

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

BIANCA BELON JANUARIO

**Avaliação de diferentes conjuntos de marcadores do
cromossomo X na população residente do estado de
São Paulo**

Araraquara
2021

BIANCA BELON JANUARIO

Avaliação de diferentes conjuntos de marcadores do cromossomo X na população residente do estado de São Paulo

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
2021

FICHA CATALOGRÁFICA

J33a Januario, Bianca Belon
Avaliação de diferentes conjuntos de marcadores do cromossomo X na população residente do estado de São Paulo / Bianca Belon Januario. – Araraquara : [s.n.], 2021
122 f. : il.

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

1. Cromossomos sexuais. 2. Genética forense.
3. Marcadores genéticos. 4. Polimorfismo (Genética). 5. DNA.
I. Título.


CERTIFICADO DE APROVAÇÃO


TÍTULO DA TESE: Avaliação de diferentes conjuntos de marcadores do cromossomo X na população residente do estado de São Paulo


AUTORA: BIANCA BELON JANUARIO


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 (Participação Virtual)
Departamento de Ciências Biológicas / Faculdade de Ciências Farmacêuticas - UNESP - Araraquara


Dr. JORGE MARCELO DE FREITAS (Participação Virtual)
Instituto Nacional de Criminalística / Diretoria Técnico-Científica - Polícia Federal (INC/DITEC/PF) – Brasília-DF


Prof. Dr. MARCELO MALAGHINI (Participação Virtual)
Laboratório de Genética Molecular Forense do Instituto de Criminalística LGMF/IC – Curitiba- PR


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


Dr. CAIO CESAR SILVA DE CERQUEIRA (Participação Virtual)
Equipe de Perícias Criminalísticas - Ourinhos / Superintendência da Polícia Técnico-Científica do estado de São Paulo

Araraquara, 24 de novembro de 2021

DEDICATÓRIA

À minha família pelo apoio sempre.

*À minha amada mãe Sonia 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
gradidão 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.

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.

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

À Flávia Alves, 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 Isabela Brunelli, pela ajuda no desenvolvimento do trabalho, pelas conversas.

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

Aos amigos do LIQ (Laboratório de Investigação de Paternidade): 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, pelos serviços prestados e pela atenção.

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

“Eu não estou interessado em nenhuma teoria, em nenhuma fantasia, nem no algo mais. A minha alucinação é suportar o dia-a-dia e meu delírio é a experiência com coisas reais”.

Belchior

Resumo

Nas últimas décadas com o avanço da biotecnologia foi possível analisar os polimorfismos genéticos, os quais acabaram tornando-se uma importante ferramenta na identificação de pessoas. Na identificação humana as análises são baseadas no perfil genético de um indivíduo, por meio da utilização de marcadores como os STRs autossômicos, que são os mais utilizados em identificação humana como em análises de paternidade. Além dos STRs autossômicos, os STRs dos cromossomos sexuais podem ser utilizados para auxiliar na identificação quando há amostras degradadas ou em casos complexos de análise de parentesco. Outros marcadores são utilizados em identificação humana como os polimorfismos de inserção e deleção (InDels) e os polimorfismos de base única SNPs (*single nucleotide polymorphisms*). O foco desta pesquisa são os marcadores do cromossomo X, pois possuem características únicas, como seu padrão de herança e recombinação. Estes aspectos peculiares fazem com que sua análise seja importante na área forense. Neste trabalho foram analisadas amostras da população residente no estado de São Paulo, sendo 141 trios (pai, mãe e filha) dos quais foram avaliados os marcadores X-STRs presentes nos conjuntos de marcadores do Decaplex (10 X-STRs - Gusmão, 2009) e no kit comercial Investigator® Argus X-12QS (Qiagen). Os resultados mostraram que ambos os conjuntos são polimórficos para a população estudada, alcançando valores significativos de diversidade gênica: no Decaplex os marcadores GATA172D05 (0,83118) e DXS6809 (0,82181) apresentaram-se com valores mais altos de diversidade. Já nos 12 marcadores do Argus destacando o grupo de ligação 1 (LG1) obtendo o maior valor de diversidade haplotípica, 0,99835. Na análise do equilíbrio de Hardy-Weinberg nenhum marcador dos dois conjuntos analisados apresentou desvio significativo, nem estratificação populacional. No que se refere ao desequilíbrio de ligação houve associação entre marcadores de grupos de ligação diferentes na análise dos 12-XSTRs do kit Argus, DXS10134 – DXS10135. Entretanto, no conjunto de marcadores de 10 X-STRs – Decaplex não foi observado desequilíbrio de ligação. Os dois conjuntos de marcadores quando utilizados para análise de parentesco mostraram-se eficiente com valores expressivos de LR (*likelihood ratio*) principalmente na análise de vínculo biológico entre supostas avó-neta. Os valores obtidos de LR com o Decaplex foi de 3,2160E+09 e com o Argus de 7,3840E+03. Portanto, os resultados obtidos neste

trabalho levam a conclusão que os marcadores X-STRs testados são informativos e interessantes para análises de populações humanas e aplicáveis na genética forense.

Palavras-chave: Cromossomo X; Genética Forense; Identificação Humana; São Paulo; X-STRs.

ABSTRACT

In recent decades, with biotechnology advances it was possible to analyze genetic polymorphisms, which ended up becoming an important tool for identifying people. In human identification (HID) analyzes are based on individual genetic profile using markers such as autosomal STRs, the most used in human identification and paternity analyses. In addition to autosomal STRs analyses, sex chromosome STRs can be used to help identification with degraded samples or in complex cases of kinship analysis. Other markers are used in human identification such as insertion and deletion polymorphisms (InDels) and single nucleotide polymorphisms (SNPs). This research focuses in X chromosome markers because they have unique characteristics such as their pattern of inheritance and recombination. These peculiar aspects make analysis with these markers interesting in forensic area. In this work, samples from population residing in São Paulo state were analyzed, X-STRs markers present in the Decaplex marker sets (Gusmão, 2009) and in the Investigator® Argus commercial kit X-12QS (Qiagen) were used to genotyping 141 trios (father, mother and daughter). Results showed that both are polymorphic for the studied population reaching gene diversity significant values, Decaplex best markers were GATA172D05 (0.83118) and DXS6809 (0.82181) reaching higher values. Already in haplotypic diversity the 12 Argus markers highlighting the linkage group 1 (LG1) obtaining the highest value, 0.99835. In Hardy-Weinberg equilibrium analysis, none markers of the two analyzed sets presented significant deviation, nor population stratification. Regarding linkage disequilibrium, there was an association between markers from different linkage groups in 12-XSTRs of argus kit analysis, DXS10134 – DXS10135. However, in 10 X-STRs markers – Decaplex was not observed linkage disequilibrium. The two sets of markers when used for kinship analysis are efficient based on expressive values of LR (likelihood ratio) mainly in kinship analysis between supposed grandmother-grandchildren. The LR values obtained with the Decaplex were 3.2160E+09 and with Argus were 7.3840E+03. Therefore, the results obtained in this work lead to the conclusion that the tested X-STRs markers are informative and interesting for the analysis of human populations and applicable in forensic genetics.

Keywords: Forensic Genetics; Human Identification; São Paulo; X chromosome; X-STRs.

LISTA DE FIGURAS

Figura 1.	Fotografia da difração de raio-X de Rosalind Franklin.....	19
Figura 2.	James Watson e Francis Crick com a estrutura de DNA.....	20
Figura 3.	Dr. Thomas Brock, em <i>Black Sand Pool</i> , Parque Nacional de <i>Yellowstone</i> em setembro de 1995.....	21
Figura 4.	a) Kary Mullis e b) PCR – reação em cadeia de polimerase, aumento exponencial do DNA.....	22
Figura 5.	DNA <i>fingerprinting</i> de Alec Jeffreys.....	23
Figura 6.	DNA <i>fingerprinting</i> da família de Gana envolvida na imigração.....	25
Figura 7.	Polimorfismo de sequência - SNP (<i>single nucleotide polymorphism</i>).....	29
Figura 8.	Polimorfismo de comprimento – STRs (<i>short tandem repeats</i>).....	29
Figura 9.	Representação esquemática dos cromossomos sexuais X e Y.....	35
Figura 10.	Padrão de herança do cromossomo X.....	39
Figura 11.	Ideograma do cromossomo X com a localização dos marcadores X-STRs.	42
Figura 12.	<i>Pedigrees</i> analisados das supostas meias – irmãs (a) e supostas avó – neta (b).....	59
Figura 13.	Eletroferograma do perfil alélico do DNA comercial 9947A obtido com o kit Investigator® Argus X-12QS(Qiagen), seguindo o protocolo original.....	62
Figura 14.	Eletroferograma do perfil alélico do DNA comercial 9947A obtido com o kit Investigator® Argus X-12QS(Qiagen), com o protocolo adaptado.....	63
Figura 15.	Eletroferograma do perfil alélico de amostra masculina obtido com painel de marcadores Decaplex.....	64
Figura 16.	Eletroferograma do perfil alélico de uma mostra feminina com os marcadores do kit Investigator® Argus X 12 QS.....	65
Figura 17.	Valores de p obtidos no pareamento do teste de desequilíbrio de ligação para todos os pares de <i>loci</i> dos 10 X-STRs (Decaplex).....	76
Figura 18.	Valores de p obtidos no pareamento do teste de desequilíbrio de ligação para todos os pares de <i>loci</i> 12 X-STRs (Argus).....	77
Figura 19.	<i>Pedigree</i> analisado de “supostas” avó (1) – neta (2) com os dois conjuntos de marcadores e os STRs autossômicos.....	78
Figura 20.	<i>Pedigree</i> analisado de “supostas” meias – irmãs (3 e 4) e respectivas mães (1 e 2) com os dois conjuntos de marcadores e os STRs autossômicos.....	79

LISTA DE TABELAS

Tabela 1.	X-STRs utilizados em estudos nos últimos anos.....	43
Tabela 2.	Conjunto de marcadores com 10 X-STRs – decaplex.....	55
Tabela 3.	Conjunto de marcadores com 12 X-STRs do kit Investigator® Argus X-12 QS (Qiagen).....	55
Tabela 4.	Componentes e volume para a realização da PCR.....	61
Tabela 5.	Frequência alélica dos 10 X-STRS na população do estado de São Paulo.....	66
Tabela 6.	Comparação entre maiores diversidades gênicas deste e outros trabalhos com os 10 X-STRs do decaplex.....	67
Tabela 7.	Diversidade gênica (DG) dos marcadores e variabilidade alélica de cada grupo de ligação baseada na frequência dos 12-XSTRs do Kit Argus na população do estado de São Paulo.....	68
Tabela 8.	Comparação da diversidade gênica deste trabalho com outras pesquisas com os 12 X-STRs do Argus.....	69
Tabela 9.	Distribuição dos haplótipos dos 12 X-STRs para os grupos de ligação em todas as amostras.....	69
Tabela 10.	Diversidade gênica e variabilidade alélica dos 2 conjuntos de marcadores 10 X-STRS – Decaplex e kit Investigator® Argus com 12 X-STRs.....	70
Tabela 11.	Parâmetros estatísticos da diversidade genética para os conjuntos de 10 marcadores X-STRs e 12 X-STRs na população feminina de São Paulo. Tamanho da amostra: 141.....	72
Tabela 12.	Parâmetros estatísticos forense para os 10 marcadores X-STRs (Decaplex) na população de São Paulo. Número de cromossomos analisados: 282.....	73
Tabela 13.	Parâmetros estatísticos forense para os 12 marcadores X-STRs (Kit Argus) na população de São Paulo. Número de cromossomos analisados: 282.....	73

LISTA DE ABREVIATURAS E SIGLAS

CODIS	<i>Combined DNA Index System</i>
DNA	Ácido desoxirribonucleico
FBI	<i>Federal Bureau of Investigation</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
LR	<i>Likelihood ratio</i> , razão de verossimilhança
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
MSY	<i>Male-specific region of the Y</i> , região específica masculina do Y
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
RFU	Unidade de fluorescência relativa
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.....	14
2. Revisão da Literatura.....	17
2.1 – Breve histórico da biotecnologia do DNA a partir de década de 50.....	18
2.2 – Identificação humana e primeiros casos forenses	24
2.3 – Polimorfismos genéticos e marcadores biológicos para identificação humana.....	26
2.3.1 – SNP (<i>Single Nucleotide Polymorphisms</i>).....	28
2.3.2 – STRs (<i>Short tandem repeats</i>).....	29
2.3.3 – InDels (<i>insertion/ deletion polymorphisms</i>).....	32
2.3.4 – DNA mitocondrial.....	33
2.3.5 – Cromossomos sexuais.....	34
2.3.6 – Cromossomo Y.....	36
2.4 – Cromossomo X.....	37
2.4.1 – Características do cromossomo X.....	37
2.4.2 – Marcadores genéticos do cromossomo X.....	40
2.4.3 – Ligação e desequilíbrio de ligação para o cromossomo X.....	44
2.4.4 – Aplicação dos marcadores genéticos do cromossomo X.....	45
2.4.5 – Ferramenta para análise do cromossomo X: Famlink X.....	48
2.5 – Justificativa.....	49
3. Objetivos.....	51
3.1 – Objetivos gerais.....	52
3.2 – Objetivos específicos.....	52
4. Materiais e Métodos.....	53
4.1 - Aspectos éticos.....	54
4.2 - Casuística.....	54
4.3 – Coleta e extração de DNA.....	54
4.4 – Reação de amplificação dos X-STRs e preparação para eletroforese capilar	55
4.5 – Análise dos produtos amplificados.....	57
4.6 – Análise de dados.....	57
4.7 – Determinação dos perfis alélicos e análise estatística.....	58
4.8 – Análise de parentesco com o famlink X.....	58
5. Resultados e Discussão.....	60
5.1 – Padronização da utilização do kit Argus X-12 QS.....	61

5.2 - Avaliação da metodologia.....	64
5.3 – Análise genética da população.....	65
5.3.1 - Variabilidade genética na população de São Paulo com os conjuntos de marcadores.....	65
5.3.2 - Análise do conjunto de marcadores com 10 X-STRS – Decaplex.....	65
5.3.3 - Análise do conjunto de marcadores com 12 X-STRS do kit Argus X-12 QS.....	68
5.3.4 - Comparação dos dois conjuntos de marcadores 10 X-STRs-Decaplex e Kit Investigator® Argus X-12 QS.....	70
5.4 - Parâmetros estatísticos de avaliação da eficiência forense.....	71
5.4.1 – Análise de equilíbrio de Hardy-Weinberg.....	71
5.4.2 – Desequilíbrio de ligação entre os marcadores X-STRs.....	74
5.5 – Aplicação de X-STRs na análise de parentesco utilizando o software Famlink X.....	78
6. Conclusões.....	80
7. Referências bibliográficas.....	82
Apêndices.....	93
Apêndice A - Frequência haplotípica dos 4 grupos de ligação (LG) dos 12 X-STRs do kit argus.....	95
Apêndice B – Relatório do Famlink X com os perfis dos dois casos genotipados e resultado do Famlink X.....	102
Anexos.....	108
Anexo A - Aprovação do Comitê de Ética em Pesquisa com Seres Humanos.....	111
Anexo B - Trabalhos publicados.....	114

1. INTRODUÇÃO

A identificação humana a partir de técnicas de biologia molecular é um avanço relativamente recente. Esses avanços ocorreram a partir do conhecimento obtido no campo da genética que experimentou mudanças biotecnológicas principalmente a partir da década de 50, no século passado, obtidos com o avanço nos estudos sobre o DNA, gerando uma nova compreensão da variabilidade no nível molecular que permitiu a exploração de novos paradigmas biológicos, trazendo a análise de polimorfismos, que são de amplo interesse para geneticistas populacionais, investigadores forenses e na genética médica (ELHAIK et al., 2017).

A partir do conhecimento sobre a biotecnologia do DNA, foi possível investigar a identificação de um indivíduo e chegasse, portanto, aos conhecimentos atuais da identificação humana.

A identificação de pessoas pode ser estudada a partir de perspectivas de dinâmica entre as populações a contar dos primórdios das civilizações em eventos como migrações, a partir de acontecimentos que demandaram a identificação de pessoas como acontece em crimes, quando ocorre intempéries como desastres naturais, terremotos, tsunamis e erupções vulcânicas, além de acidentes causados por falha humana como acidentes de carros, aviões, entre outras tragédias que requerem identificação de pessoas, além de investigações de parentesco. Todas estas ocorrências muitas vezes demandam a identificação de indivíduos, através de suas amostras biológicas além do reconhecimento por documentos, joias, e outros pertences. (ZIĘTKIEWICZ et al., 2012; ELHAIK et al., 2017).

Partindo da história mais recente de identificação humana em 1985, foi publicado o primeiro artigo descrevendo a identificação de características moleculares presentes no DNA que diferenciam os indivíduos conhecido como DNA *fingerprinting* ou impressão digital do DNA (JEFFREYS et al., 1985a), a partir deste fato teve início a identificação de indivíduos para a área de genética forense.

Outro marco que definiu a biologia molecular que possuímos atualmente ocorreu com a introdução de métodos mais específicos visando a utilização de quantidades menores de DNA, principalmente com o surgimento das novas técnicas de análises como a PCR (*Polymerase Chain Reaction*) que permite a multiplicação exponencial de DNA (MULLIS et al., 1987), essas metodologias continuam sendo utilizadas e são de suma importância para o desenvolvimento biotecnológico, além da área de ciências forenses.

Desde então, a reação em cadeia da polimerase (*PCR*) permitiu o direcionamento de uma variedade de marcadores genéticos, desde polimorfismos de nucleotídeo único (*SNP*) a número variável de repetições em tandem (*VNTRs*). Os *VNTRs* incluem minissatélites (repetições de 6–100 pb) e microsatélites (repetições de 2-6 pb), frequentemente referidos como short tandem repeats (*STR*) ou repetições de sequência simples (*SSR*). A metodologia atual em identificação genética depende do uso de análise direta da sequência de DNA e da disponibilidade de numerosos bancos de dados genéticos. (BUTLER, 2011; ZIĘTKIEWICZ et al., 2012).

Muito estudo é necessário para o entendimento dos marcadores no DNA humano, sendo que, a quantidade de dados gerados nos estudos é fundamental para isso. Assim, é imprescindível aumentar o número de marcadores, e realizar estudos em diferentes populações observando o potencial forense, assim, informações estatísticas são geradas, obtendo portando maior conhecimento da história de uma população possibilitando a identificação de indivíduos (SCHAFFNER, 2004).

2. REVISÃO DA LITERATURA

2.1 Breve histórico da biotecnologia do DNA a partir da década de 50

O termo biotecnologia foi usado pela primeira vez por Karl Erkey, um engenheiro húngaro, em 1919 (VERMA et al. 2011).

A biotecnologia é um campo multidisciplinar que tem grande impacto em nossas vidas. Esse campo multidisciplinar tem uma ampla gama de uso impactando áreas como de saúde e bem-estar, meio ambiente, em processos de biorremediação que utiliza organismos naturais ou recombinantes para a limpeza de poluentes tóxicos, no desenvolvimento de fontes de energia alternativas, renováveis e sustentáveis para a produção de biocombustíveis, envolve também a biotecnologia marinha que explora os produtos que podem ser obtidos de organismos aquáticos, entre várias outras áreas. (GUPTA et al. 2017).

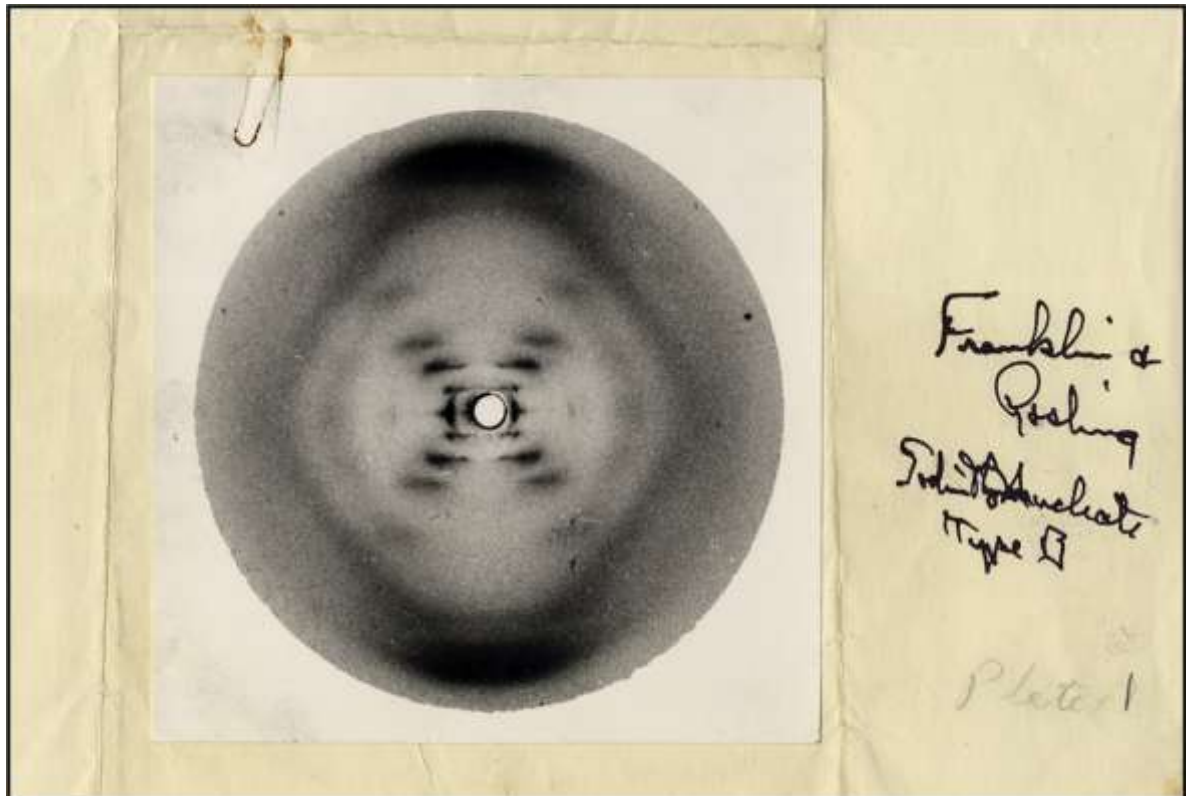
A biotecnologia chamada de moderna teve início depois da segunda guerra mundial, quando ocorreram várias descobertas importantes no campo científico, principalmente relacionadas ao DNA.

Na identificação pelo DNA, utilizamos elementos biotecnológicos, os quais foram desenvolvidos ao longo do tempo. A base da biotecnologia foi lançada após a descoberta da estrutura do DNA o material hereditário, chamado de ácido desoxirribonucléico (DNA), que contém todas as informações que ditam cada etapa da vida de um indivíduo (GUPTA et al., 2017).

A história do DNA e tudo a ele relacionado para chegarmos na identificação humana, começa muito antes da década de 50, porém foi a partir desse período que começou o que se pode chamar de revolução biotecnológica, onde vários pesquisadores deixaram sua contribuição para a ciência começando pela difração por raios X de Rosalind Franklin que forneceu os principais dados para derivar a estrutura de dupla hélice do DNA (VERMA et al.,2011).

Em novembro de 1951, Rosalind Franklin quando apresentou sua última pesquisa no King's College London, seus resultados logo mudariam o curso da ciência. Em seu estudo, Franklin apresentou fotografias de difração de raios-X onde foi possível observar que o DNA tem uma estrutura helicoidal. Para obter essas imagens, Franklin juntamente com seu aluno de graduação Raymond Gosling umedeceram as fibras de DNA e as expuseram a um feixe de raios-X. O padrão em forma de cruz resultante era certamente enigmático para a maioria dos ouvintes, mas Franklin foi capaz de interpretá-lo (CRAMER, 2020).

Figura 1 - Fotografia da difração de raio-X de Rosalind Franklin.



Fonte: CRAMER, 2020.

Na imagem (figura 1) de difração de raios X que foi gravada em filme fotográfico por Franklin e Gosling (1953) mostra um padrão específico em forma de cruz de reflexos de raios-X no centro que revelou a geometria geral da dupla hélice do DNA.

Logo após a descoberta realizada por Franklin e Gosling, James Watson e Francis Crick publicaram a estrutura de dupla hélice do DNA em 1953. O que eles fizeram foi uma continuação do que foi revelado por Franklin e eles determinaram que o DNA consiste em duas fitas complementares antiparalelas torcidas formando uma dupla hélice mantida no lugar por interações entre pares de bases complementares: adenina (A) com timina (T) e guanina (G) com citosina (C) (CARTWRIGHT; GRAUR, 2011).

Figura 2 - James Watson e Francis Crick com a estrutura de DNA



Fonte: Disponível em <https://www.sciencephoto.com/media/222943/view/watson-and-crick>

Já a descoberta da bactéria termófila *Thermus aquaticus* nos gêiseres de Yellowstone em 1969 por Thomas D. Brock e Hudson Freeze foi realmente valorizada após o desenvolvimento da técnica de PCR (*polymerase chain reaction*), pois foram a partir destas bactérias que houve o desenvolvimento da Taq DNA polimerase, enzima cuja característica principal é a resistência a altas temperaturas dos ciclos da PCR. A DNA polimerase I de *Thermus aquaticus* (Taq polimerase) é a enzima representativa mais famosa entre as DNA polimerases termoestáveis e a descoberta de sua capacidade de permanecer ativo após várias rodadas de ciclos térmicos repetidos expandiram suas aplicações para vários campos das ciências da vida. Portanto, devido à estabilidade ao calor da Taq polimerase, o tubo de reação pode permanecer na incubadora após a mistura de reação contendo a DNA polimerase ter sido preparada, e apenas mudanças de temperatura foram necessárias para a PCR. A Taq polimerase tem sido usada como a enzima padrão para PCR desde o seu início. (ISHINO, ISHINO, 2014; RAGHUNATHAN, 2019).

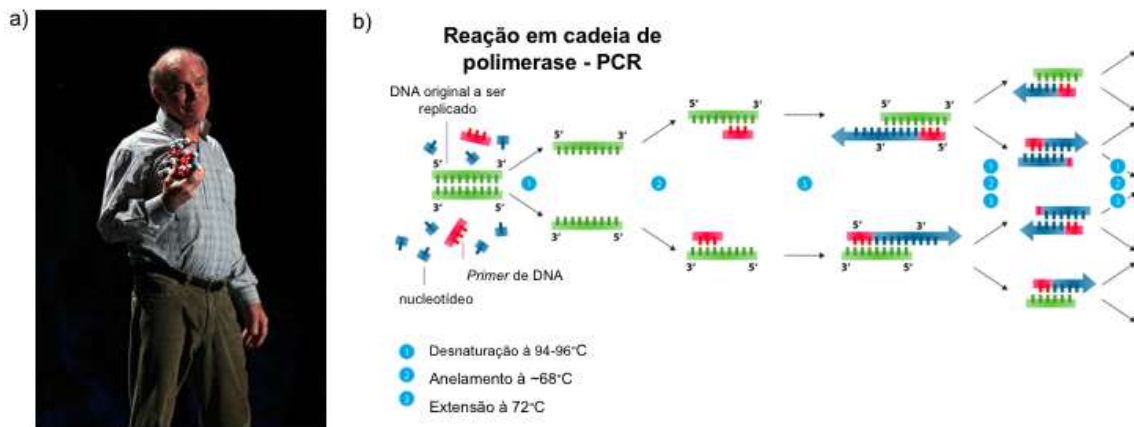
Figura 3 - Dr. Thomas Brock em Black Sand Pool, Parque Nacional de Yellowstone em setembro de 1995.



Fonte: Disponível em <https://www.nps.gov/features/yell/slidefile/researchissues/Page-1.htm>

Seguindo a ordem de descobertas importantes da biotecnologia moderna, a PCR (*polymerase chain reaction*) teve início a partir de uma técnica desenvolvida por Mullis (figura 4), principalmente para conseguir uma quantidade maior de DNA pela multiplicação exponencial do DNA. Em 1993, Kary Mullis recebeu o Prêmio Nobel de Química por inventar a reação em cadeia da polimerase após seu trabalho inicial na Cetus Corporation na década de 1980. A ideia foi concebida por Mullis em 1983, e posteriormente a metodologia foi desenvolvida através de um método consiste em ciclos repetitivos de desnaturação, hibridização e extensão da polimerase, aumentando em concentração exponencialmente, e o processo pode ser continuado por muitos ciclos e é inerentemente muito específico (MULLIS, 1986).

Figura 4- a) Kary Mullis e b) PCR – reação em cadeia de polimerase, aumento exponencial do DNA.



Fonte: Disponíveis em https://commons.wikimedia.org/wiki/File:Kary_Mullis_at_T
https://pt.m.wikipedia.org/wiki/Ficheiro:Polymerase_chain_reaction.svg

A PCR é um ensaio bem sensível, pois pouco DNA é necessário para gerar cópias suficientes para análises usando métodos convencionais de laboratório. A PCR pode ser realizada usando como fonte de DNA uma variedade de tecidos e organismos, que inclui sangue periférico, pele, cabelo, saliva e micro-organismos (GARIBYAN; AVASHIA, 2013).

Quando a PCR foi introduzida no meio científico foi revolucionário, pois na prática, era difícil obter uma molécula bem definida de DNA natural de qualquer organismo. A reação em cadeia da polimerase tornou a vida muito mais fácil para os biólogos moleculares, dando a eles um DNA específico (MULLIS, 1990).

Todas as tecnologias desde a original de Jeffreys, agora obsoleta para uso forense, passaram por importantes desenvolvimentos em termos de metodologia básica, ou seja, de *Southern blot* a PCR, de marcadores radioativos a fluorescentes, de géis em placas à eletroforese capilar.

Um evento muito importante para a identificação de pessoas foi a descoberta do DNA *fingerprinting* realizada por Jeffreys, na década de 80, evento que ocorreu muito próximo ao desenvolvimento da PCR.

A partir da pesquisa de Alec Jeffreys (1985a) a identificação humana deu um salto para que houvesse um avanço significativo em relação à identificação de pessoas e na utilização dessa técnica em casos forenses.

Jeffreys, encontrou padrões extraordinariamente variáveis e hereditários de DNA repetitivo analisado com sondas multi-locus. Este método foi chamado de impressão digital de DNA (figura 5), sua invenção abriu uma nova área da ciência.

A técnica revelou-se aplicável em muitas disciplinas biológicas, nomeadamente em estudos de diversidade e conservação entre espécies, e em estudos clínicos e antropológicos. A impressão digital de DNA tornou-se grandiosa quando os primeiros pedidos em casos civis e criminais foram publicados, pois introduziu novas evidências para investigações criminais e processos judiciais mostrando dimensão política e social desta técnica que foi além dos círculos acadêmicos. Esta técnica pode ser utilizada para comparação do DNA de uma pessoa com o material biológico identificado em uma cena de um crime ou ainda com o DNA de outra pessoa para fins de identificação ou exclusão. (ROEWER, 2012).

Figura 5 - DNA *fingerprinting* de Alec Jeffreys



Fonte: Disponível em <https://www.in-the-loop.net.au/godfather-dna-fingerprints-alec-jeffreys-revolutionized-solving-crime/>

Esse conjunto de acontecimentos e descobertas ao longo da história faz com que tenhamos o conhecimento para que a identificação de pessoas seja realizada de forma eficaz nos dias atuais. Obviamente existe muito mais sobre a história da

biotecnologia e desenvolvimento de técnicas que são utilizadas atualmente a ser abordados, porém destacamos os marcos mais importantes de acordo com o contexto do trabalho aqui realizado.

