corresponding author. Fax: +55 16 3322 0073.

E-mail address: paternidade.fcfa@gmail.com (R.M.B. Cicarelli).

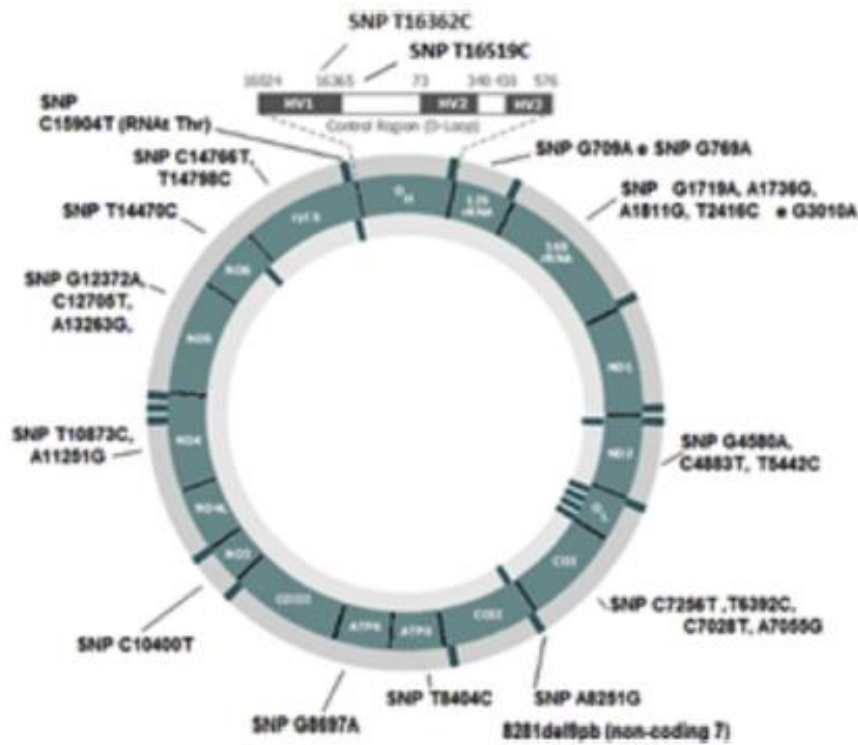


Fig. 1. General location of the mtDNA 30-SNPs.

Frequency of Haplogroups

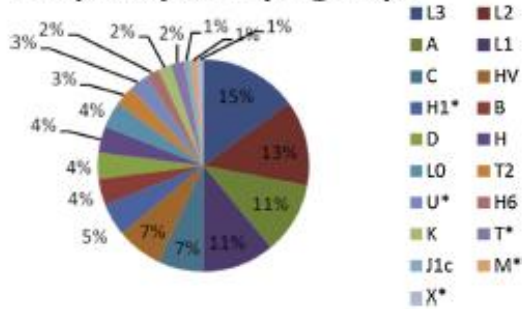


Fig. 2. Frequency of haplogroups in the population of Espírito Santo, Brazil.

References

- [1] G. Chemale, G.G. Paneto, M.A.M. Menezes, et al., Development and validation of a D-loop mtDNA SNP assay for the screening of specimens in forensic casework, *Forensic Sci. Int.: Genet.* 7 (2013) 353–358.
- [2] J. Alves-Silva, M.S. Santos, P.E.M. Guimarães, et al., The ancestry of Brazilian mtDNA lineages, *Am. J. Hum. Genet.* 67 (2000) 444–461.
- [3] M.D. Coble, R.S. Just, J.E. O’Callaghan, et al., Single nucleotide polymorphisms over the entire mtDNA genome that increase the power of forensic testing in Caucasians, *Int. J. Legal Med.* 118 (2004) 137–146.
- [4] G.G. Paneto, S. Köhnemann, J.A. Martins, et al., A single multiplex PCR and SNaPshot minisequencing reaction of 42 SNPs to classify admixture populations into mitochondrial DNA haplogroups, *Mitochondrion* 11 (2011) 296–302.
- [5] M. van Oven, M. Kayser, Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation, *Hum. Mutat.* 18 (2004) 14–18.

Acknowledgments

This study was supported by PADCF/FCF-UNESP, FAPESP and CAPES.