2.2 Identificação humana e primeiros casos forenses

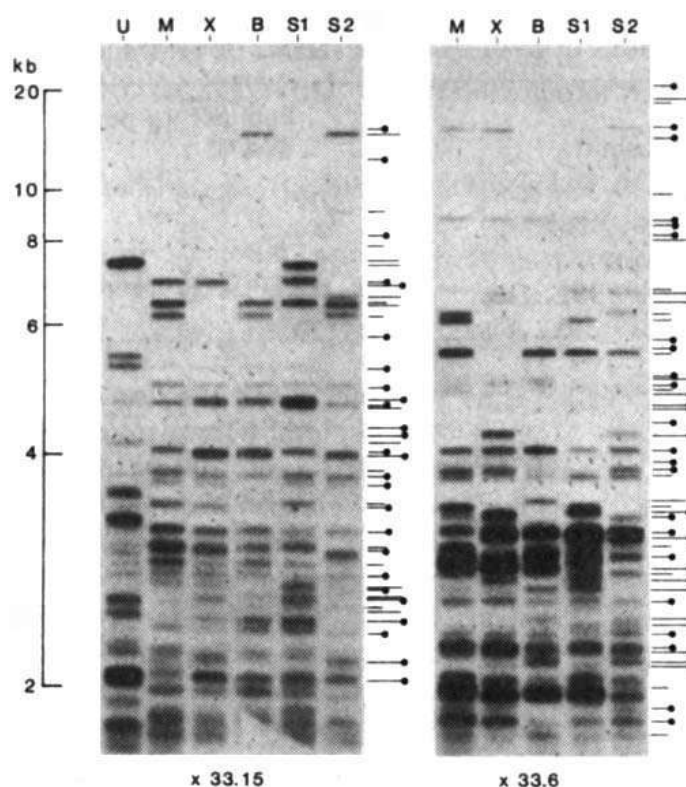
Historicamente, a identificação de pessoas no campo forense passou por três fases, começou com a análise do sistema de grupo sanguíneo ABO. Posteriormente, numa segunda fase novos marcadores para identificação de pessoas e identificação de paternidade foram baseados em variações de proteínas séricas e enzimas de glóbulos vermelhos; e também o sistema de antígeno leucocitário humano foi utilizado.

A primeira aplicação de impressões digitais de DNA salvou um menino da deportação e o método captou a simpatia do público. Nas palavras de Alec Jeffreys: “Se nosso primeiro caso fosse forense, acredito que teria sido questionado e o processo poderia ter sido prejudicado nos tribunais” (ROEWER, 2012).

O primeiro caso foi resolver um problema de imigração de uma família de Gana que vivia no Reino Unido. O filho mais novo da família visitou Gana e, ao retornar ao Reino Unido, as autoridades alfandegárias suspeitaram que o indivíduo era na verdade outra pessoa (JEFFREYS, 2013).

Na figura 6 abaixo é possível observar o DNA *fingerprinting* da família de Gana. As amostras do menino questionado (X), da mãe (M), das irmãs (S1 e S2), do irmão (B) e uma pessoa não relacionada (U).

Figura 6 - DNA *fingerprinting* da família de Gana envolvida na imigração.



Fonte: Jeffreys et al., 1985b.

Após esse caso Jeffreys foi contatado para que ajudasse a solucionar um crime. Em 1983, as autoridades descobriram o corpo de Lynda Mann, de quinze anos, em uma trilha pouco usada perto da vila de Enderby, no Reino Unido. Ela havia sido espancada, estuprada e estrangulada. Usando os melhores métodos disponíveis na época, a polícia determinou que o perfil de grupo sanguíneo do suspeito correspondia a aproximadamente dez por cento da população masculina (SAAD, 2005).

A partir das evidências deste caso, a polícia local fez a ligação dos casos em que duas estudantes foram estupradas e assassinadas com 3 anos de diferença, as semelhanças levaram a polícia a acreditar que uma pessoa era responsável por ambos. A polícia tinha um suspeito sob custódia, mas embora ele tenha confessado o segundo assassinato, ele negou o primeiro. Jeffreys foi solicitado a usar perfis de DNA para vincular o suspeito aos dois casos. Os resultados foram completamente inesperados: ambas as amostras de sêmen pertenciam ao mesmo homem, mas não eram do suspeito (ZAGORSKI, 2006). O primeiro suspeito, Buckland, se tornou a primeira pessoa exonerada pela técnica de impressão digital de DNA de Jeffreys. A técnica posteriormente foi usada para condenar o verdadeiro autor do crime, Colin

Pitchfork, que havia ocultado sua identidade. O verdadeiro criminoso foi preso e seu sangue testado, a tão procurada correspondência de DNA foi feita, e Pitchfork foi condenado pelos dois assassinatos (SAAD, 2005).

Em 1990, ele foi contatado por um promotor em Frankfurt, Alemanha, que buscava ajuda para provar que alguns restos mortais recuperados no Brasil pertenciam a Josef Mengele, o infame médico e oficial alemão nazista conhecido como anjo da morte. (JEFFREYS et al., 1992; ZAGORSKI, 2006).

A partir destes casos a utilização do DNA ganhou mais força e marcadores moleculares relacionados à genética forense para identificação humana foram sendo estudados.

2.3 Polimorfismos genéticos e marcadores biológicos para identificação humana

O DNA humano apresenta regiões nas quais ocorrem variações que faz com que os seres humanos sejam diferentes entre si. Essas regiões que apresentam variações que permitem identificar uma pessoa chegam a 0,3% a nível genético, e, os outros 99,7% do genoma humano é igual em toda a população humana (BUTLER, 2011).

Os polimorfismos genéticos, por meio de alelos múltiplos em loci individuais, fornecem um mecanismo para marcar um gene ou um pedaço de DNA, fornecendo uma ferramenta poderosa para uma variedade de investigações, dentre estas investigações podemos incluir a identificação de genótipos em estudos de paternidade e forenses, rastrear o movimento de indivíduos, experimentos de seleção em populações mapeamento de loci quantitativos que afetam características econômicas em plantas e animais, mapeamento de genes de doenças em humanos e comparações evolutivas de sequências de DNA (SINGH; KULATHINAL, 2013).

Os polimorfismos geralmente surgem como resultado de uma variação (mutação) e são normalmente referidos pelo tipo de mutação que os criou. A mutação genética pode ser definida como uma variação na sequência do DNA de ordem que ocorre em menos de 1% da população, enquanto as variantes extras prevalentes são identificadas como polimorfismos, ou seja, uma região específica do DNA para ser nomeada de “polimorfismo” tem que ter pelo menos 1% de variação

Os polimorfismos também podem ser de uma ou mais alterações de nucleotídeos, assim como as mutações. O SNP exemplifica o polimorfismo mais comum existente (KARKI et al., 2015).

A utilização de marcadores localizados no DNA fornece um número ilimitado de informação distribuídos aleatoriamente ao longo do genoma, por isso o desenvolvimento e uso de marcadores moleculares para a detecção e exploração desses polimorfismos do DNA é um avanço muito significativo no campo da genética molecular. Portanto, os marcadores moleculares são explorados e valorizados em diversas áreas que envolvem genética, biologia molecular e biotecnologia, tais como filogenia molecular, mapeamento genético, genética de populações, filogeografia, diagnósticos de doenças genéticas, testes de paternidade, genética forense (TURCHETTO-ZOLET et al., 2017).

Nas últimas três décadas, várias técnicas de marcadores moleculares foram desenvolvidas e exploradas em diferentes sistemas em todo o mundo. Diferentes técnicas para a análise de marcadores de DNA foram desenvolvidas a partir do surgimento da PCR, permitindo a amplificação de uma grande quantidade de uma sequência específica de DNA sem necessidade de clonagem, começando com apenas algumas moléculas da sequência alvo. Uma vantagem dos métodos de marcadores baseados em PCR é que requer pequenas quantidades de DNA. Entre os marcadores baseados em PCR, podemos citar, por exemplo: RAPD (*Random Amplified Polymorphism DNA*), ISSR (*Inter-simple sequence repeats*), SSR (*Simple Sequence Repeats*), AFLP (*Amplified Fragment Length Polymorphism*) (GROVER; SHARMA, 2016; TURCHETTO-ZOLET et al., 2017).

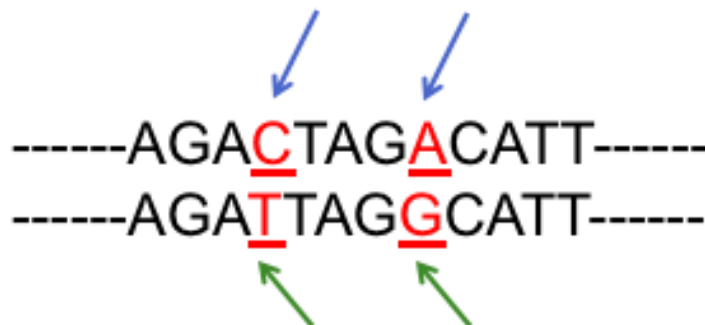
Os marcadores moleculares, devido à sua estabilidade, custo-benefício e facilidade de uso, fornecem uma ferramenta imensamente popular para uma variedade de aplicações, incluindo mapeamento de genoma, marcação de genes, diversidade genética, análise filogenética e investigações forenses (GROVER, SHARMA, 2016). Com isso, os marcadores genéticos permitem compreender a diversidade da população humana e proporciona um sistema de identificação genética de indivíduos observando seus polimorfismos. Portanto, os polimorfismos genéticos existentes em marcadores biológicos nos proporcionam a variabilidade no nosso DNA, apontando nossa individualidade.

Os polimorfismos genéticos mais utilizados na identificação de pessoas, podem ser classificadas em dois tipos: polimorfismos de sequência, originados por substituição de um ou mais nucleotídeos numa sequência de DNA como os SNPs (*single nucleotide polymorphism*) e polimorfismos de comprimento, originados por inserções ou deleções de um ou mais nucleotídeos como os marcadores genéticos VNTRs (*Variant number of tandem repeats*), conhecidos também por minissatélites, definidos por repetições consecutivas de unidades constituídas por aproximadamente 7 a 100 pares de base (pb), utilizados inicialmente nos exames de DNA, e os marcadores microssatélites, SSR ou mais conhecidos como STRs (*short tandem repeats*), repetições consecutivas de 2 a 6 pares de base (pb), que são os mais utilizados na rotina forense (BUTLER, 2011).

2.3.1 SNPs (*Single Nucleotide Polymorphisms*)

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, por isso são chamados de polimorfismos de sequência. Um SNP caracteriza-se, por exemplo, pela substituição de uma base “C” (citosina) por “T” (timina) num determinado ponto no genoma (figura 7). No projeto de análise do genoma em 2001 foram identificados mais de 1,4 milhão de polimorfismos de nucleotídeo único (SNPs) no genoma humano (LANDER et al., 2001).

Figura 7- Polimorfismo de sequência – SNP (*Single Nucleotide Polymorphism*)



Fonte: modificada de Butler, 2005.

Em geral, na área forense, os SNPs são muito utilizados, como as variações mais frequentes da sequência de DNA, podem ser genotipados por meio de

amplicons curtos, o que é uma característica interessante para melhorar o sucesso de amplificação em DNA altamente degradado, além de apresentar baixa taxa de mutação, importante para o uso em testes de parentesco. Entretanto, os SNPs bialélicos são menos informativos do que STRs, por isso, mais SNPs precisam atingir o mesmo nível de discriminação proporcionado pelos loci STR usados em ciência forense (HABIBI et al., 2019).

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 e DNA obtido de restos mortais, além de SNPs informativos de ancestralidade, bem como para inferir ascendência biogeográfica de indivíduos e desvendar características fenotípicas (AMBERS et al., 2016).

2.3.2 STRs (*Short tandem repeats*)

Os STRs (*Short Tandem Repeats*) também conhecidos como microsatélites, são repetições consecutivas de 2 a 6 pares de base (figura 8) e representam aproximadamente 3% do genoma humano comumente utilizados na rotina forense (LANDER et al., 2001; BUTLER, 2007).

Figura 8 - Polimorfismo de comprimento – STRs (*short tandem repeats*)



Fonte: modificada de Butler, 2005.

Os STRs são marcadores abundantes, encontrados em todo o genoma, fáceis de automatizar, universais, reproduzíveis, robustos e confiáveis. Os padrões de polimorfismo exibidos por microsatélites são maiores do que qualquer outro sistema de marcadores contemporâneo (GROVER; SHARMA, 2016). O número de

unidades de repetição é altamente variável entre os indivíduos, o que oferece um alto poder de discriminação quando analisado para fins de identificação (BUTLER, 2011; WYNER et al., 2020)

Além disso, são marcadores genéticos altamente polimórficos encontrados em todo o genoma humano que se tornaram rotineiramente usados em estudos forenses e genéticos populacionais com o objetivo de compreender a identidade humana e a diversidade genética específica da população (SZIBOR et al., 2003; YUAN et al., 2014; NOVROSKI, 2020). Outra característica a ser destacada dos STRs é a facilidade das técnicas que envolvem estes marcadores devido à análise rápida com tecnologia baseada em PCR e detecção fluorescente automatizada por eletroforese capilar, e também a capacidade de gerar multiplexes com pequenos comprimentos de *amplicon* para DNA degradado (GOMES et al., 2020).

Os STRs são encontrados em todo o genoma humano, podem ser encontrados em seres procariotos e eucariotos. Entretanto sua distribuição não é uniforme, eles são encontrados mais nas regiões subteloméricas. A maioria dos STRs estão localizados em regiões não codificantes, por isso, não estão implicados na expressão gênica, apenas uma pequena parte, cerca de 8% se localizam nas regiões codificantes (FAN e CHU, 2007; WYNER et al., 2020).

A denominação do locus se dá por meio de letras e números, por exemplo, o locus D3S1266, onde D representa DNA, 3 significa cromossomo 3 no qual o locus STR se localiza, S representa STR e 1266 é o identificador único (FAN e CHU, 2007).

O número de unidades de repetição dos STRs é altamente variável entre os indivíduos, o que oferece um alto poder de discriminação quando analisado para fins de identificação. A sequência de repetição dos STRs é nomeada a partir do comprimento dessa sequência que se repete, por isso com base em diferentes unidades de repetição, os STRs podem ser classificados em diferentes tipos. De acordo com o comprimento da unidade de repetição principal, os STRs são classificados em repetições mono-, di-, tri-, tetra-, penta- e hexanucleotídeo. O número total de cada tipo diminui à medida que o tamanho da unidade de repetição aumenta. Quando analisado o genoma humano os STRs mais comuns são repetições de dinucleotídeos (FAN e CHU, 2007; BUTLER, 2011).

Normalmente, 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, por isso outros marcadores genéticos são estudados e também vem sendo utilizados para complementar as análises (PINTO et al, 2013b).

Sem dúvida os marcadores genéticos mais utilizados na identificação humana são os que estão presentes nos cromossomos autossômicos, mais especificamente os STRs autossômicos (YUAN, 2014; GOMES et al., 2020)

Na identificação de uma pessoa, para gerar um perfil de DNA forense é atualmente realizado usando um painel de marcadores STRs multialélicos (*short tandem repeats*) que são estruturalmente análogos aos minissatélites originais, mas com segmentos repetidos muito mais curtos e, portanto, mais fáceis de amplificar e multiplexar com PCR. Em uma injeção de eletroforese capilar podem ser detectados até 30 STRs, isso irá gerar um código genético único para identificação de um indivíduo.

Basicamente, existem dois conjuntos de marcadores STRs em conformidade com os padrões solicitados por bancos de dados criminais em todo o mundo: o conjunto padrão europeu de 12 marcadores STR e o padrão CODIS (*Combined DNA Index System*) criado em 1998 de propriedade exclusiva do *Federal Bureau of Investigation* (FBI) unidade de polícia do departamento de justiça dos Estados Unidos da América inicialmente com 13 marcadores. Devido à sobreposição parcial, eles formam juntos um padrão de 18 marcadores STR no total (BUDOWLE et al., 1998; GILL et. al., 2006; ROWER, 2013).

Em janeiro de 2017 a unidade do *Federal Bureau of Investigation* (FBI) adicionou mais 7 loci aos 13 já existentes. Os 13 *loci* iniciais: CSF1PO; D3S1358; D5S818; D7S820; D8S1179; D13S317; D16S539; D18S51; D21S11; FGA; TH01; TPOX; vWA. E os 7 *loci* efetivos a partir de 2017: D1S1656; D2S441; D2S1338; D10S1248; D12S391; D19S433; D22S1045 formando então os 20 marcadores STRs utilizados pelo FBI.

Apesar dos STRs autossômicos serem os mais utilizados na análise forense, principalmente para identificação de pessoas, outros marcadores encontrados nos

cromossomos sexuais, e os polimorfismos no DNA mitocondrial podem ser utilizados para análise, pois cada um possui suas especificidades. Foram desenvolvidas metodologias para utilização de marcadores STRs localizados nos cromossomos sexuais usadas por muitos laboratórios forenses para complementar informações individualizadas fornecidas por sistemas STRs autossômicos (DIEGOLI, 2015).

2.3.3 InDels (*Insertion/Deletion polymorphisms*)

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.

Os InDels indicam inserções ou deleções (inserção-deleção) de fragmentos de nucleotídeos de tamanhos diferentes no mesmo local na sequência do genoma entre a mesma espécie ou espécies intimamente relacionadas. Os InDels podem ser definidos também como uma lacuna na sequência derivada do alinhamento da sequência homóloga, essas inserções e deleções estão amplamente distribuídas por todo o genoma e ocorre em alta densidade (YANG et al. 2016).

A partir do estudo de Weber e colaboradores (2002) que identificaram e caracterizaram 2.000 inserções e deleções - InDels ao longo do genoma humano, foi possível dar início a estudos destes marcadores que perduram até o momento. Os InDels destacam-se devido à abundância no genoma humano e a facilidade de análise, além de sua utilidade para estudos genéticos.

Por isso, amostras de DNA misturadas ou altamente degradadas podem ser amplificadas com sucesso com marcadores InDels e tipadas de forma eficaz (YANG et al., 2016).

Estes polimorfismos possuem características interessantes para uso na identificação humana, pois possuem baixa taxa de mutação quando comparados aos STRs; são amplamente distribuídos no genoma, sendo encontrados em média a cada 7,2 kpb (MILLS, 2006); o que representam 20% de todos os polimorfismos do genoma humano (WEBER, 2002).

Atualmente muitos estudos a fim de aumentar a possibilidade de utilização de InDels na prática forense foram publicados, tanto de autossômicos como o realizado por Pereira e colaboradores (2009) que descreveram um novo multiplex para a

identificação humana na qual foi possível analisar 38 InDels autossômicos em uma única reação, estudo realizado também para a população brasileira por Manta (2012) e Martinez (2019) que realizou o estudo com InDels do cromossomo X. Além de outros estudos publicados com o kit comercial que analisa os polimorfismos de inserção/deleção, chamado DIPplex Kit[®] (Qiagen) para cromossomos autossômicos.

2.3.4 DNA mitocondrial

A mitocôndria é uma organela encontrada no citoplasma, com muitas cópias por célula, esse é um dos motivos para o DNA mitocondrial ser utilizado em análises forenses, pois ele se conserva mais, além de possuir mais cópias. A provável origem das organelas como mitocôndria e cloroplasto é pela teoria endossimbiótica (MARTIN et al., 2015).

A herança do DNA mitocondrial é exclusivamente materna, ou seja, não há contribuição masculina no DNA mitocondrial, pois durante a fecundação o espermatozoide não colabora na transmissão das mitocôndrias.

A molécula de dupla fita circular foi descoberta por Nass & Nass, em 1963, sua sequência foi determinada em 1981 por Anderson e colaboradores e denominada *Cambridge Reference Sequence (CRS)* (ANDERSON et al., 1981), posteriormente re-analisada e revisada por Andrews e colaboradores (ANDREWS et al., 1999), sendo modificada e denominada *revised Cambridge Reference Sequence (rCRS)*. Esta sequência é utilizada como padrão nas análises dos polimorfismos do DNA mitocondrial, portanto quando encontrados polimorfismos - SNPs (*Single Nucleotide Polymorphisms*), na sequência de um DNA analisado, este polimorfismo é comparado com a sequência revisada - *revised Cambridge Reference Sequence (rCRS)*.

O DNA mitocondrial é composto por região controle, hipervariável ou não codificante (D-Loop) e região codificante. A região controle, composta por região hipervariável 1, 2 e 3, é geralmente a mais utilizada nas análises de identificação de um haplótipo, porém, quando há necessidade da utilização de dados adicionais outras regiões podem ser analisadas ao longo do DNA mitocondrial.

O DNA mitocondrial vem sendo utilizado em vários estudos desde os que resultam em dados populacionais (PANETO, 2010; BRAGANHOLI et al., 2019), os

relacionados à ancestralidade em estudo abordando como neandertais estão relacionados com os humanos modernos (KRINGS et al., 1997), migração de populações ao longo do tempo (RELETHFORD, 2004), estudos na identificação de pessoas em desastres e tragédias, guerras como a do Vietnã (HOLLAND et al., 1993), em estudos como o da identificação da família real russa (GILL et al., 1994), do fora da lei americano Jesse James (STONE et al., 2001) até no atentado das torres gêmeas do World Trade Center (BUDOWLE, 2003) onde o DNA mitocondrial auxiliou na identificação das vítimas. Isso faz com que o DNA mitocondrial seja uma ferramenta eficaz na identificação de pessoas, principalmente para traçar a linhagem materna.

O DNA mitocondrial (mtDNA) é importante para as análises quando há dificuldade para amplificar as amostras de DNA nuclear, ou seja, amostras de restos mortais como ossos não identificados, fios de cabelo sem raízes ou espécimes muito antigos onde o DNA está altamente degradado. O DNA mitocondrial e o cromossomo Y possuem um modo incomum de herança não recombinante que enfraquece o peso estatístico de uma correspondência entre amostras individuais, mas torna o método eficiente para a reconstrução da relação paterna ou materna (ROWER, 2013).

2.3.5 Cromossomos Sexuais

Os cromossomos sexuais foram originados de um par de cromossomos homólogos e seguem padrões que divergem dos autossômicos, além de eventos de supressão de recombinação relacionada ao cromossomo X que ocorreram ao longo da história evolutiva dos mamíferos. Ainda que haja diferenças entre os cromossomos sexuais e autossômicos, o cromossomo X e Y ainda apresentam divergências nas quais os tornam únicos, como a perda de genes do cromossomo Y e a tentativa de equivalência pelo silenciamento e inativação de um dos cromossomos X (POSYNICK et al., 2019).

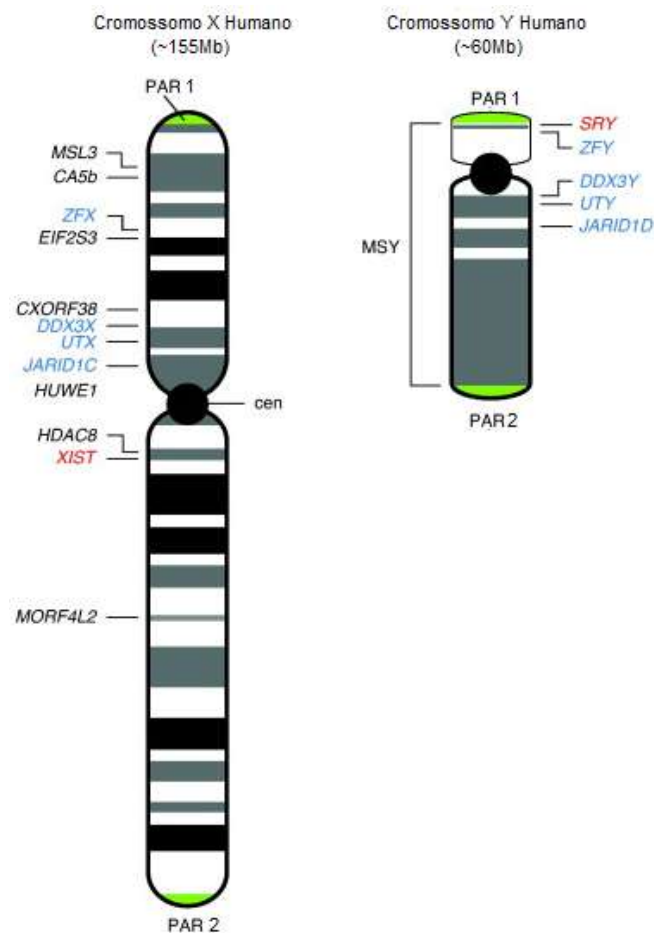
Na figura 9 abaixo, adaptada de Wijchers; Festenstein (2011), é possível observar os dois cromossomos: X e Y. Nota-se a diferença de tamanho entre estes cromossomos. As regiões pseudoautossômicas (PAR) estão representadas nas extremidades dos cromossomos em verde. Na extremidade do braço curto dos

cromossomos X e Y está localizada a PAR1, e na outra extremidade do braço longo, PAR2. Na região PAR 2 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).

Os genes representados em vermelho estão específicos da determinação do sexo. Os genes em azul são os que possuem um par homólogo no Y, a homologia entre os cromossomos 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

A região intermediária entre as PAR 1 e 2 no Y é específica do sexo masculino e corresponde a 95% do cromossomo, sendo chamada de MSY (do inglês *Male-Specific region of the Y*) (WIJCHERS; FESTENSTEIN, 2011).

Figura 9 - Representação esquemática dos cromossomos sexuais X e Y.



Fonte: modificada de Wijchers; Festenstein, 2011.

A diferença entre os padrões de herança dos cromossomos X e do Y diferem dos autossômicos e entre si, deste modo, marcadores localizados nesses cromossomos tem o potencial de oferecer informações diferentes, mas complementares. Os STRs localizados em cromossomos autossômicos foram usados na ciência forense muito antes da aplicação da análise dos cromossomos sexuais. Apesar de alguns marcadores terem sido relatados há vários anos, a intenção de usar esses marcadores como uma ferramenta para aplicações especiais desenvolveu-se tardiamente. Embora o uso forense de marcadores Y-STR esteja bem documentado e geralmente aceito pela comunidade científica e pelos tribunais em investigações e processos que utilizam análise forense, marcadores no cromossomo X são menos universalmente aplicados (DIEGOLI et al., 2011; SZIBOR et al, 2003)

No entanto, a genotipagem de marcadores do cromossomo X pode complementar a análise de marcadores autossômicos e do cromossomo Y com muita eficiência, especialmente em casos complexos de parentesco (SZIBOR et al., 2003). Assim sendo, marcadores localizados no cromossomo X podem ser uma importante ferramenta na análise forense, pois contribuem com a investigação mais complexa das relações biológicas em que a análise estatística se apresenta inconclusiva. (SZIBOR, 2007).

2.3.5.1 Cromossomo Y

O cromossomo Y está presente somente nos indivíduos do sexo masculino, por isso ele não individualiza uma pessoa, mas pode indicar a linhagem paterna do indivíduo. O cromossomo Y possui a região PAR 1 e 2 que podem sofrer recombinação, nas extremidades do cromossomo. A outra porção do cromossomo Y pode ser chamada de região não recombinante, pois não sofre recombinação e possui cerca de 60 milhões de pares de base (POSYNICK et al., 2019).

Os marcadores do cromossomo Y são os STRs encontrados na região não recombinante, essa ausência de recombinação implica em transmissão de haplótipos e dessa forma é possível traçar a linhagem paterna, ou seja, o cromossomo Y não individualiza, pois, um homem possui o mesmo haplótipo que seu genitor masculino.

Os marcadores do cromossomo Y são utilizados em investigações de paternidade, investigações criminais, ou seja, quando é identificado um haplótipo na investigação da cena do crime pode: excluir suspeitos do sexo masculino do envolvimento no crime; identificar a linhagem paterna dos estupradores do sexo masculino; destacar vários contribuintes do sexo masculino para um rastreamento e ainda fornecer pistas investigativas para encontrar estupradores desconhecidos do sexo masculino (KAYSER, 2017). Além disso, o cromossomo Y pode ser utilizado em estudos de genética de populações e evolução humana além de fluxo migratórios das populações.

Os dados sobre os marcadores genéticos do cromossomo Y podem ser encontrados em bancos de dados como o YHRD (*Y-Chromosome STR Haplotype Reference Database* - yhrd.org), que fornece informações sobre os haplogrupos do cromossomo Y, mostra a área geográfica a qual esse haplogrupo pode ser encontrado, este banco de dados possui inclusive dados da população brasileira (PALHA et al., 2010; STANGE et al., 2019; AMBROSIO et al., 2020).

O primeiro caso de destaque forense para o cromossomo Y ocorreu na Holanda, na investigação de um estupro e morte da vítima, onde o criminoso foi encontrado a partir do rastreamento da linhagem paterna (KAYSER, 2017).

Tanto o cromossomo Y quanto o DNA mitocondrial são marcadores de linhagem e têm aplicações especiais em genética forense. A análise do cromossomo Y é muito útil em casos específicos criminais e auxiliam na identificação padrão com os STRs autossômicos principalmente no rastreamento para a identificar linhagens (ROWER, 2013).

2.4 Cromossomo X

2.4.1 Características do cromossomo X

Os estudos relacionados ao cromossomo X vem avançando ao longo dos anos com vários trabalhos já publicados. Em 2001, Lander e colaboradores publicaram um estudo com o sequenciamento inicial do genoma humano, o qual indicou que o cromossomo X possui um comprimento de tamanho de aproximadamente 155 milhões de pares de bases (Mb) representando quase 5% do tamanho estimado do

genoma humano (3.200 Mb). Já em 2005, Ross (2005) e colaboradores publicaram a sequência eucromática do cromossomo, uma das primeiras publicações que cobriu aproximadamente 99,3% do cromossomo X humano (GOMES et al. 2020).

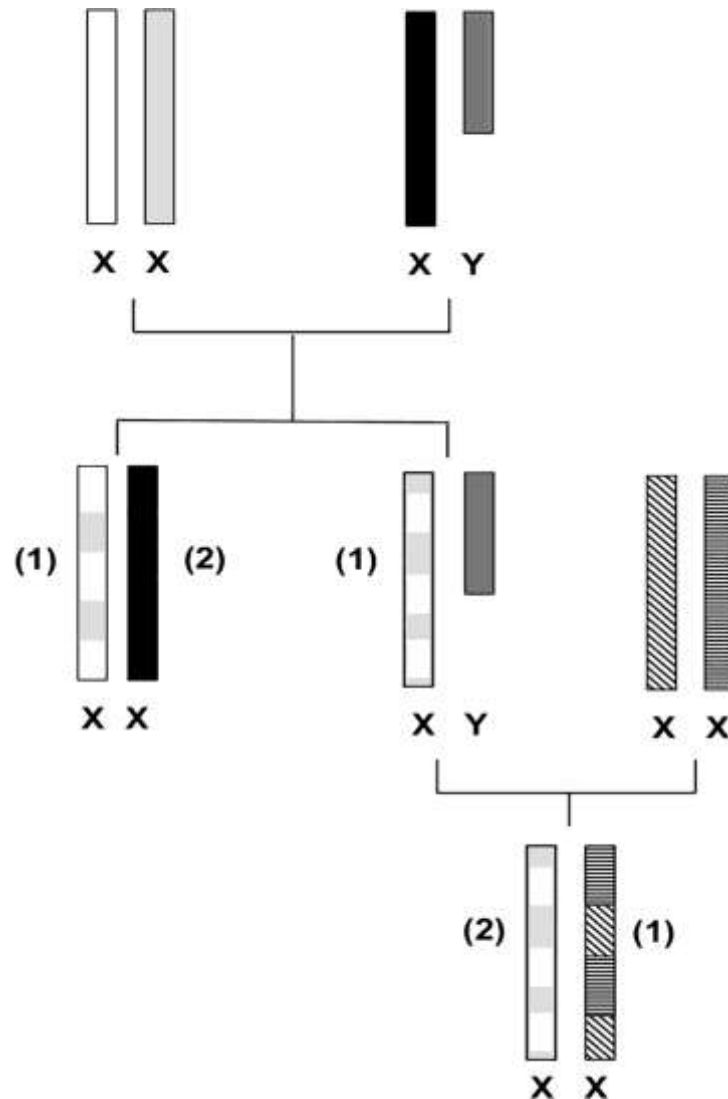
O cromossomo X possui uma biologia única na qual foi moldado para ser o cromossomo sexual que pode ser encontrado em ambos os sexos, sendo compartilhado por homens e mulheres. Essa biologia única do cromossomo X deve-se a evolução dos cromossomos sexuais dos mamíferos, onde nos últimos 300 milhões de anos os cromossomos sexuais evoluíram a partir de um par de cromossomos autossômicos. (ROSS et al., 2005).

Além disso, similar ao cromossomo autossômico ancestral, o cromossomo X sofre recombinação em mulheres, enquanto nos homens a falta de recombinação permite a identificação direta de haplótipos (POSYNICK et. al. 2019; BERGUERO et al, 2009).

O cromossomo X possui características que possibilita investigar padrões de migração de homens e mulheres ao longo da história e o padrão de mutação e recombinação entre os sexos. Na figura 10, podemos observar que os descendentes femininos e masculinos herdam um cromossomo X materno recombinado (1) que resultou da meiose feminina. As filhas herdam um cromossomo X paterno inalterado (2) devido à falta de recombinação, com exceção das regiões da região pseudoautossômica (PAR) (GOMES et al., 2020).

Por isso, em casos de investigação de parentesco no qual o descendente é do sexo feminino, a análise do cromossomo X mostra parâmetros mais eficientes que se complementam aos autossômicos tornando a análise extremamente eficiente (PEREIRA et al., 2012).

Figura 10 - Padrão de herança do cromossomo X.



Fonte: Gomes et al., 2020.

Ainda há casos em que há dificuldade de se analisar o material biológico, cujo conjunto de marcadores rotineiramente usado precisa ser complementado devido à baixa quantidade de DNA amplificado, e o material biológico a ser analisado apresenta-se degradado, como o caso de exumações de restos mortais. Outras possibilidades ocorrem quando a análise de marcadores do cromossomo X pode fornecer informações adicionais valiosas, como na ausência do material biológico do suposto pai que não está disponível, e faz-se necessário utilizar parentes (ascendentes ou descendentes) para genealogias paternas: meias-irmãs, irmãs e avó-neta, casos de maternidade, casos de violência sexual com feto do sexo feminino, análise de vestígio biológico e de ancestralidade (SZIBOR et al, 2003; BUTLER, 2011; PINTO et al., 2013).

2.4.2 Marcadores genéticos do cromossomo X

A partir de 1999 os primeiros estudos genéticos com foco na genotipagem de X-STRs para aplicação forense começaram a surgir. Desde então, e até aproximadamente o ano de 2008, houve um aumento notável de publicações de dados populacionais. No entanto, as informações relatadas sobre X-STRs humanos em diferentes populações mundiais têm estagnado nos últimos anos (GOMES et. al., 2020).

Na introdução da utilização forense do cromossomo X para obtenção de dados como frequências alélicas e parâmetros forenses, e para que estes dados pudessem ser utilizados por pesquisadores e peritos, foram criados ao longo dos anos bancos de dados que trazem informações importantes, destacando-se o *website* <http://chr-x-str.org/> especializado em informações de X – STRs (SZIBOR et al., 2006), onde está descrito os X- STRs mais utilizados na época da criação do banco (SZIBOR et. al., 2007). Além deste banco de dados com informações de populações de todo o mundo, foi desenvolvido o banco genético brasileiro do cromossomo X (<http://bgbx.com.br/?pagina=principal>) com dados da população brasileira (MARTINS et al., 2014). As informações sobre o cromossomo X que estão disponibilizadas nesses bancos de dados são importantes para auxiliar na resolução de casos forenses e de identificação humana, assim como auxiliar no conhecimento genético de populações.

É interessante ressaltar 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).

A partir do início dos estudos com o cromossomo X, foram surgindo empresas desenvolvedoras de sistemas *multiplex* inseridos em kits comerciais para a análise de marcadores do cromossomo X. O primeiro a surgir foi o kit Mentype® Argus X-UL (*Biotype Diagnostic GmbH*, Dresden, Alemanha) que amplificava 4 marcadores X-STR, este kit apesar de não ser disponibilizado no Brasil, permaneceu no mercado de 2003 a 2008. Em 2005, a mesma empresa *Biotype Diagnostic GmbH* lançou outro kit multiplex, o Mentype® Argus X-8, com 8 marcadores e incluía a amelogenina, porém não foi comercializado no Brasil. O kit mais recente e utilizado

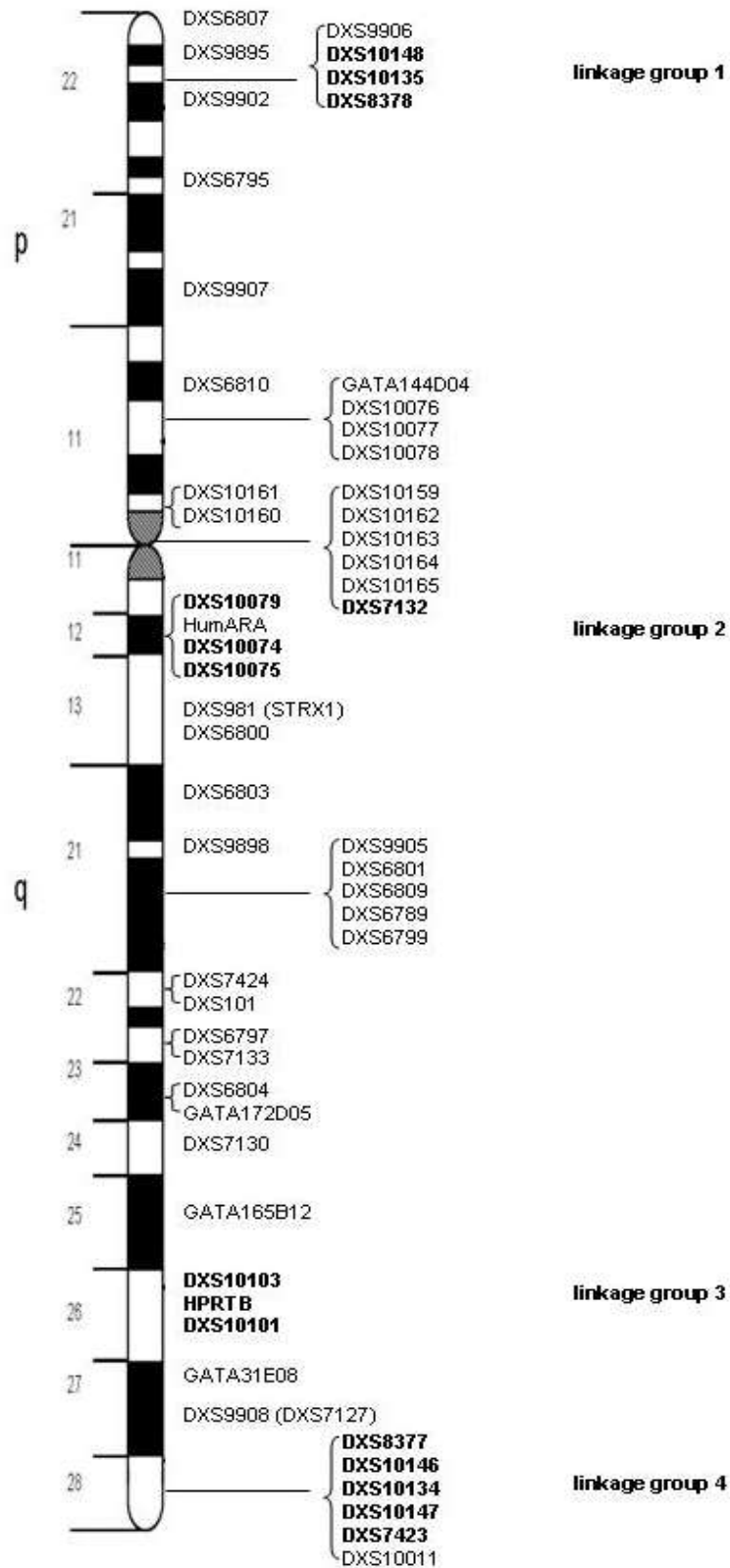
também no Brasil é o Investigator Argus X-12 (Biotype; Qiagen), que amplifica simultaneamente 12 X-STRs mais amelogenina.

Outros sistemas *multiplex* também foram desenvolvidos por grupos independentes, como o trabalho colaborativo utilizando X-STRs pelo Grupo de Línguas Espanhola e Portuguesa da Sociedade Internacional de Genética Forense (GHEP-ISFG), foi desenvolvido 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). Quando este estudo foi realizado ainda não havia o Investigator Argus X-12 (Qiagen, Hilden, Alemanha), por isso a importância deste trabalho que possibilitou a elaboração do Decaplex e a utilização destes marcadores X-STR nas populações do Brasil e outros países da América latina (MARTINS, 2011).

Ademais, 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. Em alternativa, o Decaplex (GUSMÃO et al., 2009) consiste em marcadores independentes, que são também muito informativos nos estudos de identificação.

Na figura 11 observamos aspectos que caracterizam os dois conjuntos de marcadores utilizados neste trabalho, os quatro grupos de ligação (*linkage group* ou LG) que formam os 12 marcadores do cromossomo X do kit Argus (Qiagen, Hilden, Alemanha) além dos dez marcadores independentes que estão presentes no Decaplex.

Figura 11 - Ideograma do cromossomo X com a localização de seus marcadores X-STRs.



Fonte: <http://chrX-str.org/>

Ainda um pouco mais recente Prieto-Fernández (2016) publicou o desenvolvimento de outro sistema *multiplex* com a análise de 17 X-STRs, incluindo os marcadores do sistema Decaplex, demonstrando uma confiabilidade destes marcadores para fins de genotipagem forense. Outros kits também foram disponibilizados no mercado como o Goldeneye DNA ID System 17X (*Goldeneye Technology Co., Ltd.*, Pequim, China) e o kit AGCU X19 STR (*Wuxi Sino-German Meilian Biotechnology Co.*, Jiangsu, China) também foram desenvolvidos para aplicações forenses, embora os dados disponíveis sejam virtualmente restritos à população chinesa (GOMES et al., 2020).

Além dessa publicação, na tabela 1 abaixo é possível observar os X-STRs mais utilizados nos últimos anos de acordo com a revisão de Gomes (2020).

Tabela 1 – X-STRs utilizados em estudos nos últimos anos. As referências não se referem necessariamente aos documentos de desenvolvimento originais.

Nome	Referências	Nr. e STR loci
Kit Goldeneye 17X	Gao et al. (2019)	16 (DXS6795, DXS9902, DXS8378, HPRTB, GATA165B12, DXS7132, DXS7424, DXS6807, DXS6803, GATA172D05, DXS6800, DXS10134, GATA31E08, DXS10159, DXS6789 e DXS6810)
Kit Investigator [®] Argus X-12 QS (Qiagen)	Elakkary et al. (2014)	12 (DXS7132, DXS7423, DXS8378, DXS10074, DXS10079, DXS10101, DXS10103, DXS10134, DXS10135, DXS10146, DXS10148 e HPRTB)
Kit de sistema de identificação Microreader [™] 19X	Lin et al. (2020)	19 (DXS6795, DXS6803, DXS6807, DXS9907, DXS7423, GATA172D05, DXS101, DXS9902, DXS7133, DXS6810, GATA31E08, DXS6800, DXS981, DXS10162, DXS6809, GATA165B12, DXS10079, DXS10135, HPRTB)
AGCU X19 STR Kit	Li et al. (2017)	19 (DXS10074, DXS10075, DXS10079, DXS101, DXS10101, DXS10103, DXS10134, DXS10135, DXS10148, DXS10159, DXS10162, DXS10164, DXS6789, DXS6809, DXS7132, DXS7423, DXS7424, DXS8378, HPRTB)
-	Deng et al. (2017)	19 (DXS9895, GATA144D04, DXS10077, DXS10078, DXS10161, DXS10160, DXS981, DXS6800, DXS6803, DXS9898, DXS6801, DXS6799, DXS6797, DXS7133, DXS6804, GATA172D05, DXS8377, DXS10146, DXS10147)

Nome	Referências	Nr. e STR loci
-	Prieto-Fernández et al. (2016)	17 (DXS8378, DXS9898, DXS7133, GATA31E08, GATA172D05, DXS7423, DXS6809, DXS7132, DXS9902, DXS6789, DXS10079, DXS6801, DXS6799, DXS6800, DXS10075, DXS6807, DXS6803)
GHEP-ISFG decaplex	Gusmão et al. (2009)	10 (DXS8378, DXS9898, DXS7133, GATA31E08, GATA172D05, DXS7423, DXS6809, DXS7132, DXS9902 e DXS6789)
-	Zhang et al. (2017b)	15 (DXS6807, DXS8378, DXS6795, DXS10164, DXS7132, DXS10074, DXS6803, DXS6801, DXS101, DXS7133, GATA165B12, DXS10103, HPRTB, GATA31E08 e DXS7423)
MiSeq FGx™ Forensic Genomics	Jäger et al. (2017)	7 (HPRTB, DXS7132, DXS7423, DXS8378, DXS10074, DXS10103 e DXS10135)

Fonte: modificada de Gomes et al., 2020.

Apesar da ampla quantidade de conjuntos de marcadores X-STR disponíveis e utilizados em muitos estudos populacionais que surgiram ao longo desses anos, todos os dados existentes infelizmente não estão compilados em um único banco de dados.

2.4.3 Ligação e desequilíbrio de ligação para o cromossomo X

O cromossomo X possui algumas características das quais fazem com que alguns parâmetros sejam avaliados para a utilização de seus marcadores na prática forense.

A ligação gênica ou *linkage* ocorre quando a localização de um *loci* está em um mesmo cromossomo e não se segrega de forma independente no momento da formação de um gameta. A ligação está relacionada com a dependência física de *loci* localizados próximos (TILLMAR et al, 2008).

De acordo com Tillmar e colaboradores (2017), a ligação é uma consequência do processo de recombinação que ocorre entre dois cromossomos homólogos durante a meiose. Quando acontece de mais de 50% dos gametas tiver o mesmo segmento do cromossomo parental, ou seja, no qual os dois marcadores estão localizados, os dois marcadores estarão ligados. Portanto, dois marcadores

são considerados não vinculados quando se espera que um número de recombinações ocorra em 50% das vezes durante a meiose.

Já o desequilíbrio de ligação mede a co-segregação de alelos em nível populacional. Para um determinado conjunto de marcadores, a ligação pode ser medida através da obtenção frequências de recombinação de amostras, e o desequilíbrio de ligação pode ser estimado a partir de frequências de alelos e haplótipos. (TILLMAR et al., 2008)

A taxa de recombinação meiótica é estimada por pares de marcadores e transformada em distância genética que pode ser expressa em centi-Morgan (cM), sendo que 1 cM equivale a 1 esperada recombinação por 100 meioses. A distância física e a genética não estão necessariamente relacionadas, por isso não se deve supor que marcadores estejam ligados baseando-se apenas na distância física, a taxa de recombinação é determinante para a confirmar a ligação (SZIBOR et al., 2003).

Os marcadores com uma distância maior que 50 cM são ditos não ligados e espera-se que não seja encontrado desequilíbrio de ligação, já um valor inferior a este as chances de ocorrer desequilíbrio de ligação aumentam (SZIBOR et al., 2007).

A distância genética não é o único fator que determina o desequilíbrio de ligação em uma população, fatores como seleção natural, efeito fundador, miscigenação, subestruturação populacional e deriva genética também podem influenciar no desequilíbrio, por isso, é importante que cada população seja avaliada, de modo que cada característica dessa população e esses fatores acima citados sejam levados em consideração (SZIBOR et al., 2003; ZARRABEITIA et al., 2006).

2.4.4 Aplicação dos marcadores genéticos do cromossomo X

Na análise de parentesco, existem casos considerados padrão em que as análises do cromossomo X não são necessárias, deste modo, podemos citar os casos de paternidade padrão (duos e trios) que podem ser resolvidos apenas com os marcadores autossômicos. Entretanto, em várias situações os marcadores do cromossomo X podem ser mais informativos que os autossômicos, DNA mitocondrial e cromossomo Y, devido ao seu modo específico de herança (SZIBOR et a., 2007; TILLMAR et al., 2017)).

Por isso, normalmente, quando a evidência estatística obtida é considerada insuficiente, marcadores adicionais são analisados como os marcadores de Inserção/deleção (InDels) e polimorfismos de nucleotídeo único (SNPs) foram sugeridos como marcadores para incluir nessas análises de parentesco mais detalhadas (GOMES et al.,2012).

Além destes marcadores citados acima, em casos de paternidade, onde a informação dos marcadores autossômicos não é suficiente para chegar a uma conclusão, os marcadores cromossômicos X-STRs podem representar uma ferramenta suplementar útil.

Dentre as possibilidades de análises que podemos aplicar o cromossomo X (STRs) estão:

- Casos de incompatibilidade(s) entre filhas e suposto pai (duo ou trio):

Quando na análise dos resultados poucas inconsistências são observadas entre o suposto pai e a filha, mesmo havendo alta razão de verossimilhança, uma das explicações mais prováveis são que ocorreram mutações ou que o suposto pai está intimamente relacionado ao pai verdadeiro como um irmão, um filho, ou pai; ou seja, apesar do LR (razão de verossimilhança) ser alto, ainda há incompatibilidades (PINTO et al., 2013). Portanto, todas estas dúvidas relacionadas aos resultados obtidos podem ser resolvidas com a análise de marcadores do cromossomo X (TILLMAR, 2017; GOMES, 2012).

- Casos deficientes de paternidade:

Com a análise apenas dos STRs autossômicos, estes casos podem ser inconclusivos, por isso quando o suposto pai não está presente, o DNA de parentes é analisado para a reconstrução de seu perfil.

No caso avó (paterna)-neta o cromossomo X pode ser um fator decisivo para atestar o vínculo biológico, visto que o pai herda o cromossomo X materno e a filha conseqüentemente herda o único cromossomo X que o pai possui (BUTLER et al. 2011).

Outro caso que pode ser analisado pelo DNA de parentes é o de meias-irmãs paternas, neste caso, as duas meias-irmãs obrigatoriamente devem possuir os alelos do cromossomo X idênticos, devido ao padrão de herança, o qual elas herdam o cromossomo X paterno. Assim, neste tipo de investigação de paternidade a presença de quatro diferentes alelos nos marcadores do cromossomo X permite a exclusão de paternidade (SZIBOR et al. 2003).

Outro exemplo: considere um caso de paternidade entre pai e filha, onde os supostos pais são parentes, se o pai biológico for o pai (com mães diferentes) do suposto pai, ou se o pai biológico for o filho do suposto pai, eles não compartilharão os alelos cromossômicos X idênticos por descendência. Para esses casos, uma análise usando marcadores cromossômicos X pode ser muito mais informativa do que usar marcadores autossômicos (TILLMAR, 2017; GOMES, 2012).

Nestes casos é possível perceber o alto poder de exclusão do cromossomo X, devido ao seu padrão de herança o qual a filha obrigatoriamente deve possuir o cromossomo X paterno, através do compartilhamento de alelos cromossômicos X idênticos por descendência (TILLMAR, 2017; GOMES, 2012).

- Casos com dificuldade de analisar o material biológico:

Há casos em que o conjunto de marcadores STRs autossômicos utilizados na rotina de laboratórios forenses, precisa ser complementado devido aos resultados de amplificação insatisfatórios como pode acontecer na investigação de paternidade de restos exumados. Quando se trata de duplas pai-filha, em caso de inclusão, a análise dos marcadores cromossômicos X pode fornecer informações adicionais muito importantes, uma vez que há apenas um alelo que pode ser transmitido do pai para a filha, e LRs mais informativos podem ser esperados do que para os marcadores autossômicos (TILLMAR et al., 2017).

- Casos de maternidade:

Nas investigações de maternidade, para os marcadores X, quando se trata de maternidade, é possível investigar a relação de parentesco por meio de qualquer genitor direto da suposta mãe, para que isso ocorra, a cadeia de transmissão

genética X não deve ser interrompida nos parentescos maternos “avô-neto”, “meio-irmãos” e “avuncular”. Para outros *pedigrees* como “meio-irmãos” e “avô-neto”, a utilidade dos marcadores X é a mesma das contrapartes autossômicas, pois a probabilidade de compartilhar ou não um par de alelos X por consanguinidade também é igualmente provável. Entretanto, deve-se lembrar que no *pedigree* “tio-sobrinho / sobrinha materna”, a probabilidade de compartilhar um par de alelos X por parentesco consanguíneo é ainda menor ($\frac{1}{4}$) do que a obtida considerando os cromossomos autossômicos ($\frac{1}{2}$) (GOMES et al, 2012).

Os testes para mães-filhos utilizando os X-STRs também são eficientes, pois o filho possui apenas um alelo para cada X-STR e deve estar obrigatoriamente presente na mãe, a chance de exclusão neste caso é idêntica aos testes entre pai-filha (SZIBOR et al., 2003).

Ainda é possível utilizar os X-STRs em casos de violência sexual envolvendo o feto do sexo feminino, quando ocorre o aborto e geralmente há mistura de material do embrião com o material da mãe. Neste caso é possível realizar um teste para identificar a paternidade do feto, já que os alelos não compartilhados com a mãe podem ser analisados, excluindo casos de incesto, onde o parente é muito próximo (SZIBOR et al., 2003). Outra forma de aplicar os marcadores do cromossomo X é na análise de vestígios biológicos e ancestralidade (SANTOS-LOPES et al., 2007).

2.4.5 Ferramenta para análise do cromossomo X: FamLink X

Em investigações de parentesco, como investigações de paternidade, podem ser resolvidas utilizando STRs autossômicos que são altamente polimórficos (STRs), que levam a poderosas razões de verossimilhança (LR) (GOMES et al, 2012). Entretanto, outros marcadores foram sendo inseridos para inferir relações de parentesco e, portanto, novas ferramentas de análises além das já conhecidas vêm sendo utilizadas.

A partir do conhecimento adquirido sobre os marcadores do cromossomo X, e à medida que novos marcadores se tornam disponíveis, houve a necessidade de softwares novos que englobam cálculos importantes para atuação forense. Um desses softwares é o FamLinkX, este *software* pode ser encontrado no *site* <http://www.famlink.se>. O FamLink X é um *software* desenvolvido para usuários

forenses à medida que mais marcadores foram disponibilizados, e foi implementado como um novo algoritmo para cálculos de probabilidade responsáveis por ligação, desequilíbrio de ligação e mutações para marcadores do cromossomo X (KLING, 2015).

Existem poucos *softwares* disponíveis para avaliações de parentesco, principalmente levando em conta o padrão de transmissão do cromossomo X, onde há recombinação apenas na mulher. O FamLink X é um dos *softwares* mais relevantes considerando a possibilidade de mutação, *linkage* e também desequilíbrio de ligação, haja vista que alguns dos conjuntos de marcadores mais utilizados são formados por grupos de ligação, portanto estas informações são de extrema relevância para estudos forenses (GOMES, 2020).

Além disso, a sociedade internacional de genética forense - ISFG recentemente forneceu diretrizes gerais para o uso de marcadores cromossômicos X em testes de parentesco (TILLMAR et al., 2017). Sendo assim, é importante destacar a utilidade desta ferramenta para uso forense.

Uma das análises realizadas pelo software é a razão de verossimilhança (LR – *likelihood ratio*) onde duas hipóteses mutuamente excludentes, a primeira hipótese (X) assume que o compartilhamento de alelos entre as partes resulta de um vínculo genético, enquanto a segunda hipótese (Y) assume que o compartilhamento de alelos entre as partes é resultado de eventos aleatórios. Quanto maior é o valor de LR, maior é a evidência em favor da relação do vínculo biológico investigado.

2.6. Justificativa

Os marcadores genéticos existentes no cromossomo X são uma importante ferramenta na análise forense, principalmente em situações que a análise dos STRs autossômicos não é suficiente para a resolução do caso. Por isso, em conjunto com os STRs autossômicos, os marcadores X-STRs podem superar as limitações da análise de marcadores tradicionais, solucionando casos tratados como inconclusivos.

Para ampliar as possibilidades de análises complementares aos STRs, e solucionar casos forenses, estudar a genética de uma população é de extrema importância, pois trata-se do conhecimento do perfil populacional que pode ser único.

Neste estudo, foram utilizados dois conjuntos de marcadores STRs do cromossomo X e foram calculadas as frequências alélicas e identificação do potencial forense na população do estado de São Paulo.

O estudo e a comparação entre os conjuntos de marcadores estudados na população do estado de São Paulo permite o conhecimento dessa população por meio do delineamento da variabilidade genética, e potencial para análise forense e aplicação desses dois conjuntos de marcadores.

A partir dos dados gerados, essas informações são importantes para a construção dos dados dessa população e efetivo uso do cromossomo X como ferramenta forense na identificação de pessoas.

Por fim, espera-se que este trabalho possa contribuir na construção do conhecimento e da importância do uso do cromossomo X na rotina de laboratórios que utilizam a tecnologia do DNA na identificação humana.

3. OBJETIVOS

3.1 Objetivo geral

Caracterizar a população do estado de São Paulo para os conjuntos de marcadores Decaplex (10 X-STRs) e do kit Investigator® Argus X-12QS da Qiagen (12 X-STRs) avaliando a variabilidade genética e a eficiência forense na população estudada.

3.2 Objetivos específicos

Para o estudo da população do estado de São Paulo com dois conjuntos de marcadores do cromossomo X, os objetivos específicos são:

- Determinar as frequências alélicas e haplotípicas destes marcadores para esta população;
- Determinar a variabilidade genética avaliando poder discriminatório dos dois conjuntos de marcadores;
- Calcular o poder de discriminação e o poder de exclusão em duos e trios destes dois painéis de marcadores;
- Avaliar o desequilíbrio de ligação entre os pares de marcadores destes dois painéis;
- Comparar os dois painéis e verificar a eficiência forense de ambos;
- Analisar parentescos utilizando o FamLink X para verificar a razão de verossimilhança (LR – *likelihood ratio*) entre os dois conjuntos de marcadores.

4. MATERIAIS E MÉTODOS

4.1 Aspectos éticos

Este projeto foi aprovado pelo Comitê de Ética em Pesquisa da Faculdade de Ciências Farmacêuticas da UNESP de Araraquara com o código CAAE:12150919.0.0000.5426.

Todos os participantes que aceitaram participar desta pesquisa foram esclarecidos quanto aos preceitos éticos da pesquisa e assinaram o termo de consentimento livre e esclarecido (TCLE).

4.2 Casuística

Para a caracterização genética dos dois conjuntos de marcadores do cromossomo X, foram coletadas amostras sanguíneas de 141 trios (pai, mãe e filha) 423 amostras de indivíduos brasileiros, residentes no estado de São Paulo. A seleção desses indivíduos e a coleta das amostras biológicas foram realizadas pelo laboratório de investigação de paternidade da Faculdade de Ciências Farmacêuticas da Unesp de Araraquara.

4.3 Coleta e extração de DNA

Para a elaboração do perfil genético foi realizada uma punção capilar no dedo anelar dos participantes da pesquisa para a coleta de amostras de sangue em papel FTA® (*Whatman*®) que foram armazenadas a 8°C até o momento da extração do DNA.

A extração do DNA foi realizada conforme adaptação das metodologias propostas por Singer-Sam, Tanguay e Riggs (1989), com auxílio da resina Chelex 100 (Bio-Rad Laboratories). Para a extração do DNA, 1 disco de 1,2mm de papel de filtro FTA® (*Whatman*® GE) foi cortado com o auxílio da caneta Harris Micro Punch® (*Whatman*® GE) com a amostra sanguínea do indivíduo, e colocado 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 1 hora à 56°C. Logo após esta incubação, a solução foi agitada e incubada por mais 8 minutos à 100°C. A solução foi centrifugada a 13.000 rpm por 3 minutos e transferiu-se o sobrenadante para novo microtubo que foi armazenado à 4°C.

4. 4 Reação de amplificação dos X-STRs e preparação para eletroforese capilar

Na reação em cadeia de polimerase PCR foram utilizados 2 protocolos para a amplificação dos 2 painéis de X-STRs, um citado em Gusmão et. al (2009) para o painel de 10 X-STRs – Decaplex (tabela 2), e outro com os 12 X-STRs presentes no kit Investigator® Argus X-12 QS (Qiagen) (tabela 3) seguindo o protocolo do fabricante com adaptações.

Tabela 2- Conjunto de marcadores com 10 X-STRs – Decaplex.

localização citogenética	Marcador	Grupo de ligação	localização física [Mb] NCBI 36	Rutgers Mapa v.2 [cM Kosambi]
22p31	DXS8378	X1	9.330	20,21
22p2	DXS9902		15.234	32,32
centrômero	DXS7132	X2	64.572	90,75
21q31	DXS9898		87.682	101,29
21q33	DXS6809		94.825	108,12
21q33	DXS6789		95.336	108,47
22q3	DXS7133		108.928	118,18
23q	GATA172D05		113.061	124,36
27q1	GATA31E08		140.062	160,54
28q	DXS7423	X4	149.460	184,19

Fonte: modificada de ChrX-STR.org

Tabela 3 - Conjunto de marcadores com 12 X-STRs do kit Investigator® Argus X-12 QS (Qiagen).

localização citogenética	Marcador	Grupo de ligação	localização física [Mb] NCBI 36	Rutgers Mapa v.2 [cM Kosambi]
22p31	DXS10148	X1	9.198	19,84
22p31	DXS10135	X1	9.199	20,03
22p31	DXS8378	X1	9.330	20,21
centrômero	DXS7132	X2	64.572	90,75
12q	DXS10079	X2	66.632	90,82
12q	DXS10074	X2	66.894	90,83
26q2	DXS10103	X3	133.246	149,37
26q2	HPRTB	X3	133.443	149,66
26q3	DXS10101	X3	133.482	149,75
28q	DXS10146	X4	149.335	183,72

q 28	DXS10134	X4	149.401	183,96
q 28	DXS7423	X4	149.460	184,19

Fonte: modificada de ChrX-STR.org

As reações de amplificação foram preparadas com algumas modificações, como segue abaixo.

- Componentes da reação de PCR dos 10 X-STRs – Decaplex (Gusmão, 2009):

Goldstr buffer (Promega) - 5 µl
Primer mix 10X (concentração 2 µM) – 1 µl
Platinum® Taq DNA Polymerase (Termo Fisher) – 0,3 µl (1,5 U)
DNA extraído - 2 µl (~2-5ng/ µl)
Água livre de nuclease qsp. 10 µl.

- Componentes da reação de PCR dos 12 X-STRs – Investigator® Argus X-12 QS (Qiagen):

Fast Reaction Mix 2.0 - 1,5 µl
Primer Mix - 0,5 µl
DNA extraído - 0,5 µl (~ 0,2-2ng/ µl)
Água livre de nuclease qsp. 5 µl

- Condições para a amplificação dos 10 X-STRs (Decaplex) e 12 X-STRs (Kit Investigator® Argus X-12QS) no termociclador Veriti® ThermalCycler (Applied Biosystems® – Thermo Fischer Scientific):

10 X-STRs (Decaplex)

95°C - 15 min
94°C - 30 seg } 10 ciclos
60°C - 90 seg }
72°C - 60 seg }
94°C - 30 seg } 20 ciclos
58°C - 90 seg }
72°C - 60 seg }
60°C – 80 min
4°C - ∞

12 X-STRs (Kit Argus)

98°C - 60 seg }
61°C - 100 seg } 3 ciclos
72°C - 5 seg }
96°C - 10seg } 27 ciclos
61°C - 100 seg }
72°C - 5 seg }
68°C - 2 min
10°C - ∞

Os produtos amplificados com o protocolo do decaplex foram preparados para análise de identificação humana (*Human Identification-HID*) com 8,7 µl *Hi-Di™*

formamida deionizada (*Thermo Fisher Scientific*), 0,3 µl de GeneScan™600 LIZ® (*Thermo Fisher Scientific*) e 1 µl de produto de PCR. Os produtos amplificados com o protocolo do Kit Argus X-12 QS foram preparados com 12 µl *Hi-Di™* Formamida deionizada (*Thermo Fisher Scientific*), 0,5 µl DNA Size Standard 550 (BTO) (Qiagen) e 1 µl de produto de PCR para então serem submetidos à eletroforese capilar pelo analisador genético ABI 3500 *Applied Biosystems®* (*Thermo Fisher Scientific*) utilizando polímero POP-4 da *Applied Biosystems®* (*Thermo Fisher Scientific*) e capilar de 36cm da *Applied Biosystems®* (*Thermo Fisher Scientific*).

4. 5 Análise dos produtos amplificados

Os resultados da eletroforese capilar foram analisados em eletroferogramas no *software Gene Mapper v ID-X-1.2* (*Applied Biosystems® – Thermo Fisher Scientific*). A tipagem ou classificação dos alelos foi efetuada por comparação com o *allelic ladder* fornecido com o kit, no caso dos 12 X-STRs do Argus. O Decaplex como é um conjunto de marcadores *in house*, não possui *allelic ladder*.

4. 6 Análise de dados

Para análise de dados um total de 141 trios pai / mãe / filha foram genotipados para 12 X-STRs com o kit Argus e para os 10 X-STRs do Decaplex. Um total de 141 mulheres não relacionadas foram selecionadas entre as amostras das filhas, e a fase gamética foi determinada usando as informações do haplótipo do pai, sendo no total 141 trios e 282 haplótipos femininos não relacionados.

As frequências alélicas de cada conjunto de marcadores foram calculadas com o programa Arlequin (EXCOFFIER et al., 2010), de acordo com a especificação para cada conjunto de marcador citadas em Gusmão (2009) e no kit Investigator® Argus X (Qiagen), respectivamente.

No conjunto de marcadores Decaplex foi realizada a frequência alélica de cada marcador separadamente. Entretanto, no Investigator® Argus X-12QS os marcadores foram divididos em quatro grupos de ligação (*linkage group*) gerando, assim, uma frequência haplotípica: LG1, LG2, LG3 e LG4.

Todos os parâmetros de análise foram calculados, separadamente, para os 12 X-STRs do kit Argus e para os 10 X-STRs do Decaplex.

4.7 Determinação dos perfis alélicos e análise estatística

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

O *software* Arlequin v3.5.2.2 foi utilizado para determinar as frequências alélicas da população estudada, a heterozigosidade observada e esperada, o teste exato de equilíbrio de Hardy-Weinberg e desequilíbrio de ligação.

Além do arlequin, outras ferramentas foram utilizadas para o cálculo de parâmetros estatísticos de interesse forense como o banco de dados de X-STRs (ChX-STR – <http://www.chrx-str.org/>). Para o cálculo da eficiência forense utilizou-se: o Poder de Discriminação Feminina (PD_F), Poder de Discriminação Masculina (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 (duo – pai/filha), como descrito por Desmarais et al. (1998).

4.8 Análise de parentesco com Famlink X.

O Famlink X foi utilizado para calcular valores da razão de verossimilhança - LR (*likelihood ratio*) em análises de parentesco, quando o suposto pai não está presente, para comparar os valores obtidos dos dois conjuntos de marcadores e ainda compará-los com os valores obtidos com os STRs autossômicos.

A razão de verossimilhança, LR (*likelihood ratio*), é utilizada para demonstrar a força da evidência em casos de paternidade.

A genotipagem de 6 amostras foi realizada seguindo os protocolos já descritos anteriormente com os dois conjuntos de marcadores: o kit Argus X-12QS e o Decaplex.

Os dados obtidos dos STRs autossômicos foram genotipados no laboratório de paternidade da Faculdade de Ciências Farmacêuticas da Unesp de Araraquara, utilizado o kit comercial PowerPlex® Fusion 6C System (Promega Corporation) para comparação dos resultados de LR entre X-STRs e STRs autossômicos.

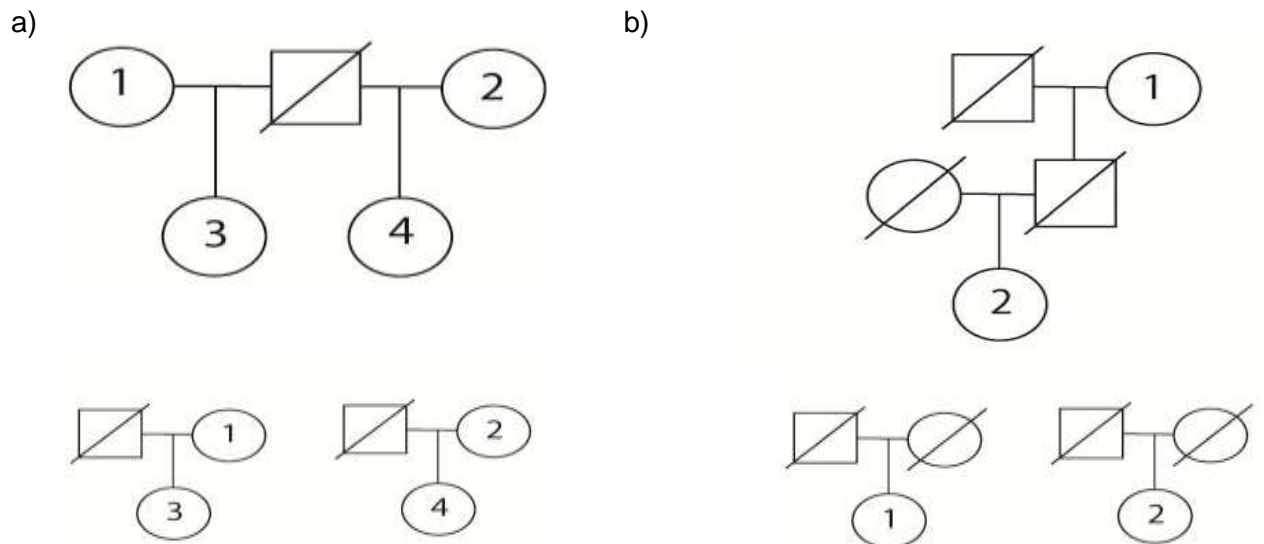
As frequências alélicas utilizadas para esta análise foram as obtidas neste trabalho.

Foram analisados os *pedigrees* de meias – irmãs (paternas) e avó-neta (paternas) - figura 12.

Na análise de meias – irmãs foram utilizados os perfis das “supostas” meias-irmãs além do perfil de suas respectivas mães.

O segundo *pedigree* analisado foi o da “supostas” avó – neta sendo utilizada apenas amostra das duas, ou seja, um parente próximo do suposto pai foi testado.

Figura 12 - *Pedigrees* analisados das supostas meias – irmãs (a) e supostas avó – neta (b).



Fonte: Famlink X *software*

5. RESULTADOS E DISCUSSÃO

5.1 Padronização da utilização do Kit Argus X-12QS

Atualmente vivenciamos tempos onde há a necessidade de economia de reagentes devido aos escassos recursos para a pesquisa. Por isso, foi realizada uma padronização para a diminuição do volume de uso para o kit comercial Argus X-12 QS utilizado neste projeto. Conforme recomendação do fabricante, utiliza-se um volume final de 25µl do *mix* utilizado na reação de PCR (tabela 4).

Tabela 4- Componentes e volume para a realização da PCR.

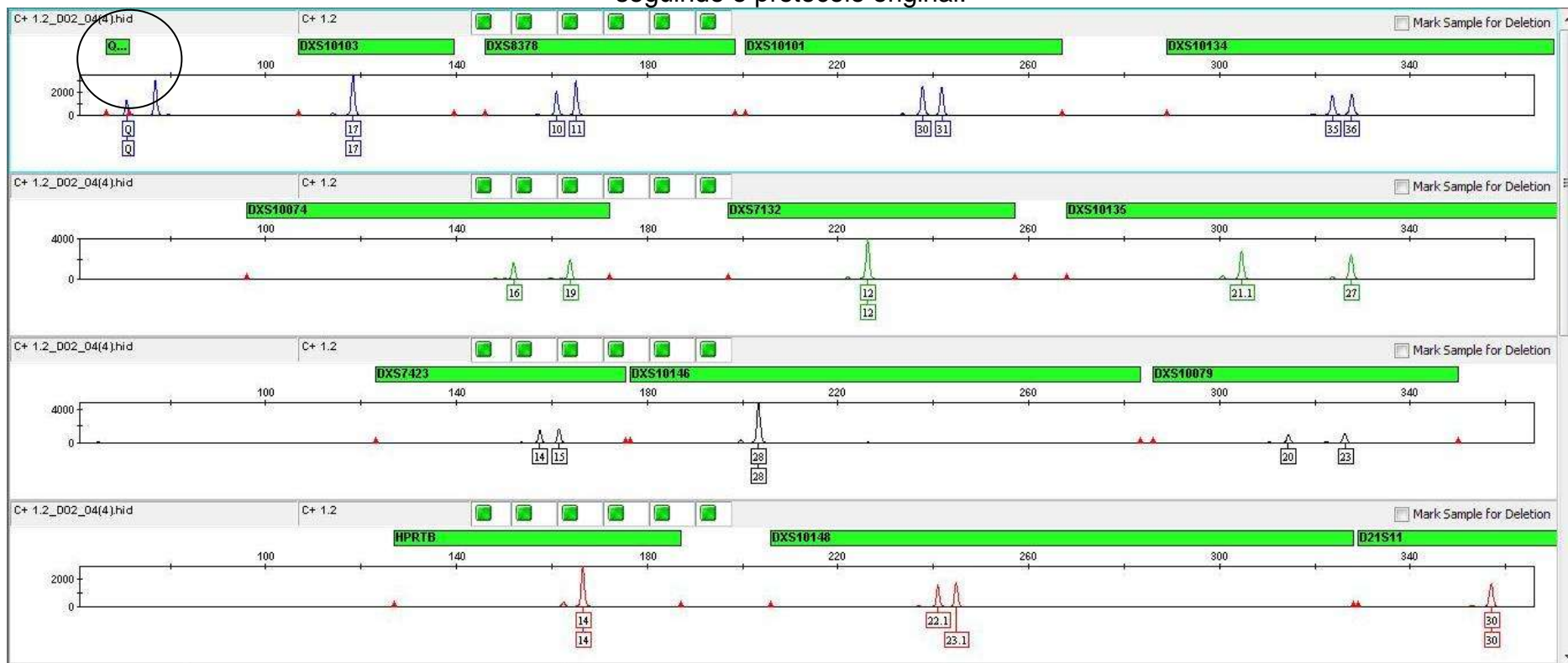
Componentes	Volume por reação	Volume por reação
	original	adaptado
<i>Fast Reaction Mix 2.0</i>	7,5 µl	1,5 µl
<i>Primer Mix</i>	2,5 µl	0,5 µl
Água livre de nuclease	Variável	Variável
DNA extraído	Variável	0,5 µl
Volume total	25 µl	5 µl

Fonte: adaptada do manual do fabricante do kit Investigator® Argus X-12QS(Qiagen)

Para a otimização dos custos na rotina do laboratório, mas sem afetar a especificidade da reação foi testada a redução do volume final da reação para 5µL, mantendo a proporção da concentração recomendada pelo fabricante dos componentes do Kit (*Faster Reaction Mix e Primer Mix*), ou seja, uma redução de 1/5 no volume final do kit comercial. Além da economia feita na utilização do kit, houve a redução na utilização de DNA que passou de 2 µl para 0,5 µl (0,5-2ng/µl). Para comparar os resultados, foram realizadas reações seguindo o protocolo do fabricante e o adaptado, como controle positivo foi utilizado a amostra feminina do DNA comercial 9947A (Qiagen).

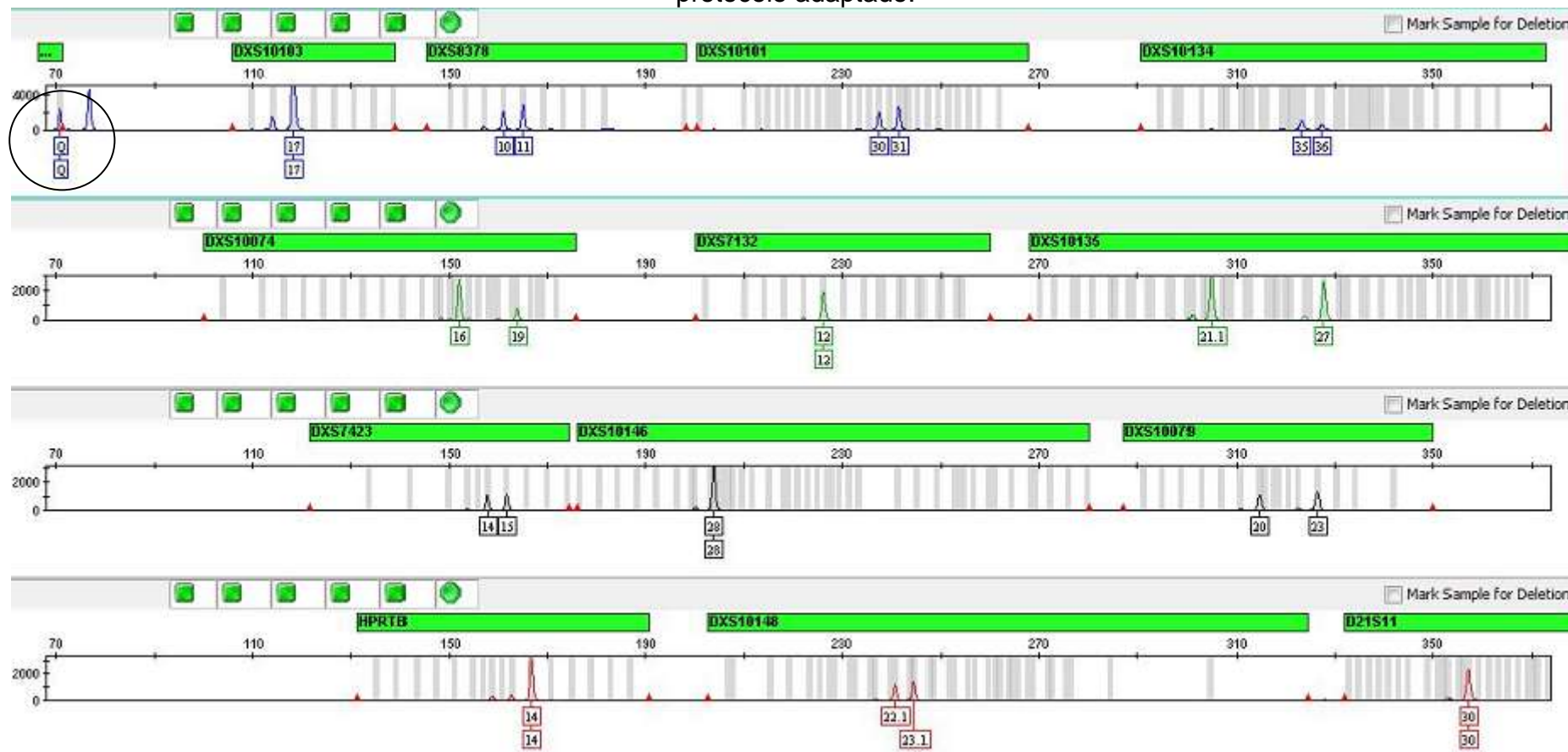
Observando os eletroferogramas do controle positivo (Figura 13) seguindo o protocolo estabelecido no laboratório e o com as adaptações realizadas neste trabalho (Figura 14), foi possível observar que houve adequada amplificação de todos os marcadores nas amostras com volume final de 5µL.

Figura 13 - Eletroferograma do perfil alélico do DNA comercial 9947A obtido com o kit Investigator® Argus X-12QS(Qiagen), seguindo o protocolo original.



Fonte: Este trabalho

Figura 14 - Eletroferograma do perfil alélico do DNA comercial 9947A obtido com o kit Investigator® Argus X-12QS(Qiagen), com o protocolo adaptado.



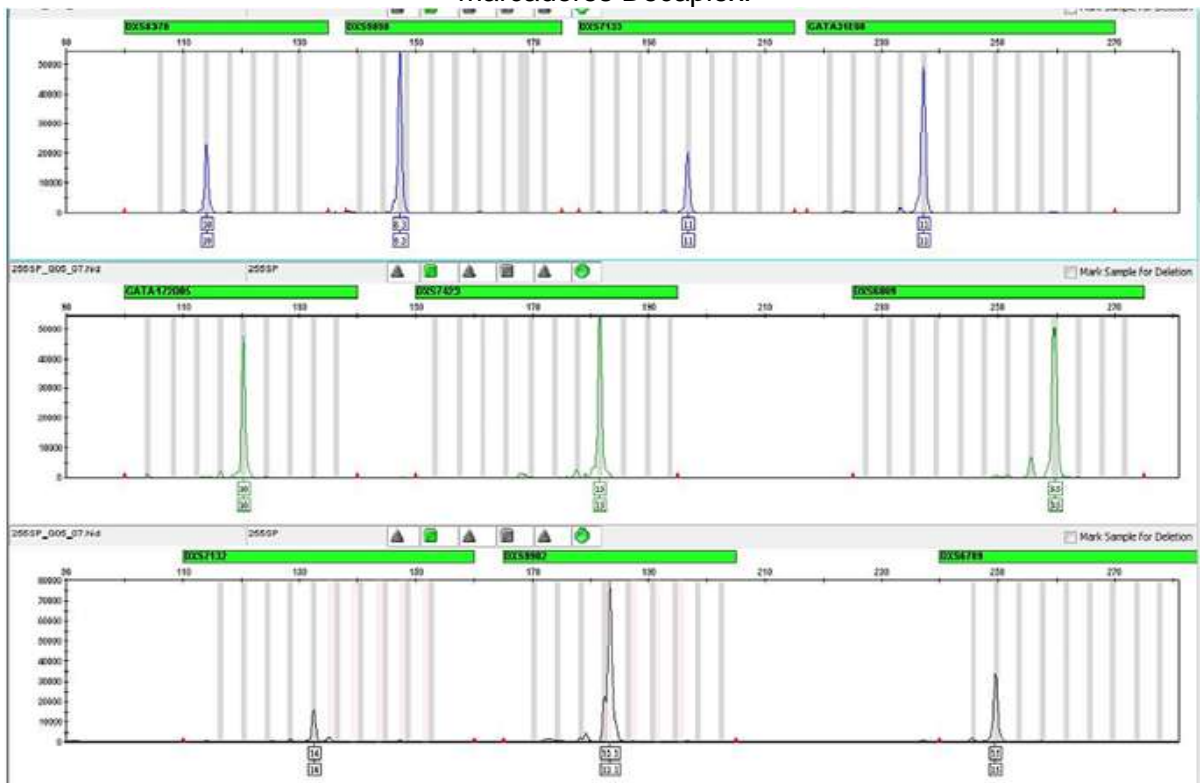
Fonte: Este trabalho

Em relação às adaptações para a utilização de menor volume, é possível observar que as amostras encontram-se preservadas devido à presença do indicador de qualidade presente no Kit Argus, indicado pela letra Q nos eletroferogramas e destacada pelo círculo nas figuras. Outra forma de verificar a qualidade da amplificação das amostras pela unidade de fluorescência relativa (RFU), a qual não houve redução significativa, apesar de ser estabelecido empiricamente o valor de 200 RFUs como o valor a partir do qual é possível identificar e distinguir um pico para que seja designado como alelo.

5.2 Avaliação da metodologia

A metodologia mostrou-se eficaz como pode-se observar pelos eletroferogramas dos perfis alélicos abaixo obtidos no analisador genético ABI3500 (figuras 15 e 16).

Figura 15- Eletroferograma do perfil alélico de amostra masculina obtido com painel de marcadores Decaplex.

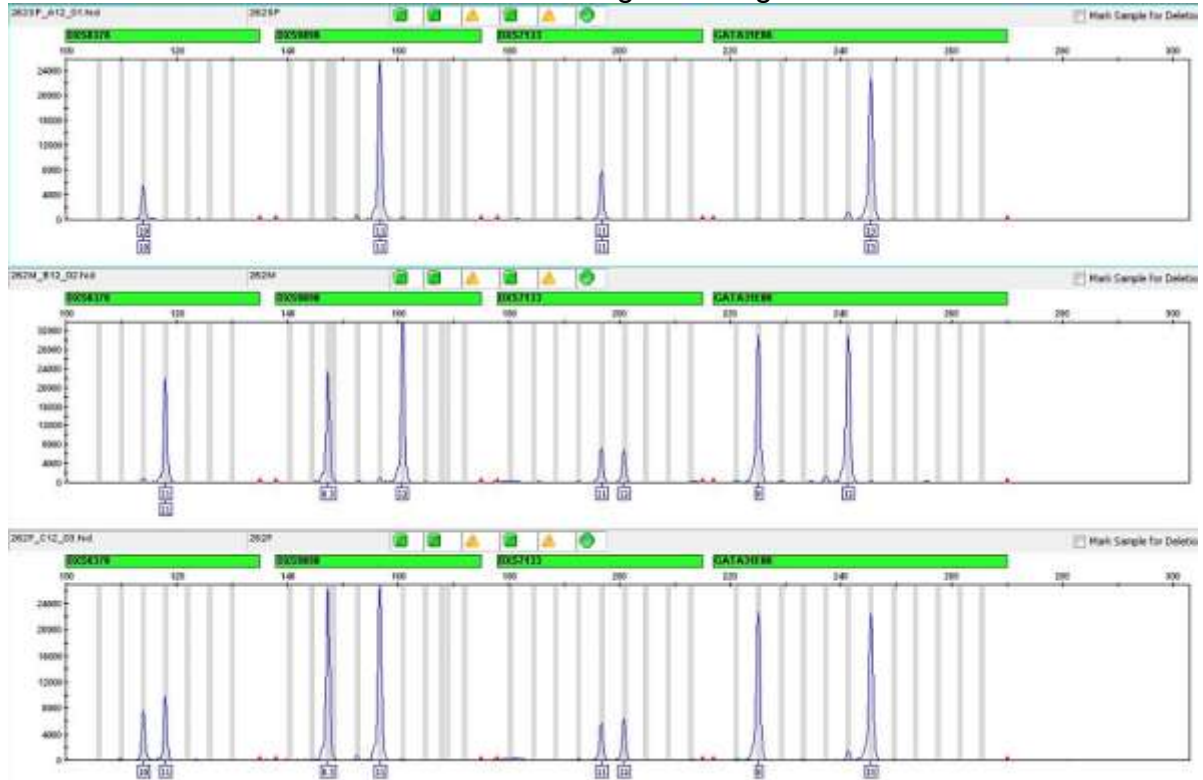


Fonte: Este trabalho

Na figura 15 é possível observar o perfil alélico de uma amostra masculina onde os 10 marcadores (Decaplex) foram amplificados e na figura 16 um perfil

alélico de uma amostra feminina amplificada com os 12 X-STRs do kit Investigator Argus X 12 QS.

Figura 16 - Eletroferograma do perfil alélico de uma mostra feminina com os marcadores do kit Investigator® Argus X 12 QS.



Fonte: Este trabalho

5.3 Análise genética da população

5.3.1 Variabilidade genética na população de São Paulo com os conjuntos de marcadores

Neste estudo, as amostras de 141 mulheres não relacionadas foram selecionadas para análises populacionais. Ao selecionar as filhas dos trios estudados, foi possível determinar a fase gamética do genótipo e criar um banco de dados que inclui 282 haplótipos não relacionados. Estes haplótipos foram utilizados para a análise da frequência dos 2 conjuntos de marcadores.

5.3.2 Análise do conjunto de marcadores com 10 X-STRS – Decaplex

A frequência alélica do Decaplex foi calculada para os 10 marcadores X-STRs. A tabela 5 ilustra as distribuições das frequências alélicas da população estudada, bem como os valores da diversidade gênica de cada marcador.

Tabela 5 - Frequência alélica dos 10 X-STRS na população do estado de São Paulo.

Marcador	DXS8378	Marcador	DXS9902	Marcador	DXS7132	Marcador	DXS9898	Marcador	DXS6809
alelo	frequência	alelo	frequência	alelo	frequência	alelo	frequência	alelo	frequência
9	0,01064	8	0,00355	11	0,00355	7	0,01064	28	0,01418
10	0,34397	9	0,02128	12	0,10638	8.3	0,1383	29	0,01064
11	0,31915	10	0,0461	13	0,2766	9	0,00709	29.1	0,00355
12	0,28723	11	0,3617	14	0,31915	10	0,05674	30	0,05674
13	0,03191	11.1	0,00355	14.3	0,00355	11	0,18794	31	0,17021
14	0,00709	12	0,3227	15	0,1844	12	0,29078	32	0,11702
		12.1	0,01773	15.3	0,01064	13	0,20922	33	0,29433
		13	0,21277	16	0,05319	13.3	0,00709	33.1	0,00355
		13.1	0,00355	16.3	0,00709	14	0,08511	34	0,19858
		14	0,00709	17	0,02837	15	0,00709	35	0,08511
				17.3	0,00709			36	0,03901
								37	0,00709
n de alelos	6	n de alelos	10	n de alelos	11	n de alelos	10	n de alelos	12
n de cópias	282	n de cópias	282	n de cópias	282	n de cópias	282	n de cópias	282
DG	0,69862	DG	0,71934	DG	0,77519	DG	0,80937	DG	0,82181
Marcador	DXS6789	Marcador	DXS7133	Marcador	GATA172D05	Marcador	GATA31E08	Marcador	DXS7423
alelo	frequência	alelo	frequência	alelo	frequência	alelo	frequência	alelo	frequência
15	0,07092	8	0,01773	6	0,16312	7	0,01418	12	0,00709
16	0,06383	9	0,33688	7	0,03191	8	0,01418	13	0,03901
17	0,00709	10	0,14894	8	0,17021	9	0,19504	14	0,38298
18	0,00355	11	0,43262	9	0,17021	10	0,05319	15	0,40426
19	0,0461	12	0,0461	10	0,23404	11	0,17021	16	0,14184
20	0,39007	13	0,00709	11	0,16667	12	0,25887	17	0,02482
21	0,20213	14	0,01064	12	0,06028	13	0,20922		
22	0,15957			13	0,00355	14	0,07447		
23	0,02482					15	0,01064		
24	0,03191								
n de alelos	10	n de alelos	7	n de alelos	8	n de alelos	9	n de alelos	6
n de cópias	282	n de cópias	282	n de cópias	282	n de cópias	282	n de cópias	282
DG	0,77133	DG	0,67696	DG	0,83118	DG	0,81621	DG	0,66997

* DG - diversidade gênica; em verde, maiores valores de DG e em vermelho, menores valores de DG.

Fonte: Este trabalho

Dentro da população de São Paulo (tabela 5) foi possível observar em cada marcador de 6 a 12 alelos com um total de 89 alelos. Os dois marcadores com a maior diversidade gênica encontram-se destacados em verde na tabela, sendo o GATA172D05 aquele que apresentou maior diversidade, de 0,83118 e DXS6809, com a diversidade de 0,82181, além de ser o marcador com maior variabilidade alélica (12 alelos).

Os dois marcadores com menor diversidade gênica estão destacados em vermelho e são DXS7133 com 0,67696 e DXS7423 com 0,66997e possuem a menor variabilidade alélica (6 alelos).

Na tabela 6 é possível observar que estes dados corroboram com estudos realizados anteriormente com a população de São Paulo, onde os marcadores GATA172D05, DXS6809 e DXS6789 apresentaram diversidade gênica de 0,81 (MARTINS, 2010); além de outras populações de regiões isoladas da Espanha e norte de Portugal com os marcadores GATA172D05 e DXS6809 apresentando 0,81 e 0,80 com maior diversidade gênica e DXS7133 com menor diversidade gênica, 0,65 (ZARRABEITIA et. al., 2009); a população do estado do Paraná com o marcador GATA172D05 0,82 e DXS7133 com 0,66 (KOBACHUK et. al., 2012); na população do Mato Grosso o marcador com maior diversidade gênica foi o GATA172D05 que apresentou o valor de 0,85; e comunidade foral de Navarra na Espanha com os marcadores GATA172D05 com 0,80 e DXS7133 com 0,60 (ILLESCAS et al., 2012), na qual os marcadores citados apresentaram maior e menor diversidade gênica, respectivamente. Já para a população da Península Ibérica e América latina, o marcador DXS6809 com maior diversidade gênica apresentou um índice de 0,82 e o marcador DXS7133 apresentou a menor diversidade gênica com 0,66 (GUSMÃO et al., 2009).

Tabela 6- Comparação entre maiores diversidades gênicas deste e outros trabalhos com os 10 X-STRs do decaplex.

Local - Fonte	DG - Marcador	DG - Marcador
	GATA172D05	DXS6809
SP (Este trabalho)	0,83	0,82
SP (MARTINS et al., 2010)	0,81	0,81
MT (MARTINS et al., 2017)	0,85	-
Paraná (KOBACHUK, 2012)	0,82	-
Espanha e norte de Portugal (ZARRABEITIA et al., 2009)	0,81	0,80
Navarra (ILLESCAS et al.,2012)	0,82	-
Península Ibérica e América Latina (GUSMÃO et al., 2009)	-	0,82

Fonte: Este trabalho

Nas populações comparadas os dois marcadores, DXS6809 e GATA172D05 variaram em apresentar as maiores diversidades gênicas, ainda assim, na maioria das populações comparadas os índices foram maiores para o marcador GATA172D05, inclusive neste trabalho.

5.3.3 Análise do conjunto de marcadores com 12 X-STRS do Kit Argus X-12 QS

A frequência haplotípica para os 12 X-STRs do kit Argus foi calculada a partir dos haplótipos formados pelos 4 grupos de ligação (*linkage group*): LG1, LG2, LG3 e LG4 (apêndice A).

O grupo de ligação que apresentou a maior diversidade gênica e variabilidade alélica foi o LG1.

O marcador DXS10135 foi o que se mostrou com maior valor de diversidade gênica 0,94624, destacado em verde na tabela 7 e 8, e o marcador com a maior variabilidade alélica foi o DXS10148.

O grupo de ligação que possui o marcador com a menor diversidade gênica (destacado em vermelho na tabela 7), é o LG4 no marcador DXS7423 apresentando diversidade de 0,66997

Tabela 7 - Diversidade gênica (DG) dos marcadores e variabilidade alélica de cada grupo de ligação baseada na frequência dos 12-XSTRs do Kit Argus na população do estado de São Paulo.

LG1			LG2		
Marcador	DG	n alelos	Marcador	DG	n alelos
DXS10148	0,92845	36	DXS 7132	0,77519	11
DXS10135	0,94624	29	DXS10079	0,81858	11
DXS 8378	0,69862	6	DXS10074	0,88486	15
Total	0,85777	71	Total	0,82621	37
LG3			LG4		
Marcador	DG	n alelos	Marcador	DG	n alelos
DXS10103	0,78221	7	DXS10146	0,92537	30
HPRTB	0,75821	9	DXS10134	0,89246	22
DXS10101	0,91661	19	DXS 7423	0,66997	6
Total	0,81901	35	Total	0,82927	58

* DG - diversidade gênica; em verde, maior valor de DG em vermelho, menor valor de DG.

Fonte: Este trabalho.

Estes dados obtidos estão de acordo com outros estudos de populações como da Argentina (GARCIA et al., 2019); dos Emirados Árabes Unidos (ALMARRI

et al., 2017), do oeste do Mediterrâneo (FERRAGUT et al., 2015), da Groênlandia, Dinamarca e Somália (TOMAS et al, 2012), e Croácia (MRŠIĆ et. al, 2018) como mostra a tabela 8.

Tabela 8 - Comparação da diversidade gênica deste trabalho com outras pesquisas com os 12 X-STRs do Argus.

Local - fonte	Marcador DXS10135
São Paulo - este trabalho	0,946
Argentina (GARCIA et al.,2019);	0,94
Emirados Árabes Unidos (ALMARRI et al.,2017)	0,946
Oeste do Mediterrâneo (FERRAGUT et al., 2015)	0,95
Groênlandia, Dinamarca e Somália (TOMAS et al.,2012)	0,946
Croácia (MRŠIĆ et al.,2018)	0,929

Fonte: Este trabalho

O grupo de ligação mais informativo foi o LG1 com o maior valor de diversidade haplotípica 0,998359456 e 236 haplótipos diferentes, além de apresentar o maior número de haplótipos únicos (tabela 9). A proporção de haplótipos únicos mostrou-se extremamente alta quando comparada com outros trabalhos, como o da população argentina (GARCIA et. al, 2019), provavelmente está relacionado à alta taxa miscigenação da população brasileira.

O grupo de ligação com menor diversidade haplotípica foi o LG3 0,994472628 e 156 haplótipos, porém o haplótipo mais comum foi o do grupo de ligação LG2, aparecendo 8 vezes, sendo o haplótipo com os alelos 18 (DXS7132), 19 (DXS10079) e 10 (DXS10074) como é possível observar na tabela 9.

Tabela 9 - Distribuição dos haplótipos dos 12 X-STRs para os grupos de ligação em todas as amostras.

Grupo de ligação	Número de haplótipos diferentes	Número de haplótipos únicos	Frequência do haplótipo mais frequente %	Diversidade haplotípica
1	236	203 (86,01%)	1,77	0,998359456
2	174	114 (65,51%)	2,83	0,995078368
3	156	87 (55,76%)	2,12	0,994472628
4	210	162 (77,14%)	2,12	0,996600174

Fonte: Este trabalho

5.3.4 Comparação dos dois conjuntos de marcadores 10 X-STRs- Decaplex e Kit Investigator® Argus X-12 QS.

Quando comparados os dois conjuntos de marcadores, o Decaplex com 10 X-STRs e o Kit Investigator® Argus com 12 X-STRs, verificou-se que este último apresentou melhores índices de eficiência de discriminação, provavelmente pelos maiores índices de diversidade gênica, possuindo marcadores mais polimórficos, além da variabilidade alélica ser maior. Além disso, deve ser considerado que os 12 X-STRs ainda possui o maior número de marcadores (tabela 10).

Tabela 10 - Diversidade gênica e variabilidade alélica dos 2 conjuntos de marcadores 10 X-STRS – Decaplex e kit Investigator® Argus com 12 X-STRs.

Decaplex			Argus		
Marcador	10 X-STRs n de alelos	DG	Marcador	12 X-STRs n de alelos	DG
DXS8378	6	0,69862	LG1-DXS10148	36	0,92845
DXS9902	10	0,71934	LG1-DxS10135	29	0,94624
DXS7132	11	0,77519	LG1-DXS 8378	6	0,69862
DXS9898	10	0,80937	LG2-DXS 7132	11	0,77519
DXS6809	12	0,82181	LG2-DXS10079	11	0,81858
DXS6789	10	0,77133	LG2-DXS10074	15	0,88486
DXS7133	7	0,67696	LG3-DXS10103	7	0,78221
GATA172D05	8	0,83118	LG3-HPRTB	9	0,75821
GATA31E08	9	0,81621	LG3-DXS10101	19	0,91661
DXS7423	6	0,66997	LG4-DXS10146	30	0,92537
			LG4-DXS10134	22	0,89246
			LG4-DXS 7423	6	0,66997

* DG - diversidade gênica; LG - *linkage group*; em verde, maiores valores de DG e em vermelho, menores valores de DG

Fonte: Este trabalho

Os dois conjuntos de marcadores possuem características e análise distintas, enquanto o decaplex possui a avaliação baseada nos marcadores independentes, o conjunto de marcadores do Kit Argus possui marcadores ligados e os parâmetros de avaliação um pouco diferentes, pois avalia a diversidade haplotípica dos marcadores que estão ligados, entretanto os dois conjuntos de marcadores conseguiram atingir altos índices de diversidade gênica e haplotípica.

5.4 Parâmetros estatísticos de avaliação da eficiência forense

Para aplicação forense de marcadores microsatélites, existe a necessidade destes marcadores possuírem propriedades genéticas existentes na população estudada que sejam adequadas, tais como, alto nível de polimorfismo, desequilíbrio de ligação conhecido, equilíbrio de Hardy-Weinberg, entre outras (SZIBOR, 2003). Por isso são aplicados os parâmetros estatísticos de avaliação forense.

5.4.1 Análise de equilíbrio de Hardy-Weinberg

O princípio de Hardy-Weinberg (HWE) é fundamental na genética de população, e é relevante em muitas áreas relacionadas das ciências da vida, epidemiologia, bioinformática e bioestatística. Na ausência de forças evolutivas como mutação, migração ou seleção, e quando os cruzamentos são aleatórios e o fluxo gênico não seja capaz de alterar a composição alélica de uma população, assim sendo, com as frequências alélicas permanecendo inalteradas ao longo das gerações atingindo proporções genotípicas estáveis, diz-se que a população está em equilíbrio de Hardy-Weinberg (GRAFFELMAN et al., 2016).

Os dados gerados das amostras femininas estudadas neste trabalho foram utilizados no teste exato para o equilíbrio de Hardy-Weinberg.

Nos dois conjuntos de marcadores analisados neste trabalho não foram detectados desvios estatisticamente significativos das expectativas de equilíbrio de Hardy-Weinberg para ($p \leq 0,05$); diante disso, entende-se que as frequências genotípicas podem ser consideradas representativas indicando não haver sinais de estratificação populacional. Além de não ser observado em nenhum dos conjuntos de marcadores excesso de homozigotos.

Foi realizada a correção de Bonferroni (1936) para testes múltiplos e também não foram encontrados desvios significativos para os genótipos femininos da população estudada (tabelas 11) com níveis de significância (p) de 0,005 (10 X-STRs – Decaplex) e de 0,00416667 (12 X-STRs – Kit Argus).

Tabela 11 - Parâmetros estatísticos da diversidade genética para os conjuntos de 10 marcadores X-STRs e 12 X-STRs na população feminina de São Paulo. Tamanho da amostra: 141.

10 X-STRs				12 X-STRs			
Locus	Ho	He	P-HWE	Locus	Ho	He	P-HWE
DXS8378	0,70922	0,69862	0,10209	DXS10103	0,80851	0,78221	0,43152
DXS6789	0,75887	0,77133	0,89622	DXS8378	0,70922	0,69862	0,10467
DXS9898	0,78723	0,80937	0,79077	DXS7132	0,79433	0,77519	0,09382
DXS7133	0,68085	0,67696	0,32774	DXS10134	0,87943	0,89246	0,11424
GATA31E08	0,78014	0,81621	0,25013	DXS10074	0,83688	0,88486	0,63896
GATA172D05	0,82979	0,83118	0,51990	DXS10101	0,86525	0,91661	0,46864
DXS7423	0,66667	0,66997	0,65111	DXS10135	0,92908	0,94624	0,06169
DXS6809	0,78014	0,82181	0,26266	DXS7423	0,66667	0,66997	0,64906
DXS7132	0,79433	0,77519	0,09549	DXS10146	0,90780	0,92537	0,31097
DXS9902	0,68085	0,71934	0,53877	DXS10079	0,85106	0,81858	0,70950
				HPRTB	0,77305	0,75821	0,98951
				DXS10148	0,94326	0,92845	0,07789

Ho, heterozigosidade observada; He, heterozigosidade esperada; P-HWE, valor de p para o equilíbrio de Hardy-Weinberg, com nível de significância de 0,005 para os 10 XSTRs e 0,00416667 para os 12 X-STRs (após correção de Bonferroni para múltiplos testes).

Fonte: Este trabalho

Apesar dos resultados não mostrarem nenhum desequilíbrio na população estudada, em outras populações brasileiras avaliadas com o conjunto de marcadores 12 X-STRs do Kit Argus foi observado um afastamento significativo do HWE no *locus* DXS10134 (PINTO et al., 2020). No entanto, o desvio observado não foi significativo ao analisar separadamente as três populações do Brasil (população geral), Rio de Janeiro e de São Paulo. De acordo com Pinto (2020) esses resultados indicam que para a utilização de um banco de dados em comum são necessários estudos maiores sobre populações brasileiras.

No conjunto de marcadores 10 X-STRs depois da correção de Bonferroni outras populações brasileiras como as do Paraná e Mato Grosso (KOBACHUK et. al., 2012; MARTINS et. al., 2017), também não mostraram desvios significativos, além de outros estudos com populações da América Latina e Península Ibérica (GUSMÃO, et. al, 2009); onde apenas na população do Rio de Janeiro foi observado um desvio nos *locus* DXS7423 e a da Galícia, no *locus* DXS6809.

Na análise dos parâmetros estatísticos forenses, nas tabelas 12 e 13 abaixo, os 2 conjuntos de marcadores provaram ser altamente discriminativos na população de São Paulo.

Tabela 12 - Parâmetros estatísticos forense para os 10 marcadores X-STRs (Decaplex) na população de São Paulo. Número de cromossomos analisados: 282.

Marcador	PD _F	PD _M	MEC _T	MEC _D
DXS8378	0,846525	0,696145	0,634998	0,489226
DXS6789	0,918425	0,768600	0,74057	0,609635
DXS9898	0,93587	0,806517	0,779823	0,656861
DXS7133	0,83659	0,674566	0,617063	0,471741
GATA31E08	0,95635	0,839453	0,821579	0,713553
GATA172D05	0,947164	0,828231	0,804899	0,688365
DXS7423	0,827641	0,667594	0,605728	0,460355
DXS6809	0,944569	0,81892	0,796279	0,678558
DXS7132	0,913947	0,772426	0,738163	0,606183
DXS9902	0,869581	0,716774	0,666572	0,524222
Total	0,999999999968	0,99999953	0,99999805	0,999895138

PD_F, poder de discriminação feminino; PD_M, poder de discriminação masculino; MEC_T, chance média de exclusão em trios envolvendo filha, MEC_D, chance média de exclusão em duo pai/filha.

Fonte: Este trabalho

Tabela 13 - Parâmetros estatísticos forense para os 12 marcadores X-STRs (Kit Argus) na população de São Paulo. Número de cromossomos analisados: 282.

Grupo de ligação	PD _F	PD _M	MEC _T	MEC _D
LG1	0,999946691	0,994819464	0,994792993	0,989675847
LG2	0,999858943	0,991550137	0,991480481	0,983205683
LG3	0,999975241	0,99647358	0,996461256	0,992966223
LG4	0,999975069	0,996463788	0,996451361	0,992943532
Total	>0,99999999999	0,99999999945	0,99999999944	0,9999999913

PD_F, poder de discriminação feminino; PD_M, poder de discriminação masculino; MEC_T, chance média de exclusão em trios envolvendo filha, MEC_D, chance média de exclusão em duo pai/filha.

Fonte: Este trabalho

Na avaliação das populações estudadas, os parâmetros testados demonstraram valores altos, principalmente o poder de discriminação feminino (PD_F) nos 12 X-STRs do kit Investigator® Argus X-12 (tabela 13).

O PD_F dos 10 X-STRs do Decaplex (tabela 12) também mostrou-se eficiente alcançando alto índice discriminativo sendo o GATA172D05 e o DXS6809 com os maiores índices, o que demonstra a aplicabilidade destes painéis para a identificação e análise de parentesco como descrito em outras populações

brasileiras (GUSMÃO et al., 2009; MARTINS et al., 2010, CAINÉ et al., 2010; CAINÉ et al., 2011; PINTO et al., 2020).

Em estudos com os 12 X-STRs na população brasileira comparada com outras populações latinas, este painel de marcadores apresentou os maiores valores de chance média de exclusão e capacidade de discriminação, seguida pela Argentina e Ibéria, com resultados semelhantes (PINTO et al., 2020).

No painel de marcadores com os 10 X-STRs - Decaplex, foi observado também maiores valores de chance média de exclusão na população brasileira, assim como descrito em Martins et al. (2010), onde essas populações foram comparadas com populações latinas e ibéricas, contudo esses parâmetros foram mais altos em alguns marcadores em comum no estudo das populações de Angola, Moçambique e Uganda (GOMES et al., 2007).

5.4.2 Desequilíbrio de ligação entre os marcadores X-STRs

O desequilíbrio de ligação pode ser definido como uma associação não aleatória de alelos e diferentes marcadores em nível populacional (SLATKIN, 2008). Como a recombinação do cromossomo X ocorre apenas em mulheres, alelos presentes nos marcadores geralmente localizados neste cromossomo, deve exibir um maior grau de desequilíbrio de ligação que marcadores autossômicos. Outro fator que contribui para um maior desequilíbrio de ligação nos marcadores do cromossomo X é a baixa taxa de mutação na linhagem feminina que se aplica a 2/3 dos cromossomos X presentes em uma população, já que a mulher possui dois cromossomos X e o homem apenas um (TILLMAR et al., 2017).

No presente trabalho, a análise do desequilíbrio de ligação foi realizada para os 2 conjuntos de marcadores (figuras 17 e 18).

Associações entre marcadores do mesmo grupo de ligação foram detectadas em no conjunto de marcadores com 12 X-STRs (kit Argus) após a correção de Bonferroni, o que é esperado entre os grupos de ligação, mas a falta de associação entre marcadores dentro do mesmo grupo de ligação pode indicar uma baixa quantidade de amostras (PINTO et al., 2020). No presente estudo foi encontrada associação entre os marcadores DXS10134 – DXS10135 que são de diferentes grupos de ligação, outros estudos também encontraram associações entre

marcadores de diferentes grupos de ligação (GOMES et al., 2017; ROBINO et al., 2018) na população da Sardenha e da Guiné-Bissau.

No conjunto de 10 X-STRs (decaplex) não foi detectada nenhuma associação entre os marcadores. De acordo com Gusmão et al. (2009), a ausência de desequilíbrio de ligação indica que não há subestruturação genética na população estudada principalmente nas populações miscigenadas e está de acordo com a distância entre os *loci* dos 10 X-STRs que forma o conjunto de marcadores do Decaplex.

Figura 17- Valores de p obtidos no pareamento do teste de desequilíbrio de ligação para todos os pares de *loci* dos 10 X-STRs (Decaplex).

	DXS8378	DXS9902	DXS7132	DXS9898	DXS6809	DXS6789	DXS7133	GATA172D05	GATA31E08	DXS7423
DXS8378	*									
DXS9902	0,1710	*								
DXS7132	0,4990	0,4610	*							
DXS9898	0,7310	0,5730	0,5010	*						
DXS6809	0,1050	0,5010	0,2910	0,9540	*					
DXS6789	0,8790	0,8660	0,4800	0,0310	0,1750	*				
DXS7133	0,8410	0,0670	0,9510	0,6280	0,1270	0,0250	*			
GATA172D05	0,5970	0,8810	0,3290	0,8150	0,3500	0,1590	0,0180	*		
GATA31E08	0,0980	0,5230	0,0180	0,9390	0,1470	0,5050	0,3250	0,9910	*	
DXS7423	0,0060	0,1230	0,4430	0,2270	0,9000	0,5850	0,1020	0,2620	0,3860	*

Teste exato do desequilíbrio de ligação usando *Markov chain* (*Chain length*:100.000; *Dememorization*:10.000; 100172 *Steps done*). Valor de p estatisticamente significativo para o teste de desequilíbrio de ligação, com nível de significância de 0,0011 após correção de Bonferroni para múltiplos testes

Fonte: Este trabalho

Figura 18 – Valores de p obtidos no pareamento do teste de desequilíbrio de ligação para todos os pares de *loci* 12 X-STRs (Argus).

	LG1			LG2			LG3			LG4		
	DXS10148	DXS10135	DXS8378	DXS7132	DXS10079	DXS10074	DXS10103	HPRTB	DXS10101	DXS10146	DXS10134	DXS7423
DXS10148	*											
DXS10135	0,3843	*										
DXS8378	0,6622	0,0265	*									
DXS7132	0,5213	0,9932	0,4172	*								
DXS10079	0,5610	0,1846	0,0537	0,2839	*							
DXS10074	0,7649	0,5650	0,8312	0,3276	0,2506	*						
DXS10103	0,2660	0,9645	0,2534	0,2945	0,7001	0,8396	*					
HPRTB	0,0229	0,5579	0,8161	0,3321	0,5265	0,2628	0,0362	*				
DXS10101	0,8305	0,3011	0,2346	0,0216	0,0683	0,2769	0,0001	0,0000	*			
DXS10146	0,8360	0,3411	0,9407	0,8411	0,1891	0,7126	0,5429	0,1190	0,2472	*		
DXS10134	0,6437	0,0000	0,8456	0,2129	0,1152	0,2168	0,9516	0,6857	0,2038	0,0000	*	
DXS7423	0,8048	0,7868	0,0114	0,3776	0,0933	0,0186	0,5873	0,2439	0,4114	0,0012	0,0430	*

Teste exato do desequilíbrio de ligação usando *Markov chain* (*Chain length*:100.000; *Dememorization*:10.000; 100172 *Steps done*). Em negrito está os valores p significativos para um nível de significância após a aplicação da correção de Bonferroni; $p \leq 0,00075$ para 66 comparações múltiplas

Fonte: Este trabalho

5.5 Aplicação de X-STRs na análise de parentesco utilizando o software Famlink X

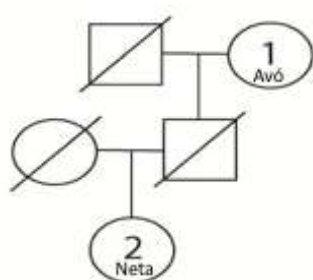
O famlink X foi utilizado para análises de parentesco entre “supostas” meias-irmãs (paternas) e avó (paterna) - neta sem o suposto pai presente, utilizando os dois conjuntos de marcadores deste trabalho (Apêndice B).

Os valores de LR (*likelihood ratio*) ou razão de verossimilhança indicam que os marcadores são eficientes para este tipo de análise forense e que os 12 marcadores do kit Argus X-12 QS obtiveram uma leve vantagem em relação aos valores de LR obtidos com o conjunto de marcadores Decaplex.

Devido ao modo de herança especial do cromossomo X, o X-STR geralmente apresenta algumas vantagens nestes tipos de análises, quando comparado com os autossômicos, na figura 19 é possível observar principalmente na análise de avó-neta, o quanto o cromossomo X elevou o nível da análise para que a mesma não se convertesse em inconclusiva apenas com os valores dos STRs autossômicos.

Portanto, para as situações em que os marcadores do cromossomo X têm o poder de excluir, valores elevados de LR são obtidos (GOMES et al., 2012).

Figura 19 - Pedigree analisado de “supostas” avó (1) – neta (2) com os dois conjuntos de marcadores e os STRs autossômicos.



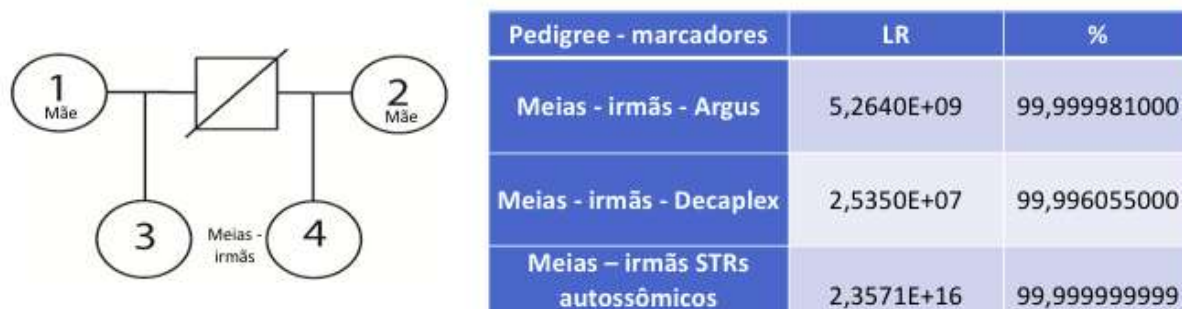
Pedigree - marcadores	LR	%
Avó - neta - Argus	3,2160E+09	99,999969000
Avó - neta - Decaplex	7,3840E+03	99,986458000
Avó - neta - STRs autossômicos	1,2871E+00	56,276405129

Fonte: Este trabalho

Novamente analisando sob a perspectiva do padrão de herança do cromossomo X, no caso das duas supostas meias-irmãs (figura 20) elas devem compartilhar os alelos cromossômicos X paternos idênticos por descendência, o que se não acontecesse excluiria a paternidade (GOMES et al., 2012). Neste caso os valores de LR foram significativos para o cromossomo X, dando destaque para os

valores do conjunto de 12 marcadores do Argus e levando em consideração que a análise incluiu as respectivas mães.

Figura 20 - Pedigree analisado de “supostas” meias – irmãs (3 e 4) e respectivas mães (1 e 2) com os dois conjuntos de marcadores e os STRs autossômicos.



Fonte: Este trabalho

Os dois conjuntos de marcadores mostraram-se eficientes com valores de LR altos, portanto, significativos, principalmente para o conjunto de 12 marcadores do Argus.

Quando utilizados em reconstruções genéticas para análise de parentesco esses valores de LR combinam-se com valores obtidos de LR obtidos dos STRs dos cromossomos autossômicos como recomenda a ISFG (TILLMAR et al., 2017) validando, portanto, a importância do cromossomo X neste tipo de análise.

6. CONCLUSÕES

As frequências alélicas e haplotípicas foram determinadas e os dois conjuntos de marcadores mostraram-se polimórficos para a população do estado de São Paulo indicando que são bastante informativos.

Os dois conjuntos de marcadores foram discriminativos e apresentaram altos índices de diversidade gênica, além de marcadores com alta taxa de variabilidade alélica, principalmente os 12 X-STRs do Argus.

As análises estatísticas para avaliar a eficiência forense dos dois conjuntos de marcadores mostraram que os dois conjuntos utilizados foram bastante discriminativos na população estudada.

Apesar da associação entre marcadores de grupos de ligação diferentes na análise dos 12-XSTRs do kit Argus, DXS10134 – DXS10135, este resultado indica apenas que mais amostras devem ser analisadas. Já no conjunto de marcadores de 10 X-STRs – Decaplex não foi observado desequilíbrio de ligação.

Na comparação entre os dois painéis, o de 12 marcadores X-STRs (Kit Investigator® Argus X-12QS) mostrou-se mais eficiente que os 10 X-STRs do Decaplex.

Os resultados obtidos no *software* famlink X indicam que os dois conjuntos de marcadores são eficientes, porém os 12 marcadores do kit Argus indicaram ser mais eficientes observando os valores de LR nas análises de parentesco realizadas neste trabalho.

7. REFERÊNCIAS BIBLIOGRÁFICAS

ALMARRI, M. A.; LOOTAH, R. A. Allelic and haplotype diversity of 12 X-STRs in the United Arab Emirates. **Forensic Science International: Genetics**, v. 33, p. e4–e6, 2017.

AMBERS, A. D.; CHURCHILL, J. D.; KING, J. L.; STOLJAROVA, M.; GILL-KING, H.; ASSIDI, M.; ABU-ELMAGD, M.; BUHMEIDA, A.; AL-QAHTANI, M.; BUDOWLE, B. More comprehensive forensic genetic marker analyses for accurate human remains identification using massively parallel DNA sequencing. **BMC genomics**, v.17, p.21-30, 2016.

AMBROSIO, I. B.; BRAGANHOLI, D. F.; ORLANDO, L. B. M.; ANDREKENAS, N. C.; MOTA, D.; SILVA, D.; ASTOLFI, FILHO S.; CARVALHO, D.. CICARELLI, R. M. B.; GUSMÃO, L. Mutational data and population profiling of 23 Y-STRs in three Brazilian populations. **For Sci Int Genet** , v.48, p. 102348, 2020.

ANDERSON, S.; BANKIER, A.T.; BARRELL, B.G. et al. Sequence and organization of the human mitochondrial genome. **Nature**, v. 290, p. 457-465, 1981.

ANDREWS, R.; KUBACKA, I.; CHINNERY, P.F. et al. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. **Nat. Genet.**, v. 23, p. 147, 1999.

ASAMURA, H.; SAKAI, H.; KOBAYASHI, K.; OTA, M.; FUKUSHIMA, H. MiniX-STR multiplex system population study in Japan and application to degraded DNA analysis. **Int. J. Legal Med.**, v. 120, p. 174-181, 2006.

BERGERO, R.; AND CHARLESWORTH, D. The evolution of restricted recombination in sex chromosomes. **Trends Ecol. Evol**, v. 24, p. 94–102, 2009.

BONFERRONI, C. E. Teoria statistica delle classi e calcolo delle probabilità. **Publicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze**, v. 8, p. 3-62, 1936.

BRAGANHOLI, D. F.; FREITAS, J. M.; ANDREKENAS, N. C.; JANUÁRIO, B. B.; AMBROSIO, I. B.; POLVERARI, F. S., CICARELLI, R. M. B. Analysis of the complete mtDNA in Brazilian samples: Haplotypes differentiation and identification of new variations in Native American haplogroups. **Forensic Science International: Genetics Supplement Series**, v. 7, p. 345-346, 2019.

BUDOWLE, B.; ALLARD, M.W.; WILSON, M.R.; CHAKRABORTY, R. Forensics and mitochondrial DNA: applications, debates, and foundations. **Annu. Rev. Genomics Hum. Genet**, v.4, p.119–141, 2003.

BUDOWLE, B.; MORETTI, T. R.; NIEZGODA, S. J.; BROWN, B. L. CODIS and PCR-based short tandem repeat loci: law enforcement tools., Proceedings of the Second European Symposium on Human Identification. **Madison, WI: Promega Corporation**, pp. 73–88, 1998.

BUTLER, J. M. **Forensic DNA typing**. 2nd ed. London: Elsevier Academic Press, 2005.

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. London: Elsevier Academic Press, 2011, 704p.

CAINÉ, L. M.; ZARRABEITIA, M. T.; RIANCHO, J. A.; PINHEIRO, F. M.; Genetic data of a Brazilian population sample (Santa Catarina) using an X-STR decaplex. **J Forensic Legal Med**, v.17, p. 272–274, 2010.

CAINÉ, L.; CARVALHO, R.; COSTA, S.; PEREIRA, M. F.; PINHEIRO, M. F. Interest of X chromosome (Argus X-12 kit) in complex kinship analysis. **Forensic Science International: Genetics Supplement Series**, v.3, p.206–207, 2011.

CRAMER, P. Rosalind Franklin and the Advent of Molecular Biology. **Cell**, v. 182, n. 4, p. 787-789, 2020.

CARTWRIGHT, R. A.; GRAUR, D. The multiple personalities of Watson and Crick strands. **Biology direct**, v.6, p. 1-7., 2011.

DENG, C.; SONG, F.; LI, J.; HOU, Y.; LUO, H. Multiplex PCR for 19 X-chromosomal STRs in Chinese population. **Forensic Science International: Genetics Supplement Series**, v.6, p. e24–e26, 2017.

DESMARAIS, D.; ZHONG, Y.; CHAKRABORTY, R.; PERREAULT, C.; BUSQUE, L. Development of a highly polymorphic STR marker for identity testing purposes at the human androgen receptor gene (HUMARA), **J. Forensic Sci.**, v.43, p.1046–1049, 1998.

DIEGOLI, T. M.; COBLE, M. D. Development and characterization of two mini-X chromosomal short tandem repeat multiplexes. **Forensic Science International: Genetics**, v. 5, p. 415–421, 2011.

DIEGOLI, T. M. Forensic typing of short tandem repeat markers on the X and Y chromosomes. **Forensic Science International: Genetics**, v.18, p.140–151, 2015.

ELAKKARY, S.; HOFFMEISTER-ULLERICH, S.; SCHULZE, C.; SEIF, E.; SHETA, A.; HERING, S., et al. Genetic polymorphisms of twelve X-STRs of the investigator Argus X-12 kit and additional six X-STR centromere region loci in an Egyptian population sample. **Forensic Sci Int Genet.**, v.11, p. 26–30, 2014.

ELHAIK, E.; YUSUF, L.; ANDERSON, A. I. J.; PIROOZANIA, M.; DIMITRIOS ARNELLOS, D.; VILSHANSKY, G.; ERCAL, G.; LU, Y.; WEBSTER, T.; BAIRD, M. L.; ESPOSITO, U. The Diversity of REcent and Ancient huMan (DREAM): A New Microarray for Genetic Anthropology and Genealogy, Forensics, and Personalized Medicine. **Genome Biol. Evol.**, v.9, p. 3225–3237, 2017.

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.

FEDERAL BUREAU OF INVESTIGATION. CODIS System, USA. Disponível em www.fbi.gov Acesso em: 08 dez. 2021.

FAN, H.; CHU, J. Y. A brief review of short tandem repeat mutation. **Genomics Proteomics Bioinformatics**, v. 5, n.1, p.7-14, 2007.

FERRAGUT, J.; BENTAYEBI, K.; CASTRO, J.; CRISTIAN, R.; PICORNELL, A. Genetic analysis of 12 X-chromosome STRs in Western Mediterranean populations, *Int. J. Legal Med.* 129. **Forensic Science International: Genetics**, v.41, p. 253–255, 2015.

GAO, H.; WANG, C.; ZHANG, R. WU H., SUNS, XIAO D., et al. Application of CPI cutoff value based on parentage testing of duos and trios typed by four autosomal kits. **PLoS ONE** ,v.14 (11)p. e0225174 ,2019.

GARCIA, M. G.; CATANESI, C. I.; PENACINO, G. A.; L. GUSMÃO, L.; PINTO, N. X-chromosome data for 12 STRs: Towards an Argentinian database of forensic haplotype frequencies. **Forensic Science International: Genetics**, v.41, p. 8-13, 2019.

GARIBYAN, L.; AVASHIA, N. Polymerase chain reaction. **J Invest Dermatol.**, v.133, n.3, p.1-4, 2013.

GILL, P. et al. Identification of the remains of the Romanov family by DNA analysis. **Nature Genetics**, v. 6, p. 130-135, 1994.

GILL, P.; FEREDAY, L.; MORLING, N.; SCHNEIDER, P. M. The evolution of DNA databases--recommendations for new European STR loci. **Forensic Sci Int.**, v. 156(2-3), p.242-4, 2006.

GJERTSON, D.W. ; BRENNER , C.H. ; BAUR ,M.P. ; CARRACEDO, A.; GUIDET, F.; LUQUE, J.A.; LESSIG, R. ; MAYR, V W.R.; PASCALI, V. L.; PRINZ, M.; SCHNEIDER, P.M. ; MORLING, N.; ISFG: recommendations on biostatistics in paternity testing. **Forensic Sci. Int. Genet.**, v.1 ,p. 223-231, 2007.

GOMES, I.; ALVES, C.; MAXZUD, K.; PEREIRA, R.; PRATA, M. J.; SÁNCHEZ-DIZ, P.; CARRACEDO, A.; AMORIM, A.; GUSMÃO, L. Analysis of 10 X-STRs in three African populations. **Forensic Sci Int Genet** , v.1, p. 208–211, 2007.

GOMES, C.; MAGALHÃES, M.; ALVES, C.; AMORIM, A.; PINTO, N.; GUSMÃO, L. Comparative evaluation of alternative batteries of genetic markers to complement autosomal STRs in kinship investigations: autosomal indels vs. X-chromosome STRs. **Int. J. Legal Med.**, v.126, p. 917-921, 2012.

GOMES, I.; PINTO, N.; ANTÃO-SOUSA, S.; GOMES, V.; GUSMÃO, L.; AMORIM, A. Twenty Years Later: A Comprehensive Review of the X Chromosome Use in Forensic Genetics. **Frontiers in Genetics**, v. 11, p.926, 2020.

GRAFFELMAN, J.; WEIR, B. S. Testing for Hardy–Weinberg equilibrium at biallelic genetic markers on the X chromosome. **Heredity**, v. 116, p. 558–568, 2016.

GROVER, A.; SHARMA, P. C. Development and use of molecular markers: past and present, **Critical Reviews in Biotechnology**, v.36, n. 2, p.290-302, 2016.

GUPTA, V.; SENGUPTA, M.; PRAKASH, J.; TRIPATHY, B. C. An Introduction to Biotechnology. In: **Basic and Applied Aspects of Biotechnology**. Springer, Singapore, 2017, p.1-21.

GUSMÃO, L.; SÁNCHEZ-DIZ, P.; ALVES, C.; GOMES, I.; ZARRABEITIA, M. T.; ABOVICH, M.; ATMETLLA, I.; BOBILLO, C.; BRAVO, L.; BUILES, J.; CAINÉ, L.; CALVO, R.; CARVALHO, E.; CARVALHO, M.; CICALARELLI, R.; CATELLI, L.; CORACH, D.; ESPINOZA, M.; GARCÍA, O.; MALAGHINI, M.; MARTINS, J.; PINHEIRO, F.; PORTO, M. J.; RAIMONDI, E.; RIANCHO, J. A.; RODRÍGUEZ, A.; RODRÍGUEZ, A.; CARDOZO, B. R.; SCHNEIDER, V.; SILVA, S.; TAVARES, C.; TOSCANINI, U.; VULLO, C.; WHITTLE, M.; YURREBASO, I.; CARRACEDO, A.; AMORIM, A. A GEP-ISFG collaborative study on the optimization of an X-STR decaplex: data on 15 Iberian and Latin American populations. **Forensic Science International: Genetics**, v. 123, p. 227–234, 2009.

HABIBI, S.; AHMADI, A.; BEHMANESH, M.; MIRI, A.; TAVALLAIE, M. Evaluation of ten SNP Markers for Human Identification and Paternity Analysis in Persian Population. **Iranian journal of biotechnology**, v.17, n.3, p. e2148, 2019.

HOLLAND, M.M., et al. Mitochondrial DNA sequence analysis of human skeletal remains: Identification of remains from the Vietnam War. **Journal of Forensic Sciences**, v.38, p.542-553, 1993.

ILLESCAS, M. J.; PEREZ, A.; AZNAR, J. M.; VALVERDE, L.; CARDOSO, S.; ALGORTA, J.; PANCORBO, M. M. Population genetic data for 10 X-STR loci in autochthonous Basques from Navarre (Spain). **Forensic Science International: Genetics**, v. 6, p.146–148, 2012.

ISHINO, S.; ISHINO, Y. DNA polymerases as useful reagents for biotechnology – the history of developmental research in the field. **Front. Microbiol.**, v. 5, p.465, 2014.

JÄGER A. C., ALVAREZ M. L., DAVIS C. P., GUZMÁN E., HAN Y., WAY L., et al. Developmental validation of the MiSeq FGx Forensic Genomics System for Targeted Next Generation Sequencing in Forensic DNA Casework and Database Laboratories. **Forensic Sci Int Genet.**, v. 28, p. 52–70, 2017.

JEFFREYS, A. J.; WILSON, V.; THEIN, S. L. Hypervariable “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(6040), p. 818–819, 1985b.

JEFFREYS, A. J.; ALLEN, M. J.; HAGELBERG, E.; SONNBERG, A. Identification of the skeletal remains of Josef Mengele by DNA analysis, **Forensic Science International**, v. 56, n. 1, p. 65-76, 1992.

JEFFREYS, A. J. The man behind the DNA fingerprints: an interview with Professor Sir Alec Jeffreys. **Investigative genetics**, v.4, n. 21, p. 1-7, 2013.

KARKI, R.; PANDYA, D.; ELSTON, R.C. et al. Defining “mutation” and “polymorphism” in the era of personal genomics. **BMC Med Genomics**, v. 8, n.37, p.1-7, 2015.

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

KLING, D.; DELL'AMICO, B.; TILLMAR, A. O. FamLinkX – implementation of a general model for likelihood computations for X-chromosomal marker data. **Forensic Science International: Genetics**, v. 17, p. 1-7, 2015.

KOBACHUK, L. D. G.; M., MALAGHINI; SILVA, J. M.; GUSMÃO, L.; BICALHO, M. G. Genetic profile characterization of ten X-STRs in a sample from Paraná, Brazil. **Int J Legal Med**, v.126, p. 975–976, 2012.

KRINGS, M. et al. Neandertal DNA sequences and the origin of modern humans. **Cell**, 90,p.19-30, 1997.

LANDER, E. S.; LINTON, L. M.; BIRREN, B.; NUSBAUM, C.; ZODY, M. C.; BALDWIN, J., et al. Initial sequencing and analysis of the human genome. **Nature** v. 409, p. 860–921, 2001.

LI, J.; DENG, C.; LUO, H.; SONG, F.; HOU, Y. Analyzing an “Off-Ladder” allele at DXS10135 from the AGCU X19 STR kit. **Forensic Science International: Genetics Supplement Series** v.6, p. e92–e93, 2017.

LIN, L.; LI, J.; HU, Y.; WANG, H.; MARAH, F. A.; MOSERAY, M., et al. Genetic characterization of 19 X-STRs in Sierra Leone population from Freetown. **Int. J. Legal Med**. v. 34, p.1659–1661, 2020.

LIU, Q.L.; LU, D.J.; QUAN, L.; CHEN, Y.F.; SHEN, M.; ZHAO, H. Development of multiplex PCR system with 15 X-STR loci and genetic analysis in three nationality populations from China. **Electrophoresis**, v. 33, p. 1299-1305, 2012.

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.

MARTIN, W. F.; GARG, S.; ZIMORSKI, V. Endosymbiotic theories for eukaryote origin. **Philosophical Transactions of the Royal Society B: Biological Sciences**, v. 370, p.1678, 2015.

MARTINEZ, J; POLVERARI F. S; SILVA, F. A. J; BRAGANHOLI, D. F; FERRAZ, J. A. M. F; GUSMÃO, L; CICARELLI, R. M. B. Genetic characterization of 32 X-InDels in a population sample from São Paulo State (Brazil). **Int J Legal Med.**, v.133, p.1385–1388, 2019.

MARTINS, J.A.; COSTA, J. C.; PANETO, G. G.; FIGUEIREDO, R. F.; GUSMÃO, L.; SÁNCHEZ-DIZ, P.; CARRACEDO, A.; CICARELLI, R. M. B. Genetic profile characterization of 10 X-STRs in four populations of the southeastern region of Brazil. **Int J Legal Med**, v. 124, p.427–432, 2010.

MARTINS, J. A.; KAWAMURA, B.; CARDOSO, A. E.; CICARELLI, R. M. Brazilian genetic database of chromosome X. **Mol Biol Rep**, v. 41, p. 4077-4080, 2014.

MARTINS, J. A.; MARTINS, D.; OLIVEIRA-BRANCATI, C. I. F.; JULIANA MARTINEZ, J.; CICARELLI, R. M. B.; SOUZA, D. R. S. 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, p. 1523–1530, 2017.

MATISE, T. C.; CHEN, F.; CHEN, W.; DE LA VEGA, F. M.; HANSEN, M.; HE, C.; HYLAND, F. C. L.; KENNEDY, G. C.; KONG, X.; MURRAY, S. S.; ZIEGLE, J. S.; STEWART, W. C. L.; BUYSKE, S. A second-generation combined linkage–physical map of the human genome. **Genome Research**. v.17, p.1783-1786, 2007.

MRŠIĆ, G.; OZRETIĆ, P.; CRNJAC, J.; MERKAŠ, S.; SUKSER, V.; RAČIĆ, I.; ROŽIĆ, S.; BARBARIĆ, L.; POPOVIĆE, M.; KOROLIJA, M. Expanded Croatian 12 X-STR loci database with an overview of anomalous profiles. **Forensic Science International: Genetics**. v. 34, p. 249–256, 2018

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.

MULLIS, K., FALOONA, F., SCHARF, S., SAIKI, R., HORN, G., & ERLICH, H. Specific Enzymatic Amplification of DNA In Vitro: The Polymerase Chain Reaction. **Cold Spring Harbor Symposia on Quantitative Biology**, v. 51(0), p. 263–273, 1986. doi:10.1101/sqb.1986.051.01.032

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

MULLIS, K. B. The Unusual Origin of the Polymerase Chain Reaction. **Scientific American**, v.262, n.4, p. 56–65, 1990.

NASS, S., & NASS, M. M. Intramitochondrial fibers with DNA characteristics. Enzymatic and other hydrolytic treatments. **The Journal of cell biology**, v. 19, n.3, p. 613–629, 1963.

NOVROSKI, N. M. M. Exploring new short tandem repeat markers for DNA mixture deconvolution. **WIREs Forensic Sci**, v.3, p.e1390, 2020.

PALHA, T. J.; RODRIGUES, E. M.; SANTOS, D. 'Y-STR haplotypes of Native American populations from the Brazilian Amazon region.', **Forensic Sci Int Genet** v.4(5), p.e121-3, 2010.

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.

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.

PINTO, N.; GUSMÃO, L.; EGELAND, T.; AMORIM, A. Paternity exclusion power: Comparative behaviour of autosomal and X-chromosomal markers in standard and deficient cases with inbreeding. **Forensic Science International: Genetics**, v.7, p. 290–295, 2013a.

PINTO, N.; MAGALHÃES, M.; CONDE-SOUSA, E.; GOMES, C.; PEREIRA, R.; ALVES, C.; GUSMÃO, L.; AMORIM, A. Assessing paternities with inconclusive STR results: the suitability of bi-allelic markers. **Forensic Sci Int Genet**. V. 7, n. 1, p. 16-21, 2013b.

PINTO, N.; PEREIRA, V.; TOMAS, C.; LOIOLA, S.; CARVALHO, E. F.; MODESTI, N.; MAXZUD, M.; MARCUCCI, V.; CANO, H.; CICARELLI, R.; JANUARIO, B.; BENTO, A.; BRITO, P.; BURGOS, G.; PAZ-CRUZ, E.; DÍEZ-JUÁREZ, L.; VANNELLI, S.; PONTES, M. L.; BERARDI, G.; FURFURO, S.; FERNANDEZ, A.; SUMITA, D.; BOBILLO, C.; GARCÍA, M. G.; GUSMÃO, L. Paternal and maternal mutations in X-STRs: A GHEP-ISFG collaborative study. **Forensic Science International: Genetics**, v. 46, p. 228-230, 2020.

POSYNICK, B. J.; BROWN, C. J. Escape From X-Chromosome Inactivation: An Evolutionary Perspective. **Front. Cell Dev. Biol.**, v. 7, p. 241, 2019.

PRIETO-FERNANDEZ, E.; BAETA, M.; NÚÑEZ, C.; ZARRABEITIA, M. T.; HERRERA, R. J., BUILES, J. J. ; PANCORBO, M. M. DE. Development of a new highly efficient 17 X-STR multiplex for forensic purposes **Electrophoresis**. v. 37, p. 1651-1658, 2016.

RAGHUNATHAN, G.; MARX, A. Identification of *Thermus aquaticus* DNA polymerase variants with increased mismatch discrimination and reverse

transcriptase activity from a smart enzyme mutant library. **Scientific reports**, v.9(1), n. 590, 2019.

RELETFORD, J.H. Reflections of our past: How human history is revealed in our genes. **Boulder Colorado:Westview Press**, 2004. 269p.

ROBINO, C.; LACERENZA, D.; ANELI, S.; DI GAETANO, C.; MATULLO, G.; ROBLEDO, R.; CALÒ, C. Allele and haplotype diversity of 12 X-STRs in Sardinia, **Forensic Science International: Genetics**, v. 33, p. e1-e2, 2018.

ROSS, M. T. et al. The DNA sequence of the human X chromosome. **Nature**. v.434, n. 7031, p.325–337,2005.

SAAD, R. Discovery, development, and current applications of DNA identity testing. **Proceedings (Baylor University. Medical Center)**, v. 18, n. 2, p. 130-3, 2005.

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. **Nature Reviews Genetics**, v.5, p.43–51, 2004.

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

SINGH, R. S.; KULATHINAL, R.J. Polymorphism. Editor(s): Stanley Maloy, Kelly Hughes in *Brenner's Encyclopedia of Genetics (Second Edition)*, **Academic Press**, p. 398-399, 2013.

SLATKIN, M. Linkage disequilibrium — understanding the evolutionary past and mapping the medical future. **Nat Rev Genet**, v.9, p.477–485, 2008.

STANGE, V. S.; SILVA, R.; MARIANO, F.; LIMA, M.; MAYUMI, C.; DE PAULA, F.; DOS SANTOS, E. V. W., MADEIRA, A.; DRUMOND, I.; GUSMÃO, L. Stratification among European descent and admixed Brazilian populations of Espírito Santo for 27 Y-STRs, **For Sci Int Genet**, v. 41, p. e20-e22, 2019.

STONE, A. C. et al., Mitochondrial DNA analysis of the presumptive remains of Jesse James. **Journal of Forensic Sciences**, v. 46, p.173-176, 2001.

SZIBOR, R.; KRAWCZAK, M.; HERING, S.; · J. EDELMANN, J.; KUHLSCH, E.; KRAUSE, D. Use of X-linked markers for forensic purposes. **Int J Legal Med**, v. 117, n. 2, p. 67-74, 2003

SZIBOR, R.; HERING, S.; EDELMANN, J. A new Web site compiling forensic chromosome X research is now online. **Int J Legal Med**, v. 120, p. 252-254, 2006.

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

TOMAS, C.; PEREIRA, V.; MORLING, N. Analysis of 12 X-STRs in Greenlanders, Danes and Somalis using Argus X-12. **Int. J. Legal Med**, v. 126, p. 121–128, 2012.

TILLMAR, A. O.; MOSTAD, M.; EGELAND, T.; LINDBLUM, B.; HOLMLUND, G.; MONTELIUS, K. Analysis of linkage and linkage disequilibrium for eight X-STR markers, **Forensic Science International: Genetics**, v. 3, n.1, p.37-41, 2008.

TILLMAR, A. O.; KLING, D.; BUTLER, J. M.; PARSON, W.; PRINZ, M.; SCHNEIDER, P. M.; EGELAND, T.; GUSMÃO, L. '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. 269-275, 2017.

TURRINA, S.; ATZEI, R.; FILIPPINI, G.; DE LEO, D. Development and forensic validation of a new multiplex PCR assay with 12 X-chromosomal short tandem repeats. **Forensic Sci. Int. Genet.**, 1 (2007), pp. 201-214, 2007

YANG, J.; HE, J.; WANG, D.; SHI E.; YANG, W; GENG, Q.; WANG, Z. Progress in research and application of InDel markers. **Biodiv Sci**, v.24(2), p.237-243, 2016.

YANG, N.; Li, H.; CRISWELL, L. A.; GREGERSEN, P. K.; ALARCON-RIQUELME, M. E.; KITTLES, R.; SHIGETA, R. et al., Examination of ancestry and ethnic affiliation using highly informative diallelic DNA markers: application to diverse and admixed populations and implications for clinical epidemiology and forensic medicine. **Hum. Genet.**, 118, p. 382-392, 2005.

YUAN, L.; OU, Y.; LIAO, Q.; GUI, J.; XUE BAI, X.; GE, J.; YE, J.; ZHANG, L. Population genetics analysis of 38 STR loci in the she population from Fujian province of China. **Legal Medicine**, v. 16, p. 314-318, 2014.

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

WIJCHERS, P.J.; FESTENSTEIN, R.J. Epigenetic regulation of autosomal gene expression by sex chromosomes. **Trends Genet**. v. 27. n. 4. p. 132-40, 2011.

WYNER, N.; BARASH, M.; MCNEVIN, D. Forensic autosomal short tandem repeats and their potential association with phenotype. **Frontiers in Genetics**, v. 11, p 884, 2020.

ZAGORSKI, N. Profile of Alec J. Jeffreys. **Proceedings of the National Academy of Sciences**. v. 103, n. 24 p. 8918-8920, 2006.

ZARRABEITIA, M. T.; Alonso, A.; Martin, J. et al. Study of six X-linked tetranucleotide microsatellites: population data from five Spanish regions. **Int J Legal Med** ,v.120, p.147–150, 2006.

ZARRABEITIA, M. T.; PINHEIRO, F.; PANCORBO, M. M.; CAINE, L.; CARDOSO, S.; LEONOR GUSMÃO, L.; RIANCHO, J. A. Analysis of 10 X-linked tetranucleotide markers in mixed and isolated populations. **Forensic Science International: Genetics**, v. 3, p.63–66, 2009.

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

TURCHETTO-ZOLET, A. C.; TURCHETTO, C.; ZANELLA, C. M.; PASSAIA, G. Marcadores moleculares na era genômica: Metodologias e Aplicações – **Ribeirão Preto: Sociedade Brasileira de Genética**, 2017. 181 p.

APÊNDICES

APÊNDICE A - FREQUÊNCIA HAPLOTÍPICA DOS 4 GRUPOS DE LIGAÇÃO (LG) DOS 12 X-STRS DO KIT ARGUS.

LG1			LG2			LG3			LG4		
No.	Haplótipo	Frequência	No.	Haplótipo	Frequência	No.	Haplótipo	Frequência	No.	Haplótipo	Frequência
1	13.3 20 11	0,003546	1	13 18 17	0,007092	1	18 11 33	0,003546	1	26 33 15	0,007092
2	18 29 11	0,003546	2	14 21 18	0,01773	2	20 14 33	0,003546	2	27 36 16	0,003546
3	25.1 19 12	0,003546	3	14 19 13	0,010638	3	16 13 31.2	0,007092	3	26 30 14	0,003546
4	29.1 33 12	0,007092	4	15 21 17	0,028369	4	20 14 31.2	0,010638	4	33 37 14	0,003546
5	24.1 18 10	0,003546	5	13 20 17	0,010638	5	18 12 24.2	0,003546	5	30 36 14	0,007092
6	28.1 23 11	0,003546	6	13 20 18	0,010638	6	16 12 30.2	0,007092	6	35.2 38.3 15	0,003546
7	19 24 10	0,003546	7	12 21 14	0,003546	7	20 12 31	0,003546	7	28 34 15	0,003546
8	25.1 27 11	0,007092	8	14 22 14	0,003546	8	16 13 30	0,01773	8	29 35 15	0,014184
9	26.1 19 10	0,003546	9	14 18 8	0,007092	9	18 13 33	0,007092	9	28 33 14	0,003546
10	29.1 25 11	0,003546	10	15 20 18	0,010638	10	16 11 28.2	0,007092	10	29 37 15	0,014184
11	18 18.1 10	0,007092	11	13 21 15	0,014184	11	18 14 33	0,021277	11	27 35 15	0,007092
12	23.1 24 12	0,003546	12	14 21 15	0,010638	12	17 11 32	0,003546	12	31 38 15	0,010638
13	24.1 27 10	0,003546	13	12 16 17	0,003546	13	19 15 33.2	0,003546	13	35.2 36 14	0,007092
14	30.1 30 12	0,003546	14	12 21 19	0,003546	14	19 9 24.2	0,003546	14	41.2 39 14	0,003546
15	24.1 20.1 12	0,003546	15	15 23 18	0,003546	15	19 12 31	0,007092	15	27 30 14	0,003546
16	24.1 24 12	0,003546	16	14 14 15	0,003546	16	16 13 31	0,014184	16	31 39.3 14	0,003546
17	25.1 23 10	0,007092	17	15 19 18	0,007092	17	16 11 32	0,007092	17	42.2 36 15	0,003546
18	27.1 25 10	0,007092	18	13 17 8	0,007092	18	18 13 32.2	0,003546	18	25 38.1 15	0,003546
19	23.1 24 11	0,007092	19	16 20 14	0,003546	19	16 13 28.2	0,003546	19	26 39 17	0,003546
20	28.1 19 10	0,003546	20	13 18 15	0,01773	20	18 13 32	0,007092	20	27 33 14	0,010638
21	29.1 29 11	0,003546	21	14 20 17.2	0,003546	21	17 12 29	0,007092	21	44.2 33 15	0,003546
22	23 19 9	0,003546	22	13 20 15	0,007092	22	19 14 30	0,010638	22	28 36 13	0,003546
23	26.1 31 12	0,007092	23	14 21 14	0,003546	23	17 13 30	0,010638	23	28 37 14	0,007092
24	18 18 10	0,007092	24	14 21 19	0,010638	24	19 13 30	0,007092	24	25 35 13	0,007092
25	26.1 28 11	0,014184	25	13 21 16	0,007092	25	18 9 28	0,003546	25	29 36 14	0,014184
26	18 17 10	0,003546	26	14 22 20	0,003546	26	19 12 32	0,007092	26	26 36 14	0,003546
27	28 22 10	0,003546	27	12 16 18	0,003546	27	18 12 30.2	0,010638	27	43.2 37 14	0,003546

28	19 26 11	0,003546	28	14 20 19	0,010638	28	19 14 33	0,010638	28	30 37 16	0,007092
29	24.1 15 10	0,003546	29	13 19 17	0,010638	29	17 14 31.2	0,003546	29	28 34.2 14	0,003546
30	26 25 11	0,003546	30	14 20 18	0,01773	30	18 15 31.2	0,007092	30	41.2 35 15	0,003546
31	28.1 30 11	0,003546	31	14 19 7	0,01773	31	19 11 27.2	0,010638	31	28 35 16	0,003546
32	18 24 10	0,014184	32	14 22 15	0,003546	32	19 12 30.2	0,01773	32	25 35 14	0,003546
33	24.1 27 12	0,003546	33	12 20 14	0,010638	33	17 11 24.2	0,003546	33	30 38 15	0,007092
34	25.1 21 10	0,003546	34	14 20 16	0,014184	34	19 12 29	0,01773	34	30 35 15	0,010638
35	25.1 22.1 10	0,003546	35	14 20 17	0,01773	35	17 12 28.2	0,003546	35	43.2 35 16	0,003546
36	25.1 29 11	0,003546	36	13 19 7	0,007092	36	20 14 31	0,014184	36	28 38 15	0,003546
37	18 25 10	0,007092	37	17.3 20 18	0,003546	37	18 12 26.2	0,003546	37	43.2 38 15	0,003546
38	24.1 28 12	0,007092	38	14 20 15	0,01773	38	19 13 32	0,021277	38	25 37 15	0,003546
39	18 19 10	0,01773	39	12 21 15	0,003546	39	16 12 30	0,007092	39	25 34 15	0,003546
40	20 19 12	0,003546	40	14 21 16	0,003546	40	19 13 31	0,010638	40	27 37 15	0,007092
41	13.3 18 10	0,007092	41	12 19 14	0,003546	41	16 14 31	0,007092	41	30 37 15	0,007092
42	25.1 18.1 10	0,003546	42	13 21 19	0,007092	42	20 16 33	0,003546	42	26 34 14	0,014184
43	19 20 10	0,007092	43	13 18 14	0,007092	43	19 12 33.2	0,003546	43	27 36 14	0,007092
44	28.1 21.1 12	0,003546	44	13 21 12	0,003546	44	18 11 31	0,003546	44	26 35 14	0,01773
45	24.1 23 11	0,003546	45	13 22 18	0,003546	45	20 13 32	0,01773	45	27 35 17	0,003546
46	25.1 25 12	0,003546	46	13 18 12	0,007092	46	20 11 30	0,003546	46	30 32 15	0,003546
47	38.1 27 12	0,003546	47	15 21 14	0,007092	47	16 11 31	0,007092	47	40.2 36.3 15	0,003546
48	19 23 11	0,007092	48	13 16 14	0,003546	48	20 13 33	0,010638	48	27 36 15	0,003546
49	24.1 26 11	0,003546	49	17 21 15	0,003546	49	16 11 27.2	0,003546	49	28 38 16	0,003546
50	24.1 20 10	0,003546	50	12 17 16	0,003546	50	19 12 30	0,010638	50	28 40 15	0,003546
51	27.1 22 12	0,003546	51	14 23 18	0,003546	51	18 15 30	0,003546	51	24 35 15	0,007092
52	25.1 18 10	0,007092	52	14 23 19	0,003546	52	18 12 29.2	0,014184	52	27 38 15	0,010638
53	28.1 19 12	0,003546	53	13 20 16	0,010638	53	19 13 30.2	0,014184	53	23 36 15	0,007092
54	23.1 26 12	0,003546	54	14 22 16	0,007092	54	18 13 27	0,003546	54	43.2 36.3 15	0,003546
55	18 25 11	0,003546	55	12 20 12	0,003546	55	16 13 29	0,007092	55	32 35 14	0,003546
56	18 30 12	0,003546	56	14 21 17.3	0,003546	56	19 15 34	0,003546	56	35.2 37 14	0,003546
57	24.1 30 12	0,003546	57	13 16 15	0,003546	57	16 13 33	0,007092	57	28 39 15	0,003546

58	25.1 30 12	0,007092	58	15 18 16	0,003546	58	18 14 34	0,007092	58	40.2 41.3 15	0,007092
59	25.1 18 12	0,003546	59	14 21 17	0,014184	59	17 7 32	0,003546	59	29 34 15	0,003546
60	30.1 19 12	0,003546	60	16 21 17	0,003546	60	17 13 33	0,007092	60	40.2 39.3 15	0,003546
61	24.1 16.1 11	0,003546	61	13 20 13	0,007092	61	16 12 26.2	0,014184	61	33 37 16	0,003546
62	25.1 21 12	0,010638	62	14 20 13	0,003546	62	21 13 32	0,007092	62	39.2 35 15	0,003546
63	23.1 18 11	0,003546	63	14 18 7	0,003546	63	18 13 31.2	0,003546	63	43.2 36 16	0,003546
64	26.1 26 11	0,007092	64	14 19 14	0,007092	64	20 15 31.2	0,003546	64	40.2 34 15	0,003546
65	25 18.1 11	0,003546	65	13 20 14	0,014184	65	18 13 28.2	0,007092	65	43.2 41.3 16	0,003546
66	26.1 19 13	0,003546	66	13 21 18	0,007092	66	19 14 31	0,007092	66	28 35 14	0,007092
67	18 16 10	0,003546	67	14 22 17	0,007092	67	19 13 29.2	0,003546	67	30 39.3 15	0,003546
68	38.1 17.1 12	0,003546	68	13 15 15	0,007092	68	19 14 32	0,010638	68	29 34 14	0,007092
69	19 21 12	0,003546	69	13 18 16	0,007092	69	17 12 30	0,003546	69	29 37 14	0,007092
70	27.1 23 12	0,003546	70	15 22 18	0,007092	70	17 14 32	0,003546	70	34.2 36 14	0,003546
71	25.1 20 10	0,003546	71	16 20 20	0,003546	71	16 12 27.2	0,003546	71	45.2 41.3 15	0,003546
72	27.1 31 11	0,003546	72	16 21 16	0,003546	72	18 15 31	0,007092	72	28 36 14	0,007092
73	13.3 20.1 10	0,007092	73	14 20 14	0,007092	73	16 12 29.2	0,010638	73	33 40.3 15	0,003546
74	27.1 31 13	0,003546	74	15 21 16	0,003546	74	19 9 30	0,003546	74	27 35 14	0,021277
75	18 21 11	0,003546	75	12 19 12	0,003546	75	21 12 30.2	0,003546	75	29 35 14	0,010638
76	21.1 25 12	0,003546	76	12 20 16	0,003546	76	19 15 31.2	0,007092	76	43.2 36 15	0,007092
77	24 24 12	0,003546	77	16.3 20 18	0,003546	77	19 14 30.2	0,003546	77	26 33 14	0,003546
78	24.1 27 13	0,003546	78	16 21 18	0,007092	78	19 12 29.2	0,014184	78	35.2 36 15	0,003546
79	24.1 26 10	0,003546	79	15 19 17	0,014184	79	20 13 32.2	0,003546	79	28 32 14	0,007092
80	25.1 26 11	0,003546	80	17 21 17	0,003546	80	16 14 30	0,003546	80	38.2 36 14	0,003546
81	13.3 22 10	0,007092	81	12 19 8	0,007092	81	19 15 30	0,003546	81	26 37 14	0,010638
82	18 21 10	0,010638	82	17 20 15	0,003546	82	18 12 27.2	0,007092	82	33 37 15	0,003546
83	26.1 23 12	0,003546	83	13 18 8	0,007092	83	17 13 30.2	0,003546	83	24 36 15	0,003546
84	25.1 26 10	0,003546	84	16 21 15	0,003546	84	20 15 32	0,010638	84	28 38 14	0,007092
85	26.1 27 11	0,014184	85	12 16 16	0,003546	85	18 13 29	0,010638	85	42.2 38 14	0,003546
86	22.1 22 10	0,003546	86	14 17 7	0,003546	86	19 14 31.2	0,014184	86	27 32 14	0,003546
87	27.1 18.1 11	0,003546	87	15 20 16	0,014184	87	17 12 24.2	0,003546	87	30 35 14	0,003546

88	28 27 11	0,003546	88	13 19 16	0,007092	88	16 13 28	0,003546	88	24 35 14	0,003546
89	18 24 11	0,003546	89	16 20 16	0,003546	89	19 12 28	0,003546	89	29 36 15	0,003546
90	18 26 11	0,003546	90	12 16 11	0,003546	90	21 12 32.2	0,003546	90	24 32 14	0,007092
91	13.3 19 11	0,007092	91	12 19 16	0,003546	91	19 12 27.2	0,003546	91	41.2 38 15	0,003546
92	23 24 13	0,003546	92	13 21 17	0,003546	92	15 11 32	0,003546	92	27 34.2 14	0,003546
93	28 21.1 11	0,003546	93	14 15 17	0,007092	93	20 13 31.2	0,01773	93	35.2 34 14	0,003546
94	31 23 13	0,003546	94	14 18 17	0,003546	94	19 14 27	0,003546	94	44.2 35 15	0,003546
95	18 19.1 12	0,003546	95	13 16 18	0,003546	95	20 15 33	0,003546	95	44.2 34 13	0,003546
96	25.1 22 12	0,007092	96	14 18 13	0,003546	96	17 13 32.2	0,003546	96	45.2 36 15	0,003546
97	13.3 19 10	0,003546	97	14 20 8	0,003546	97	16 13 30.2	0,010638	97	27 37 14	0,007092
98	23.1 20 14	0,003546	98	15 21 13	0,007092	98	19 15 30.2	0,007092	98	28 43.3 17	0,003546
99	23 20.1 10	0,003546	99	12 16 19	0,003546	99	20 13 30.2	0,007092	99	31 34 14	0,003546
100	25.1 25 10	0,003546	100	13 18 9	0,003546	100	19 13 33	0,007092	100	32 36 15	0,003546
101	25.1 17 10	0,003546	101	15 19 16	0,007092	101	18 12 31	0,007092	101	28 35 15	0,003546
102	27.1 19 12	0,003546	102	12 19 7	0,003546	102	20 16 32.2	0,003546	102	32 37 16	0,003546
103	18 19 11	0,003546	103	13 20 7	0,007092	103	16 12 31.2	0,003546	103	30 36 16	0,003546
104	29 33 12	0,003546	104	15 21 18	0,003546	104	16 13 34	0,003546	104	35 34 12	0,003546
105	23 22.1 10	0,003546	105	13 17 12	0,003546	105	16 12 31	0,007092	105	39.2 36 15	0,003546
106	29.1 27 11	0,003546	106	15.3 21 19	0,003546	106	18 13 31	0,003546	106	24 34 15	0,003546
107	18 23 10	0,003546	107	14 16 19	0,003546	107	21 13 31.2	0,003546	107	27 34.2 15	0,003546
108	22 24 12	0,003546	108	14 22 19	0,003546	108	19 11 29.2	0,007092	108	30 34 15	0,007092
109	24.1 22 11	0,003546	109	14 20 7	0,003546	109	19 14 32.2	0,003546	109	42.1 39 16	0,003546
110	24.1 29 11	0,003546	110	15 20 17	0,010638	110	18 11 29.2	0,014184	110	28 41.3 15	0,003546
111	18 21.1 10	0,003546	111	13 18 7	0,007092	111	18 13 29.2	0,003546	111	29 38.3 14	0,003546
112	25.1 28 12	0,003546	112	15 18 14	0,003546	112	19 12 31.2	0,007092	112	46.2 40.3 16	0,003546
113	26.1 19.1 10	0,003546	113	15 20 19	0,007092	113	16 13 29.2	0,003546	113	27 39 14	0,003546
114	26.1 21.1 10	0,003546	114	14.3 21 16	0,003546	114	16 14 32	0,003546	114	29 39 15	0,007092
115	24.1 17 10	0,003546	115	13 19 11	0,003546	115	19 14 35	0,007092	115	28 36.1 13	0,003546
116	27.1 18 11	0,003546	116	14 20 11	0,003546	116	17 15 30	0,003546	116	34.2 37 14	0,003546
117	26.1 21 11	0,003546	117	16 21 19	0,003546	117	18 12 29	0,007092	117	27 32 13	0,007092

118	27.1 21 12	0,003546	118	12 17 17	0,003546	118	18 12 34	0,003546	118	29 36.1 14	0,003546
119	23.1 20 10	0,003546	119	13 18 13	0,007092	119	17 12 30.2	0,003546	119	25 39 14	0,003546
120	26.1 29 12	0,003546	120	14 15 14	0,003546	120	16 12 32	0,003546	120	26 40.3 15	0,003546
121	22.1 18 10	0,003546	121	14 19 8	0,003546	121	21 12 32	0,003546	121	25 36.1 15	0,003546
122	27.2 25 12	0,003546	122	15 22 15	0,007092	122	16 9 30.2	0,003546	122	29 38 15	0,003546
123	13.3 27 10	0,003546	123	12 17 7	0,003546	123	19 13 35	0,003546	123	43.2 34 14	0,003546
124	22.1 31 10	0,003546	124	15 20 12	0,003546	124	19 11 31.2	0,003546	124	43.2 37 16	0,007092
125	24.1 21 12	0,003546	125	14 22 18	0,010638	125	19 13 29	0,010638	125	28 33 15	0,007092
126	18 17 11	0,007092	126	15.3 19 17	0,003546	126	18 12 28.2	0,003546	126	36.2 37.3 15	0,003546
127	26.1 24 13	0,003546	127	17.3 22 18	0,003546	127	20 14 32	0,007092	127	32 30 16	0,003546
128	18 15 10	0,003546	128	14 21 8	0,003546	128	18 14 31.2	0,003546	128	36.2 41.3 16	0,003546
129	24.1 31 11	0,003546	129	17 22 16	0,003546	129	16 11 30	0,003546	129	26 28 14	0,003546
130	24 18 10	0,003546	130	12 18 13	0,003546	130	17 12 31.2	0,003546	130	30 38 16	0,003546
131	27.1 27 10	0,003546	131	17 20 17	0,003546	131	16 12 28.2	0,007092	131	25 38 15	0,003546
132	18 22 11	0,003546	132	13 16 17	0,003546	132	16 12 29	0,010638	132	30.2 34 14	0,003546
133	22 26 12	0,003546	133	14 19 18	0,007092	133	20 13 29.2	0,007092	133	31 37 15	0,003546
134	23 25 10	0,003546	134	16 20 18	0,007092	134	17 14 35	0,003546	134	25 34 14	0,003546
135	25.1 30 11	0,003546	135	15 20 14	0,003546	135	17 12 29.2	0,003546	135	25 40 15	0,003546
136	18 21 12	0,003546	136	15 18 13	0,003546	136	20 14 32.2	0,003546	136	26 30 15	0,003546
137	23.1 21 10	0,003546	137	11 19 7	0,003546	137	19 13 31.2	0,014184	137	31 37.3 15	0,003546
138	25.1 23 12	0,003546	138	12 21 18	0,003546	138	18 11 26.2	0,003546	138	31 36 14	0,010638
139	18 19.1 10	0,010638	139	12 18 16	0,003546	139	19 13 33.2	0,007092	139	26 36 15	0,003546
140	24.1 17 11	0,003546	140	15.3 20 17	0,003546	140	19 11 30.2	0,003546	140	28 37 16	0,010638
141	24.1 21 13	0,003546	141	16.3 21 18	0,003546	141	18 14 26	0,003546	141	24 28 14	0,003546
142	23 18 11	0,003546	142	15 20 7	0,003546	142	21 14 27	0,003546	142	27 38 14	0,003546
143	36.1 22 12	0,003546	143	12 20 18	0,003546	143	17 12 31	0,003546	143	27 34 15	0,003546
144	19 23 13	0,003546	144	16 16 8	0,003546	144	17 13 32	0,003546	144	26 34 15	0,003546
145	29.1 24 10	0,003546	145	17 18 16	0,003546	145	16 14 31.2	0,003546	145	28 36 17	0,003546
146	35.1 29 11	0,003546	146	14 19 12	0,003546	146	19 15 33	0,003546	146	30 39 15	0,003546
147	18 17 12	0,003546	147	15 22 19	0,003546	147	20 12 29.2	0,003546	147	42.2 35 14	0,003546

148	19 23 12	0,003546	148	12 15 16	0,003546	148	19 12 28.2	0,003546	148	45.2 40.3 14	0,003546
149	19 17 11	0,003546	149	15 22 17	0,007092	149	20 12 31.2	0,003546	149	39.2 38 15	0,003546
150	41.1 30 11	0,003546	150	16 24 17	0,003546	150	18 11 28	0,003546	150	26 36.3 15	0,003546
151	41.1 22 11	0,003546	151	16 20 15	0,003546	151	17 16 32	0,003546	151	36.2 36 15	0,007092
152	23 21 10	0,003546	152	12 15 8	0,003546	152	19 12 32.2	0,003546	152	24 30 14	0,003546
153	20 22 10	0,003546	153	14 19 15	0,003546	153	15 13 31.2	0,003546	153	27 31 15	0,003546
154	24.1 20 11	0,003546	154	13 19 14	0,003546	154	16 12 33	0,003546	154	28 31 15	0,003546
155	41.1 23 12	0,007092	155	14 18 18	0,003546	155	18 12 30	0,003546	155	30 36.1 15	0,003546
156	18 16.1 10	0,003546	156	12 20 15	0,003546	156	18 14 30	0,003546	156	30.2 33 15	0,003546
157	40.1 39.1 12	0,003546	157	17 22 19	0,003546				157	36.1 36 15	0,003546
158	13.3 12.1 11	0,003546	158	13 20 11	0,003546				158	25 35 15	0,003546
159	27 26 12	0,003546	159	15 19 12	0,003546				159	43.2 38 16	0,003546
160	18 20 10	0,003546	160	13 17 18	0,003546				160	31 35 16	0,003546
161	25.1 35 12	0,003546	161	15 18 19	0,003546				161	30 36 12	0,003546
162	24.1 21 11	0,003546	162	13 19 18	0,003546				162	37.2 37 17	0,003546
163	40.1 21 12	0,003546	163	14 17 16	0,003546				163	26 35 13	0,003546
164	25 20 10	0,003546	164	13 17 15	0,003546				164	28 36 15	0,003546
165	25 22 11	0,003546	165	15 17 17	0,003546				165	25 37 14	0,003546
166	25.1 24 11	0,003546	166	14 19 17	0,003546				166	25 39.3 15	0,003546
167	27.1 21.1 11	0,003546	167	13 17 16	0,003546				167	41.2 40.3 15	0,003546
168	41.2 29 11	0,003546	168	12 21 12	0,003546				168	29 33 14	0,003546
169	23.1 25 12	0,003546	169	13 17 9	0,003546				169	39.2 38 16	0,003546
170	23.1 23 10	0,003546	170	13 20 8	0,003546				170	28 34 14	0,007092
171	27.1 27 11	0,003546	171	16 18 16	0,003546				171	35.2 35 15	0,003546
172	23.1 26 9	0,003546	172	17 20 16	0,003546				172	24 33 13	0,003546
173	27 27 13	0,003546	173	13 15 18	0,003546				173	31 35 15	0,003546
174	18 20.1 10	0,003546	174	15 17 16	0,003546				174	29 37 16	0,007092
175	26.1 24.1 12	0,003546							175	28 36 16	0,003546
176	25 21 12	0,010638							176	27 38.3 16	0,003546
177	36.1 21 12	0,003546							177	28 39 17	0,003546

178	25 22 10	0,003546	178	30 34 14	0,003546
179	40.1 23 11	0,003546	179	35.2 38 14	0,003546
180	44.1 26 12	0,003546	180	26 32 14	0,003546
181	26 29 14	0,003546	181	34.2 34 14	0,003546
182	18 22 10	0,007092	182	28 37 15	0,007092
183	19 20 11	0,003546	183	29 38 14	0,003546
184	28.1 24 12	0,003546	184	36.2 38 15	0,003546
185	13.3 18.1 11	0,003546	185	40.2 41.3 16	0,010638
186	26.1 22 12	0,003546	186	24 39 14	0,003546
187	19 24 9	0,003546	187	26 35 16	0,007092
188	36.1 22 10	0,003546	188	27 39.3 15	0,003546
189	39.1 31 11	0,003546	189	39.2 41.3 15	0,003546
190	23.1 30 12	0,007092	190	35.2 35 14	0,003546
191	24.1 21.1 10	0,003546	191	38.2 38 14	0,003546
192	25.1 26 12	0,003546	192	27 37 16	0,003546
193	19 21.1 12	0,003546	193	33 36 15	0,003546
194	24 23 10	0,003546	194	28 38.3 16	0,003546
195	23.1 21.1 10	0,003546	195	31 37 13	0,003546
196	21.1 21 11	0,003546	196	26 36 13	0,003546
197	31 21 11	0,003546	197	36.2 36 16	0,003546
198	27.1 22 10	0,003546	198	24 31 14	0,003546
199	36.1 24 11	0,003546	199	27 43.3 15	0,003546
200	25.1 20 11	0,003546	200	44.2 38.3 16	0,003546
201	28.1 25 11	0,003546	201	41.2 36 17	0,003546
202	23.1 22 10	0,003546	202	31 33 14	0,003546
203	31 24 11	0,003546	203	25 36 14	0,003546
204	26.1 33 12	0,003546	204	36.2 35 16	0,003546
205	20 21 12	0,003546	205	21 35 14	0,003546
206	37.1 22 12	0,003546	206	24 38.3 15	0,003546
207	25 29 10	0,003546	207	41.2 39.3 16	0,003546

208	28.1 31 12	0,003546	208	29 35 16	0,003546
209	37.1 21 12	0,003546	209	32.2 37 15	0,003546
210	39.1 23 12	0,003546	210	23 36 14	0,003546
211	18 16 11	0,003546			
212	20.1 22.1 11	0,003546			
213	17 21 10	0,007092			
214	23.1 27 11	0,003546			
215	26 23 12	0,003546			
216	24.1 23 10	0,003546			
217	24 19 10	0,003546			
218	28.1 20 11	0,003546			
219	39.1 22 11	0,003546			
220	27.1 22 11	0,003546			
221	17 22 12	0,003546			
222	20 20 11	0,003546			
223	23.1 21.1 11	0,003546			
224	21.1 23 11	0,003546			
225	34.1 30 12	0,003546			
226	22.1 24 10	0,003546			
227	24.1 25 11	0,003546			
228	24.1 25 10	0,003546			
229	28.1 28 11	0,003546			
230	24.1 22.1 10	0,003546			
231	25.1 31 10	0,003546			
232	23.1 19 11	0,003546			
233	24 27 12	0,003546			
234	23 22 11	0,003546			
235	26.1 30 12	0,003546			
236	18 28 11	0,003546			

Fonte: Este trabalho

APÊNDICE B - RELATÓRIO DO FAMLINK X COM OS PERFIS DOS DOIS CASOS GENOTIPADOS E RESULTADO DO FAMLINK X

CASO 1 – “SUPOSTA” AVÓ-NETA –DECAPLEX

Report generated by FamLinkX version 2.9

Reference 1: Kling D, Tillmar A, Egeland T, Mostad P. Int J Legal Med. 2015 Sep;129(5):943-54.

Reference 2: Kling D, Dell'Amico B Tillmar AO. Forensic Sci Int Genet. 2015 Jul;17:1-7.

Timestamp: Wed Jun 16 13:52:34 2021

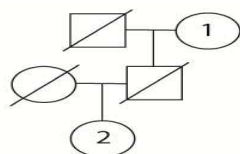
Database: DECAPLEX_SP_BR

File: C:/Users/Bianca/Desktop/Decaplex_SP_BR.sav

Pedigrees

NB! The genders of the individuals depicted in the pedigrees (illustrations) may not correspond to the specified genders. Please see table of persons for correct gender

Pedigree 1 : (Grandmother)



Pedigree 2 : (Unrelated)



Persons

	1. DP-85_MSP (Female)	2. DP-85_F (Female)
DXS8378	10, 10	10, 10
DXS9902	12, 12	12, 12
DXS7132	15, 15.3	13, 15
DXS9898	8.3, 8.3	8.3, 11
DXS6809	30, 32	30, 34
DXS6789	19, 23	19, 21
DXS7133	9, 11	11, 11
GATA172D05	9, 10	9, 12
GATA31E08	13, 14	12, 14
DXS7423	15, 15	14, 15

Results (Exact LR)

Pedigree	lnLikelihood	LR versus pedigree 2 (W)
Pedigree 1	-51.8037	7384 (0.99986458)
Pedigree 2	-60.7108	1 (0.00013541859)

Results (Cluster approach/Merlin)

Pedigree	lnLikelihood	LR versus pedigree 2 (W)
Pedigree 1	-51.8037	7384 (0.99986459)
Pedigree 2	-60.7107	1 (0.00013540815)

LR* exact per marker (Likelihood)

Cluster	Marker	LR for ped 1	LR for ped 2
DXS8378	DXS8378	2.907 (0.0407)	1 (0.014)
DXS9902	DXS9902	3.099 (0.001368)	1 (0.0001518)
DXS7132	DXS7132	1.346 (7.581e-007)	1 (6.251e-008)
DXS9898	DXS9898	3.615 (2.725e-009)	1 (6.216e-011)
DXS6809	DXS6809	4.406 (3.593e-012)	1 (1.86e-014)
DXS6789	DXS6789	5.423 (8.31e-016)	1 (7.933e-019)
DXS7133	DXS7133	1.156 (5.24e-017)	1 (4.328e-020)
GATA172D05	GATA172D05	1.469 (1.258e-019)	1 (7.076e-023)
GATA31E08	GATA31E08	3.357 (5.075e-022)	1 (8.501e-026)
DXS7423	DXS7423	1.237 (3.176e-023)	1 (4.302e-027)

*The LR is computed as the marginal value, i.e. the increase in LR for each consecutive marker in the Markov chain

Settings

Lambda: 1.000

Use cluster specific lambdas: Disabled

Zero mutation rates: Disabled

CASO 1 – “SUPOSTA”AVÓ-NETA – ARGUS

Report generated by FamLinkX version 2.9

Reference 1: Kling D, Tillmar A, Egeland T, Mostad P. Int J Legal Med. 2015 Sep;129(5):943-54.

Reference 2: Kling D, Dell'Amico B Tillmar AO. Forensic Sci Int Genet. 2015 Jul;17:1-7.

Timestamp: Wed Jun 16 11:25:14 2021

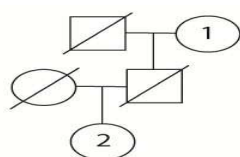
Database: BRASIL-SP

File: C:/Users/Bianca/Desktop/Argus_SP_BR LG1-LG2-LG3-LG4.sav

Pedigrees

NB! The genders of the individuals depicted in the pedigrees (illustrations) may not correspond to the specified genders. Please see table of persons for correct gender

Pedigree 1 : (Grandmother)



Pedigree 2 : (Unrelated)



Persons

	1. 85msp (Female)	2. 85F (Female)
DXS10148	18, 18	18, 40.1
DXS10135	17.1, 19	17.1, 27
DXS8378	10, 10	10, 10
DXS7132	15, 15.3	13, 15
DXS10079	13, 19	17, 19
DXS10074	16, 16	16, 17
DXS10103	17, 19	18, 19
HPRTB	13, 13	13, 14
DXS10101	28.2, 32	28.2, 31.2
DXS10146	27, 28	27, 39.2
DXS10134	36, 38	36, 42.1
DXS7423	15, 15	14, 15

Results (Exact LR)

Pedigree	lnLikelihood	LR versus pedigree 2 (W)
Pedigree 1	-107.1980	3.101e+006 (0.99999968)
Pedigree 2	-122.1452	1 (3.2250029e-007)

Results (Cluster approach/Merlin)

Pedigree	lnLikelihood	LR versus pedigree 2 (W)
Pedigree 1	-107.1617	3.216e+006 (0.99999969)
Pedigree 2	-122.1452	1 (3.1099345e-007)

LR* exact per marker (Likelihood)

Cluster	Marker	LR for ped 1	LR for ped 2
LG1	DXS10148	3.046 (0.0002813)	1 (9.235e-005)
	DXS10135	1.979e+004 (1.027e-013)	1 (1.704e-018)
LG2	DXS8378	2.903 (1.01e-014)	1 (5.771e-020)
	DXS7132	1.348 (5.45e-018)	1 (2.31e-023)
	DXS10079	1.246 (1.404e-023)	1 (4.776e-029)
LG3	DXS10074	4.844 (5.286e-026)	1 (3.711e-032)
	DXS10103	0.7575 (3.122e-028)	1 (2.893e-034)
	HPRTB	1.735 (5.792e-030)	1 (3.095e-036)
	DXS10101	2.585 (9.19e-036)	1 (1.9e-042)
LG4	DXS10146	1.899 (2.871e-039)	1 (3.126e-046)
	DXS10134	0.002613 (1.152e-045)	1 (4.799e-050)
	DXS7423	129.2 (2.783e-047)	1 (8.975e-054)

*The LR is computed as the marginal value, i.e. the increase in LR for each consecutive marker in the

Markov chain

Settings

Lambda: 1.000

Use cluster specific lambdas: Disabled

Zero mutation rates: Disabled

CASO 2 – “SUPOSTAS” MEIAS-IRMÃS -DECAPLEX

Report generated by FamLinkX version 2.9

Reference 1: Kling D, Tillmar A, Egeland T, Mostad P. Int J Legal Med. 2015 Sep;129(5):943-54.

Reference 2: Kling D, Dell'Amico B Tillmar AO. Forensic Sci Int Genet. 2015 Jul;17:1-7.

Timestamp: Wed Jun 16 14:06:45 2021

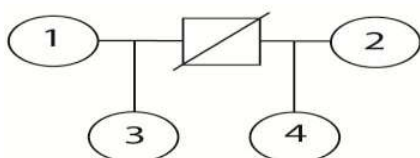
Database: DECAPLEX_SP_BR

File: C:/Users/Bianca/Desktop/Decaplex_SP_BR.sav

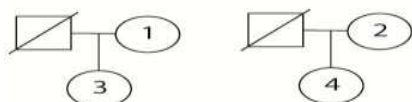
Pedigrees

NB! The genders of the individuals depicted in the pedigrees (illustrations) may not correspond to the specified genders. Please see table of persons for correct gender

Pedigree 1 : (Half Siblings (Data mothers))



Pedigree 2 : (Unrelated (Data mothers))



Persons

	1. 729_MFB (Female)	2. 729_M (Female)	3. 729_FB2 (Female)	4. 729_F (Female)
DXS8378	11, 12	11, 12	10, 12	10, 12
DXS9902	12, 12	13, 14	12, 13	13, 14
DXS7132	13, 15	13, 14	13, 15	13, 14
DXS9898	8.3, 12	12, 13	12, 12	12, 13
DXS6809	33, 34	31, 31	32, 34	31, 32
DXS6789	22, 22	20, 22	22, 22	20, 22
DXS7133	9, 10	10, 11	10, 10	10, 11
GATA172D05	10, 10	6, 11	10, 11	6, 11
GATA31E08	9, 13	11, 12	13, 13	11, 13
DXS7423	15, 15	13, 14	15, 15	14, 15

Results (Exact LR)

Pedigree	lnLikelihood	LR versus pedigree 2 (W)
Pedigree 1	-77.6088	2.535e+004 (0.99996055)
Pedigree 2	-87.7492	1 (3.9453665e-005)

Results (Cluster approach/Merlin)

Pedigree	lnLikelihood	LR versus pedigree 2 (W)
Pedigree 1	-77.6086	2.535e+004 (0.99996055)
Pedigree 2	-87.7491	1 (3.9450711e-005)

LR* exact per marker (Likelihood)

Cluster	Marker	LR for ped 1	LR for ped 2
DXS8378	DXS8378	2.907 (0.002891)	1 (0.0009943)
DXS9902	DXS9902	4.548 (9.66e-008)	1 (7.306e-009)
DXS7132	DXS7132	1 (1.247e-010)	1 (9.428e-012)
DXS9898	DXS9898	2 (8.868e-014)	1 (3.353e-015)
DXS6809	DXS6809	8.545 (1.757e-017)	1 (7.775e-020)
DXS6789	DXS6789	1.819 (4.444e-021)	1 (1.081e-023)
DXS7133	DXS7133	1.72 (2.14e-024)	1 (3.027e-027)
GATA172D05	GATA172D05	3.032 (5.311e-028)	1 (2.478e-031)
GATA31E08	GATA31E08	4.78 (1.998e-031)	1 (1.95e-035)
DXS7423	DXS7423	2.474 (1.972e-034)	1 (7.781e-039)

*The LR is computed as the marginal value, i.e. the increase in LR for each consecutive marker in the Markov chain

Settings

Lambda: 1.000

Use cluster specific lambdas: Disabled

Zero mutation rates: Disabled

CASO 2 – “SUPOSTAS” MEIAS-IRMÃS - ARGUS

Report generated by FamLinkX version 2.9

Reference 1: Kling D, Tillmar A, Egeland T, Mostad P. Int J Legal Med. 2015 Sep;129(5):943-54.

Reference 2: Kling D, Dell'Amico B Tillmar AO. Forensic Sci Int Genet. 2015 Jul;17:1-7.

Timestamp: Wed Jun 16 12:04:11 2021

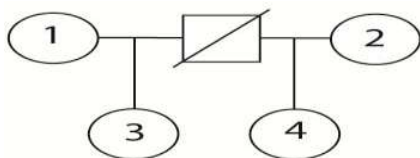
Database: BRASIL-SP

File: C:/Users/Bianca/Desktop/Argus_SP_BR LG1-LG2-LG3-LG4.sav

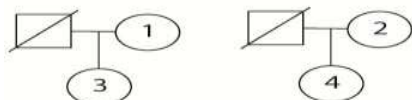
Pedigrees

NB! The genders of the individuals depicted in the pedigrees (illustrations) may not correspond to the specified genders. Please see table of persons for correct gender

Pedigree 1 : (Half Siblings (Data mothers))



Pedigree 2 : (Unrelated (Data mothers))



Persons

	1. 729m (Female)	2. 729mfb (Female)	3. 729f (Female)	4. 729fb2 (Female)
DXS10148	18, 24	23, 28.1	18, 18	18, 23
DXS10135	19.1, 26	19, 19	19.1, 23	19, 23
DXS8378	11, 12	11, 12	10, 12	10, 12
DXS7132	13, 14	13, 15	13, 14	13, 15
DXS10079	21, 21	18, 21	20, 21	18, 20
DXS10074	14, 16	8, 17	7, 16	7, 17
DXS10103	16, 19	19, 19	19, 19	19, 19
HPRTB	12, 15	11, 12	12, 13	12, 13
DXS10101	29, 33	29.2, 30.2	29, 32	30.2, 32
DXS10146	25, 33	28, 29	25, 28	28, 29
DXS10134	37, 37	34, 34	32, 37	32, 34
DXS7423	13, 14	15, 15	14, 15	15, 15

Results (Exact LR)

Pedigree	lnLikelihood	LR versus
----------	--------------	-----------

		pedigree 2 (W)
Pedigree 1	-159.6800	2.491e+007 (0.99999996)
Pedigree 2	-176.7109	1 (4.0140578e-008)

Results (Cluster approach/Merlin)

Pedigree	lnLikelihood	LR versus pedigree 2 (W)
Pedigree 1	-168.0093	5.264e+006 (0.99999981)
Pedigree 2	-183.4858	1 (1.89953e-007)

LR* exact per marker (Likelihood)

Cluster	Marker	LR for ped 1	LR for ped 2
LG1	DXS10148	6.092 (4.687e-007)	1 (7.694e-008)
	DXS10135	45.21 (3.825e-015)	1 (1.389e-017)
	DXS8378	1.006 (4.6e-022)	1 (1.659e-024)
LG2	DXS7132	1.007 (5.703e-025)	1 (2.044e-027)
	DXS10079	1.093 (5.922e-029)	1 (1.941e-031)
LG3	DXS10074	0.03437 (9.203e-039)	1 (8.774e-040)
	DXS10103	3.023 (2.226e-041)	1 (7.019e-043)
	HPRTB	3.248 (2.505e-047)	1 (2.432e-049)
LG4	DXS10101	4.752 (4.539e-051)	1 (9.274e-054)
	DXS10146	4.076 (8.493e-057)	1 (4.257e-060)
	DXS10134	34.25 (2.691e-061)	1 (3.937e-066)
	DXS7423	364.6 (4.486e-070)	1 (1.801e-077)

*The LR is computed as the marginal value, i.e. the increase in LR for each consecutive marker in the Markov chain

Settings

Lambda: 1.000

Use cluster specific lambdas: Disabled

Zero mutation rates: Disabled

ANEXOS

ANEXO A - APROVAÇÃO DO COMITÊ DE ÉTICA EM PESQUISA COM SERES HUMANOS



PARECER CONSUBSTANCIADO DO CEP

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: AVALIAÇÃO DE DIFERENTES CONJUNTOS DE MARCADORES DO CROMOSSOMO X NA POPULAÇÃO RESIDENTE DO ESTADO DE SÃO PAULO

Pesquisador: BIANCA BELON JANUARIO

Área Temática: Genética Humana:

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

Versão: 2

CAAE: 12150919.0.0000.5426

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

Patrocinador Principal: Financiamento Próprio

DADOS DO PARECER

Número do Parecer: 3.456.569

Apresentação do Projeto:

A aplicabilidade dos marcadores X-STRs é ampla e importante para a genética forense e populacional, pois esses marcadores são utilizados em casos com incompatibilidade entre filha e suposto pai, casos em que há dificuldade de análise do material biológico, em casos de maternidade e até em casos onde há violência fetal. Por isso, coordenados com os STRs autossômicos, os marcadores X-STRs podem superar as imitações da análise de marcadores tradicionais, solucionando casos tratados como inconclusivos. Neste estudo, será inédita a identificação de variabilidade genética de alguns dos marcadores do X-STRs com interesse forense utilizando e comparando conjuntos diferentes de marcadores. Determinar a variabilidade genética avaliando o poder discriminatório dos painéis de marcadores; Calcular o poder de discriminação e o poder de exclusão em duos e trios destes dois painéis de marcadores em conjunto e separadamente; Avaliar o desequilíbrio de ligação entre os pares de marcadores destes dois painéis; Determinar as frequências alélicas e a taxa de mutação destes marcadores; A partir dos dados obtidos, comparar a população em estudo com outras do Brasil e do mundo, e calcular a distância genética F_{st} .

Objetivo da Pesquisa:

Avaliar a funcionalidade dos conjuntos de marcadores Decaplex (10 X-STRs) e do kit Investigator

Endereço: Rodovia Araraquara Jaú, km 1

Bairro: Campus Universitário

UF: SP

Município: ARARAQUARA

CEP: 14.801-902

Telefone: (16)3301-4657

E-mail: sta@fctar.unesp.br



Continuação do Parecer: 3.456.569

Argus X-12QS verificando a variabilidade genética e a eficiência forense na população estudada.

Avaliação dos Riscos e Benefícios:

Os riscos e benefícios da pesquisa foram apresentados adequadamente.

Riscos: a coleta de sangue que será realizada com material estéril, será necessário que o participante disponibilize uma gota de sangue do dedo indicador coletada por caneta lancetadora G.Tech. A previsão de riscos é mínima, ou seja, consiste em leve dor após a introdução da lanceta automática. O material será utilizado exclusivamente para essa pesquisa, não podendo ser reutilizado em pesquisa posterior. duração máxima de 10 minutos.

Benefícios: não há benefício direto ao participante, apenas a contribuição voluntária para a obtenção de dados genéticos na população brasileira

Comentários e Considerações sobre a Pesquisa:

A pesquisa é relevante para a obtenção de dados genéticos da população do estado de São Paulo.

Considerações sobre os Termos de apresentação obrigatória:

O TCLE está adequado e de acordo com a legislação vigente

Conclusões ou Pendências e Lista de Inadequações:

O Vice-coordenador do Comitê de Ética em Pesquisa da Faculdade de Ciências Farmacêuticas analisou e aprovou "ad referendum" o projeto de pesquisa em questão. Os relatórios parciais deverão ser entregues em JANEIRO de 2020 e JULHO de 2020 e o Relatório Final junto aos Termos de Consentimento Livre Esclarecido (originais e assinados em todas as folhas) deverão ser entregues em NOVEMBRO de 2020.

Considerações Finais a critério do CEP:

Este parecer foi elaborado baseado nos documentos abaixo relacionados:

Tipo Documento	Arquivo	Postagem	Autor	Situação
Informações Básicas do Projeto	PB_INFORMAÇÕES_BÁSICAS_DO_PROJETO_1287204.pdf	26/06/2019 11:08:20		Aceito
Outros	carta_resposta.pdf	26/06/2019 11:04:59	BIANCA BELON JANUARIO	Aceito
TCLE / Termos de Assentimento /	TCLECrX_corr.pdf	26/06/2019 10:59:38	BIANCA BELON JANUARIO	Aceito

Endereço: Rodovia Araraquara Jaú, km 1

Bairro: Campus Universitário

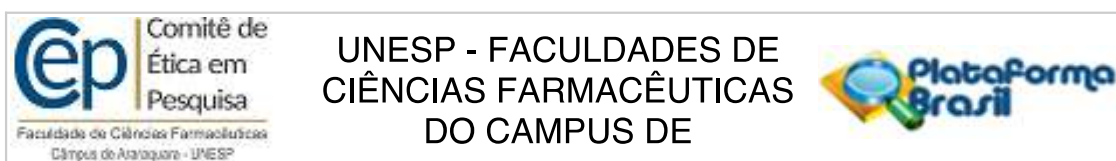
CEP: 14.801-902

UF: SP

Município: ARARAQUARA

Telefone: (16)3301-4657

E-mail: sta@fcar.unesp.br



Continuação do Parecer: 3.456.569

Justificativa de Ausência	TCLECrX_corr.pdf	26/06/2019 10:59:38	BIANCA BELON JANUARIO	Aceito
Projeto Detalhado / Brochura Investigador	Projeto_comite_v2.pdf	18/04/2019 10:39:48	BIANCA BELON JANUARIO	Aceito
Folha de Rosto	folha_de_rosto.pdf	29/01/2019 15:52:17	BIANCA BELON JANUARIO	Aceito

Situação do Parecer:

Aprovado

Necessita Apreciação da CONEP:

Não

ARARAQUARA, 16 de Julho de 2019

Assinado por:
PAULO INACIO DA COSTA
(Coordenador(a))

Endereço: Rodovia Araraquara Jaú, km 1
Bairro: Campus Universitário **CEP:** 14.801-902
UF: SP **Município:** ARARAQUARA
Telefone: (16)3301-4657 **E-mail:** sta@fcar.unesp.br

ANEXO B – TRABALHOS PUBLICADOS



Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Paternal and maternal mutations in X-STRs: A GHEP-ISFG collaborative study



Nádia Pinto^{a,b,c,*}, Vânia Pereira^d, Carmen Tomas^d, Silvia Loiola^e, Elizeu F. Carvalho^e, Nidia Modesti^f, Mariana Maxzud^f, Valeria Marcucci^g, Hortensia Cano^g, Regina Cicarelli^h, Bianca Januario^h, Ana Bentoⁱ, Pedro Britoⁱ, Germán Burgosⁱ, Elius Paz-Cruz^k, Laura Díez-Juárez^l, Silvia Vannelli^m, Maria de Lurdes Pontesⁿ, Gabriela Berardi^o, Sandra Furfuro^p, Alberto Fernandez^q, Denilce Sumita^r, Cecilia Bobillo^s, Maria Gabriela García^t, Leonor Gusmão^e

^a Institute of Pathology and Molecular Immunology from University of Porto (IPATIMUP), Portugal

^b Instituto de Investigação e Inovação em Saúde, I3S, Universidade do Porto, Portugal

^c CMUP, Centro de Matemática da Universidade do Porto, Portugal

^d Section of Forensic Genetics, Department of Forensic Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Denmark

^e Laboratório de Diagnóstico por DNA (LDD), Universidade do Estado do Rio de Janeiro, Brazil

^f Centro de Genética Forense, Poder Judicial de Córdoba, Argentina

^g Laboratorio Regional de Investigación Forense, Tribunal Superior de Justicia de Santa Cruz, Argentina

^h UNESP-Universidade Estadual Paulista, Faculdade de Ciências Farmacêuticas, Laboratório de Investigação de Paternidade-NAC, São Paulo, Brazil

ⁱ Instituto Nacional de Medicina Legal e Ciências Forenses, I.P., Serviço de Genética e Biologia Forenses, Delegação do Centro, Portugal

^j Escuela de Medicina, Facultad de Ciencias de la Salud, Universidad de Las Américas (UDLA), Quito, Ecuador

^k Laboratorio de ADN de la Fiscalía General del Estado, Quito, Ecuador

^l Departamento de Biología, Servicio de Criminalística de la Guardia Civil, Spain

^m Laboratorio Regional de Genética Forense, Poder Judicial de Río Negro, Argentina

ⁿ Instituto Nacional de Medicina Legal e Ciências Forenses, I.P., Serviço de Genética e Biologia Forenses, Delegação do Norte, Portugal

^o PRICAI-Fundación Favaloro, Buenos Aires, Argentina

^p Laboratorio de Análisis de ADN Facultad de Ciencias Médicas Universidad Nacional de Cuyo, Mendoza, Argentina

^q LabGenetics: Laboratorio de Genética Clínica S.L., Madrid, Spain

^r Genomic Engenharia Molecular Molecular, São Paulo, Brazil

^s Laboratorio de Genética Forense, Poder Judicial de la Provincia de La Pampa, Argentina

^t Laboratorio MANLAB, Área de Filiaciones, Buenos Aires, Argentina

ARTICLE INFO

Keywords:

Argus kit
Mutation rate
X chromosome
Population database

ABSTRACT

The GHEP-ISFG organized a collaborative study to estimate mutation rates for the markers included in the Investigator Argus X-12 QS kit Qiagen. A total of 16 laboratories gathered data from 1,612 father/mother/daughter trios, which were used to estimate both maternal and paternal mutation rates, when pooled together with other already published data. Data on fathers and mothers' age at the time of birth of the daughter were also available for ~93 % of the cases. Population analyses were computed considering the genetic information of a subset of 1,327 unrelated daughters, corresponding to 2,654 haplotypes from residents in several regions of five countries: Argentina, Brazil, Ecuador, Portugal and Spain. Genetic differentiation analyses between the population samples from the same country did not reveal signs of significant stratification, although results from Hardy-Weinberg and linkage disequilibrium tests indicated the need of larger studies for Ecuador and Brazilian populations. The high genetic diversity of the markers resulted in a large number of haplotype combinations, showing the need of huge databases for reliable estimates of their frequencies.

It should also be noted the high number of new alleles found, many of them not included in the allelic ladders provided with the kit, as very diverse populations were analyzed. The overall estimates for locus specific average mutation rates varied between 7.5E-04 (for DXS7423) and 1.1E-02 (for DXS10135), the latter being a troublesome figure for kinship analyses. Most of the found mutations (~92 %) are compatible with the gain or loss of a single repeat. Paternal mutation rates showed to be 5.2 times higher than maternal ones. We also found that older fathers were more prone to transmit mutated alleles, having this trend not been observed in the case of the mothers.

* Corresponding author at: IPATIMUP/i3S, R. Alfredo Allen, 4200-135 Porto, Portugal.

E-mail addresses: npinto@ipatimup.pt, nmgapinto@gmail.com (N. Pinto).

<https://doi.org/10.1016/j.fsigen.2020.102258>

Received 19 June 2019; Received in revised form 6 October 2019; Accepted 29 January 2020

Available online 05 February 2020

1872-4973/ © 2020 Elsevier B.V. All rights reserved.

1. Introduction

The analysis of markers located in the X chromosome can be useful to solve complex kinship cases in forensic genetics [1–5]. For this purpose, different genotyping methods have been described, including sets of X-chromosomal specific STRs, Indels and SNPs (e.g [6–14]).

In forensic genetics, the most widely used X-STR kit is the Investigator® Argus X-12 kit (Qiagen, Hilden, Germany), which comprises 12 markers organized in 4 linkage groups. Population data for this kit have been gathered from many populations around the world (e.g. [15–27]), although haplotype reference databases are still lacking for forensic use in many countries.

Because linked loci are more prone to show linkage disequilibrium (LD) in populations, a proper application of marker sets in the same chromosome usually requires much larger population databases than for unlinked markers. In the presence of LD, the need of large databases also increases with the inclusion of highly variable STRs, since it is necessary to estimate frequencies for many rare haplotypes that are absent or poorly represented in reference population samples.

Large databases are also important to detect LD between markers [5,28]. Since loci included in the Argus kit are closely located (three markers in each linkage group), and the X chromosome only recombines in female meiosis, it is expected to find higher LD levels between markers. Using haplotypic data from Swedish males, Kling et al. [28] showed a high probability of not detecting LD in samples of less than 400 male individuals, and LD could not be detected for linkage group 2 in 17 % of the cases where samples with 600 males were considered.

Moreover, for the application of X-STRs in kinship analyses, it is important to have reliable estimates of recombination and mutation rates [4,29]. In both cases, it is necessary to study allele/haplotype segregation between relatives. Mutation rates are usually estimated through the proportion of allele transfers resulting in mendelian incompatibilities in duos or trios from paternity cases or using information from large pedigrees. As in the current study, most approaches rely on the number of observed rather than occurring mutations. Indeed, these estimates are conservative as mutations can occur without leading to Mendelian incompatibilities – the so called “hidden” or “covered” mutations. For biases affecting these estimates, statistical adjustments were presented for autosomal markers [30–32]. The bias is greater when duos, instead of trios, are analysed and, in the case of X-chromosomal transmission, bias is greater when estimating mutation in mother-daughter than in father-daughter or mother-son duos [54].

Germline mutations were investigated for different X-STRs, and population and marker-specific rates were reported, although, in most cases, a limited amount of data prevents statistically accessing significant differences among markers or populations (e.g. [25,33–36]). To date, few data are available for the full set of 12 X-STRs from the Argus kit. Two studies reported average mutation rates of 3.3×10^{-3} and 5.1×10^{-3} per allele transmission in 513 (mother/son, mother/daughter) and 345 (father-daughter) duos, respectively [25,37]. The different observed values can be explained by a significantly higher male than female germline mutation rate, which was found by Tomas et al. [25].

For the reasons mentioned above, available data on X-STRs of the Argus kit are still insufficient to obtain good estimates of forensically relevant parameters.

Therefore, the Spanish and Portuguese speaking working group of the International Society for Forensic Genetics (GHEP-ISFG) organized a collaborative study to collect data for the 12 X-STR markers included in the Investigator Argus X-12 QS kit. This study aimed at improving the quality of the estimates that are relevant in forensic genetics, namely haplotype frequencies in different populations, LD between markers, and locus-specific mutation rates.

2. Material and methods

After approval at the general assembly, a working group was

Table 1

Population of origin and distribution of the 1,612 analyzed trios.

	Population	N° of trios ^{a)}	N° of unrelated haplotypes
Africa	Somalia - General population	73	b)
America	Greenland - General population	104	b)
	Argentina - Santa Cruz	140	250
	Argentina - Mendoza	70	140
	Argentina - Rio Negro	74	142
	Argentina - La Pampa	69	136
	Argentina - Buenos Aires	70	140
	Argentina - Cordoba	143	286
	Brazil - Rio de Janeiro	149	266
	Brazil - General population	70	140
	Brazil - São Paulo	141	282
	Ecuador - General population	180	300
Europe	Portugal - General population	140	278
	Portugal - North region	72	144
	Spain - General population	106	150
	Other populations	11	c)
	Total	1,612	2,654

a) Includes 97 family clusters; b) Haplotype data from Somalia and Greenland were not considered for population analyses since they are partially published [25]; c) Haplotype data were not considered for population analyses since concerns to few information scattered by several populations.

created with the purpose of carrying out the present study, which was open to all members of the GHEP-ISFG. The participating laboratories had to: (i) present a certificate of the proficiency test of the GHEP-ISFG, showing correct results for the Argus kit markers; and (ii) use the latest version of the Argus kit - Investigator® Argus X-12 QS (Qiagen) - to genotype at least 70 father/mother/daughter trios.

2.1. Samples collection and genotyping

A total of 16 laboratories participated in this study, providing data from 1,612 father/mother/daughter trios. The trios were collected from 15 populations from Africa, America and Europe (Table 1). Most samples belonged to paternity cases, although some healthy volunteers were also included in the samples from Santa Cruz (Argentina), Rio Negro (Argentina), and Spain. In most cases, the samples encompassed residents in each of the 15 analyzed populations. In samples from paternity investigations, the biological relationship was previously confirmed using autosomal STRs ($LR > 10^5$). Each laboratory ensured the anonymization of the samples and the accomplishment of the legal and ethical requirements for their use in this research project.

All samples were genotyped using the Investigator Argus X-12 QS (Qiagen), following manufacturer's instructions (Investigator® Argus X-12 QS Handbook), except for 299 samples where laboratories already had profiles obtained with the older version of the kit. In these cases, only profiles that were heterozygote for the markers DXS10101, DXS10146 and DXS10148 were accepted. The remaining samples had to be retyped with the new version of the kit to avoid null alleles. Indeed, according to the manufacturer, additional primers were included in the new version to prevent the high frequency of null alleles observed for these three markers in African populations [38].

2.2. Data analysis

A total of 1,612 father/mother/daughter trios, genotyped for 12 X-STRs, were collected and used to estimate locus specific mutation rates. Confidence intervals for mutation rates were estimated from the binomial standard deviation. Information on the parents' age at the time of birth of the daughter was available for ~93 % of the trios. A total of

1,327 unrelated females were selected among the daughters, and gametic phase was determined using the haplotype information from the father (Table 1). These unrelated females were used for population analyses and to calculate forensic statistics. In each family cluster, only one (randomly selected) daughter was considered for population analyses.

Estimation of allele and haplotype frequencies and pairwise F_{ST} genetic distances were calculated using the software Arlequin ver 3.5.2.2 [39]. The same software was used to test population differentiation, Hardy-Weinberg equilibrium and pairwise linkage disequilibrium between loci. The significance level of 0.05 was adjusted by applying the Bonferroni's correction for multiple tests, namely by considering the number of markers used to test Hardy-Weinberg equilibrium (0.05/12), and the total number of pairwise comparisons in population differentiation (0.05/78) and LD (0.05/66) tests.

Statistics for forensic efficiency evaluation, namely mean exclusion chance in trios involving daughters (MECT) as well as in father/daughter duos (MECD), power of discrimination in females (PDF) and in males (PDM) were calculated using the formulae from Desmarais et al. [40].

3. Results and discussion

3.1. Population genetics analysis

A sample of 1,327 unrelated females was selected for population analyses. By selecting the daughters from the studied trios, it was possible to determine the genotype gametic phase, and to create a database comprising 2,654 unrelated haplotypes.

Pairwise comparisons for the 12 X-STRs between all samples showed low, non-statistically significant differences among populations from the same country, namely Argentina, Brazil and Portugal ($F_{ST} \leq 0.0023$; $p \geq 0.0239$; Supplementary Table S1A). Non-significant differences were also observed between the samples from Portugal and Spain ($F_{ST} \leq -0.0012$; $p \geq 0.5388$). Similar results were obtained when performing the test for markers in each linkage group (LG1 to LG4), separately (Supplementary Tables S1B-E). When comparing samples from different countries, the larger distances were observed for markers at LG2 and LG3, and Ecuador was the population that most

departure from the remaining.

Therefore, Hardy-Weinberg equilibrium and LD tests were performed after pooling the samples from the same country, as well as by joining Portuguese and Spanish samples in a single population from Iberia.

3.1.1. Hardy-Weinberg equilibrium analysis

No statistically significant deviations from the Hardy-Weinberg equilibrium (HWE) expectations were detected in the Argentinian sample (Supplementary Table S2), indicating no signs of population stratification for the significance level considered, which supports the use of a single database for the analysed markers.

Concerning the samples from Brazil, a significant departure from the HWE was observed at DXS10134 locus (Supplementary Table S2). However, the observed deviation was not significant when analysing separately the three samples from Brazil (general population), Rio de Janeiro and São Paulo. Therefore, these results indicate that larger studies on Brazilian populations are required to investigate if a common database can be used for these markers.

Ecuador did not reveal statistically significant deviations from HWE, for the 12 X-STRs.

The HWE test showed a statistically significant deviation for DXS10079 in the Iberian population sample (Supplementary Table S2), due to a lower frequency of heterozygotes than the expected (observed 0.787 and expected 0.828). When the test was performed for the Spanish and Portuguese populations separately, the observed and expected values of heterozygosity were almost the same (0.787 and 0.838 for Spain, 0.787 and 0.825 for Portugal), although the deviation was not statistically significant due to a lower sample size. Therefore, the results indicate that the observed excess of homozygotes is not due to differences between Spain and Portugal and can most likely be explained by the presence of undetected null alleles in Iberian populations.

3.1.2. Linkage disequilibrium analysis

Pairwise linkage disequilibrium analysis was performed in all populations: Argentina with a total of 1,094 haplotypes, Brazil 688 haplotypes, Ecuador 300 haplotypes, and Iberia 572 haplotypes. The results are shown in Supplementary Table S3.

Table 2

Forensic efficiency statistics for each linkage group and overall values for the Argus system: Mean exclusion chance in trios involving daughters (MECT) and in father/daughter duos (MECD); power of discrimination in males (PDM) and in females (PDF).

	MECT	MECD	PDF	PDM
Argentina				
LG1	99.615 %	99.236 %	99.997 %	99.617 %
LG2	99.337 %	98.691 %	99.991 %	99.341 %
LG3	98.590 %	97.254 %	99.963 %	98.608 %
LG4	99.427 %	98.866 %	99.994 %	99.430 %
Total	99.999998 %	99.99997 %	99.99999999999994 %	99.999998 %
Brazil				
LG1	99.708 %	99.418 %	99.998 %	99.709 %
LG2	99.380 %	98.774 %	99.992 %	99.384 %
LG3	99.237 %	98.494 %	99.989 %	99.243 %
LG4	99.570 %	99.147 %	99.996 %	99.572 %
Total	99.9999994 %	99.999991 %	99.99999999999994 %	99.9999994 %
Ecuador				
LG1	99.165 %	98.353 %	99.986 %	99.171 %
LG2	98.917 %	97.874 %	99.977 %	98.928 %
LG3	96.666 %	93.710 %	99.805 %	96.756 %
LG4	99.069 %	98.168 %	99.983 %	99.078 %
Total	99.99997 %	99.9996 %	99.99999999990 %	99.99997 %
Iberia				
LG1	99.631 %	99.266 %	99.997 %	99.632 %
LG2	99.080 %	98.190 %	99.984 %	99.088 %
LG3	98.924 %	97.887 %	99.978 %	98.935 %
LG4	99.396 %	98.805 %	99.993 %	99.400 %
Total	99.999998 %	99.99997 %	99.9999999999993 %	99.999998 %

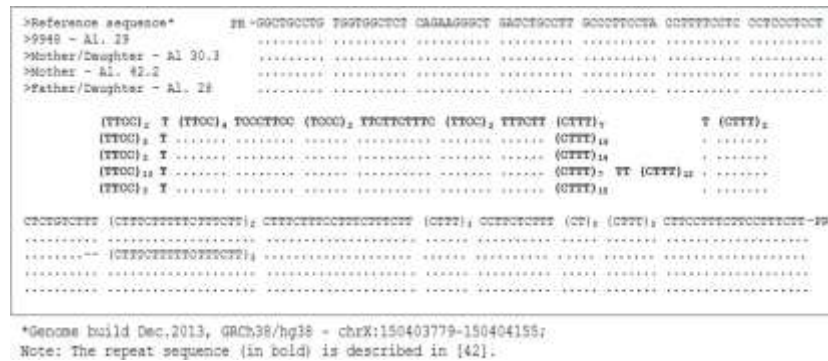


Fig. 1. Sequencing results obtained for DXS10146 alleles in a trio showing a Mendelian incompatibility due to a null allele shared by the mother and the daughter.

Associations between markers from the same linkage group were detected in all populations, although in Ecuador it was only detected inside LG1. This can be explained by its lower sample size, which reduces the chance of detecting the presence of LD [28]. Moreover, Ecuador revealed significant associations in 3 pairwise comparisons involving markers from different linkage groups. LD between non-linked markers can occur due to population stratification, which is also associated with an excess of homozygotes. The average value of observed heterozygotes in the Ecuadorian sample was indeed lower than the expected for a population in Hardy-Weinberg equilibrium (0.771 and 0.796, respectively), although deviations were not statistically significant, which can, once more, be due to the small sample size (150 genotypes). In conclusion, the overall results do not allow excluding the presence of a genetic stratification inside the sample from Ecuador, that needs to be investigated in more detail in larger samples with well-defined origin.

3.1.3. Haplotype frequencies and forensic parameters

In Supplementary Table S4, we present the haplotype frequencies per linkage group for each population group: Argentina, Brazil, Ecuador and Iberia. The parameters of forensic interest, calculated for each linkage group and the overall values for the 12 X-STRs included in the Investigator Argus X-12 QS kit, are presented in Table 2. The population from Brazil showed the highest values of mean exclusion chance and discrimination capacity, followed by Argentina and Iberia, with similar results. Ecuador was the population with the lowest values of forensic efficiency for the investigated set of X-STR markers.

3.2. Analysis of mutations

In a total of 38,688 allele transfers, 157 Mendelian incompatibilities were found, one of which explained either by mutation or by the presence of a silent allele (Supplementary Table S5). In this case, at DXS10146 locus, the mother was 42.2 and both the father and the daughter were 28. Therefore, a mutation from an allele 42.2 to 28 would be necessary to explain the genotypic configurations, being more plausible to assume the presence of a null allele. To test this hypothesis, the DXS10146 locus was amplified in singleplex, using the PCR amplification protocol described in [41]. The use of the new primers revealed no Mendelian incompatibilities among the trio, with both the mother and the daughter being heterozygotes for the DXS10146 locus. Sequencing results are presented in Fig. 1, showing a TT deletion and a 17 bp insertion, at the repeat flanking region, 8 bases downstream of the repeat.

Our results were analysed together with others already published for 9 out of the 12 X-STRs from this study, for father/mother/daughter trios [33,42–53] (Table 3). Although there are other reports on X-STR mutations, they consider only duos or a mixture of different pedigrees (not susceptible of being discriminated), preventing a joint analysis

Table 3

Number of mutations considering the parental origin and the minimal number of gains (+) or losses (–) of repeats or bases that can explain the inconsistencies between parental and filial genotypes. Data from our study and from others previously published [33,42–53] were analysed together.

Most Parsimonious Parental origin	Type of mutation					Total	
	The same (repeat variation)						Different (bp variation)
	{–2}	{–1}	{+1}	{+2}	{–1} V {+1}		
Paternal	2	75	54	2	0	134	
Maternal	0	8	15	1	1	26	
Undetermined	0	4	11	0	6	21	
Total	2	87	80	3	7	181	

with our data.

Out of the 181 mutations, 174 are compatible with the gain or loss of a single repeat, 5 with the gain or loss of two repeats, and two by the gain of one base pair. In 7 cases, inconsistencies are explainable by either the gain or loss of a single repeat. It is worth nothing that, in accordance with the other mutation rate studies, we assumed single-step mutations whenever they explain the genotypic configurations, which means that multistep mutations may be underestimated [43]. In a recent study, simulations of two-step mutations in 8 X-STRs showed that 14 % of these would be attributed to a single-step mutation [54].

Considering the parental origin, we found 5.2 more paternal than maternal mutations. A total of 134 out of the 181 genotypic configurations were compatible with a paternal mutation, 26 with a maternal one, and for 21 cases both origins are possible (Table 3). In humans, it is well known that mutations are more frequent in paternal than in maternal germline [55], and X-STR studies agree with this [25,43,46]. However, the increase of mutation rate with age is higher in males than in females, and thus, the ratio between paternal and maternal mutations is age-dependent [56]. This prevents a straightforward comparison of the estimate of ratio of male to female mutations we observed with those from studies with no age information available.

In our dataset, we found one case of simultaneous paternal and maternal mutations at the same locus and, in five trios, mutations were observed in two different loci.

3.2.1. Mutation rates and parental age

From the 1,612 analyzed trios, age information at the year of the daughter's birth was obtained for 1,492 mothers and 1,488 fathers. The age distribution, as well as the proportion of cases resulting in at least one mutation, are presented in Fig. 2. Globally, fathers tend to be older than mothers (average age at time of the birth of the daughter ~30.1

and ~25.8 years, respectively), and show higher mutation rates, the likelihood of which seems to increase also with the age. Particularly, note the case of the fathers between 51 and 55 years-old, for which at least one undoubted paternal mutation was observed in nearly 14 % of the meioses. In mothers, no correlation between age and mutation rate was detected, maybe because it is too subtle to be perceived from the currently available data, as suggested by Ségurel et al. [56].

The increase with age of male germline mutations is well documented (e.g. [55–57]) and it should be accounted for likelihood calculations in kinship cases. Nevertheless, as for other parameters that also influence mutation rates in STRs (namely repeat structure and allele length, and the sex), a large amount of information is required for age specific estimates, by joining data from different studies. For the X-STRs included in this study, this is still not possible due to low amount and large heterogeneity of the information published on mutation rates.

3.2.2. Locus-specific mutations

After pooling our data with other available in the literature for father/mother/daughter trios [33,42–53], mutation rates and their 95 % confidence intervals were estimated for each marker (Table 4). Table 4 also shows the expected paternal and maternal mutation rates estimated resorting to trios and after the proportional distribution of the mutations with undetermined origin.

Detailed information on paternal and maternal allele transmissions in our samples is provided in supplementary material (Tables S5–S7). Globally, observed mutation rates varied from 7.5E-04 (for DXS7423)

to 1.1E-02 (for DXS10135).

Note that the 21 observed mutations with undetermined origin (more than 10 % of the total) are not detected when father/daughter or mother/daughter duos are analyzed. Indeed, it is known that the use of duos instead of trios increases the possibility of hidden mutations and, consequently, resulting in a greater bias in estimates of actual mutation rates. As expected, in our data, the number of observed mutations in trios were higher or equal to those found in duos (Table S6), since the consideration of the other parent may uncover mutations. The greatest difference between the mutation rates estimated through duos or trios was achieved for the HPRTB marker, for both males and females. For this marker the estimated mutation rate almost doubled when trios instead of duos were considered.

4. Conclusions

The present collaborative study of the GHEP-ISFG working group allowed the compilation of a large amount of population and segregation data for the 12 X-STRs most widely used in forensics, contributing with valuable information for their application in kinship investigations.

In this study, haplotype data were obtained for different populations from Argentina, Brazil, Ecuador, Portugal and Spain. Differentiation analyses, Hardy-Weinberg equilibrium and linkage disequilibrium tests did not reveal signs of significant stratification in Argentina, supporting the use of a national database. The same was found between Portuguese

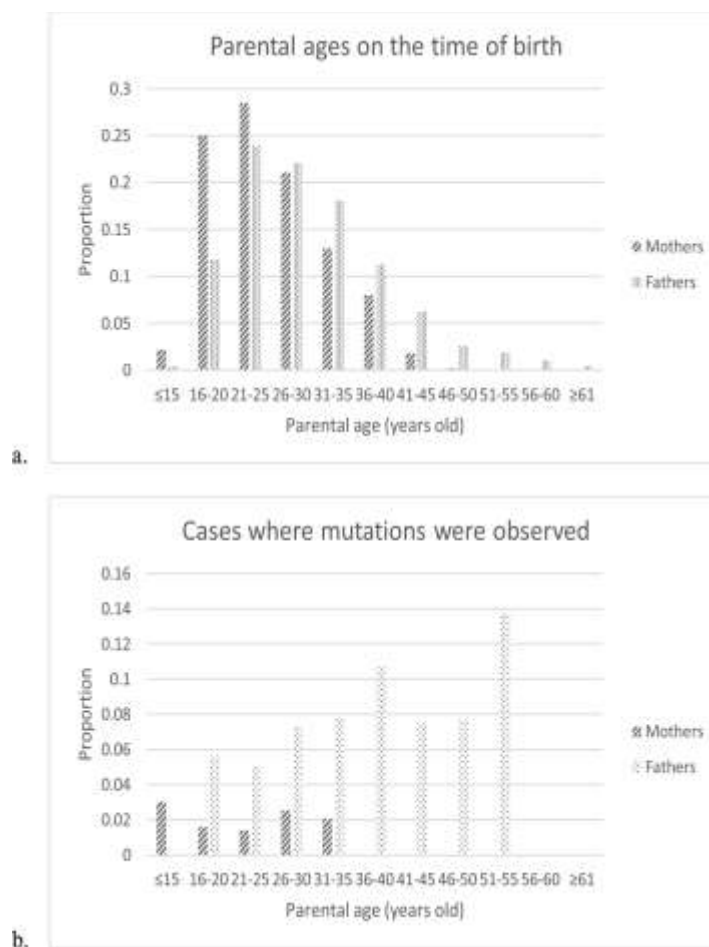


Fig. 2. a. Distribution of the parental ages at the time of the daughter's birth for the cases where such information was available. b. Proportion of cases where the meiosis resulted in at least one mutation for both males and females. Note that only mutations where the parental origin was possible to determine were considered. This implies that proportions presented are necessarily conservative.

Table 4

Estimates for locus specific mutation rates for male and female meiosis, considering the analysis of trios. The total number of 2*1612 meiosis considered in this study were gathered with other published data [33,42–53]. A level of confidence equal to 95 % were considered in the calculations.

# Markers	#Meiosis	Mutation Rate				Paternal Mutation Rate			Maternal Mutation Rate		
		# Mutations	MR	Lower	Upper	PMR	Lower	Upper	MMR	Lower	Upper
DXS10103	3224	6	1.9E-03	6.8E-04	4.0E-03	3.7E-03	1.4E-03	8.1E-03	0.0E+00	0.0E+00	2.3E-03
DXS8378	4082	5	1.2E-03	4.0E-04	2.9E-03	7.3E-04	5.3E-05	3.1E-03	7.3E-04	5.3E-05	3.1E-03
DXS10101	3452	11	3.2E-03	1.6E-03	5.7E-03	5.7E-03	2.7E-03	1.1E-02	7.1E-04	1.8E-05	3.6E-03
DXS10134	3524	17	4.8E-03	2.8E-03	7.7E-03	6.6E-03	3.4E-03	1.2E-02	3.0E-03	9.8E-04	7.0E-03
DXS10074	3776	19	5.0E-03	3.0E-03	7.8E-03	1.0E-02	6.1E-03	1.6E-02	0.0E+00	0.0E+00	2.0E-03
DXS7132	4726	31	6.6E-03	4.5E-03	9.3E-03	1.2E-02	7.7E-03	1.7E-02	1.5E-03	3.0E-04	4.1E-03
DXS10135	3452	37	1.1E-02	7.6E-03	1.5E-02	1.7E-02	1.2E-02	2.5E-02	4.2E-03	1.7E-03	8.5E-03
DXS7423	4024	3	7.5E-04	1.5E-04	2.2E-03	9.9E-04	1.2E-04	3.6E-03	5.0E-04	1.3E-05	2.8E-03
DXS10146	3224	9	2.8E-03	1.3E-03	5.3E-03	4.2E-03	1.6E-03	8.8E-03	1.4E-03	1.7E-04	4.8E-03
DXS10079	3534	22	6.2E-03	3.9E-03	9.4E-03	1.1E-02	6.7E-03	1.7E-02	1.4E-03	1.7E-04	4.7E-03
HPRTB	4132	7	1.7E-03	6.8E-04	3.5E-03	2.0E-03	5.5E-04	5.1E-03	1.4E-03	2.3E-04	4.2E-03
DXS10148	3224	14	4.3E-03	2.4E-03	7.3E-03	6.8E-03	3.4E-03	1.2E-02	1.9E-03	3.8E-04	5.4E-03

and Spanish samples, which supports the use of a single database for Iberia. For Brazil and Ecuador, the results evidenced the need of larger databases to investigate the genetic diversity and a possible stratification of the populations.

In all the studied populations, the statistical parameters of forensic relevance showed a high discrimination capacity of the full set of 12 X-STRs, both in males and females, as well as high values of a priori exclusion chance in paternity father/daughter duos and father/mother/daughter trios.

The inclusion in this study of very diverse population samples allowed the detection of a high number of new alleles, many of them not yet included in the allelic ladders provided with the kit. The high diversity of the markers led to a large number of haplotype combinations, highlighting the need of huge databases to obtain reliable estimates of haplotype frequencies. Indeed, the maximum number of observed haplotypes in the 4 population samples varied from 248 to 478, for LG3 and LG1, respectively. When considering the proportion of different haplotypes in the total sample (number of different haplotypes/total number of samples), these varied between 44%–67% for LG1, 30%–53%, for LG2, 22%–36%, for LG3, and 37%–58% for LG4, showing that other studies are still necessary to achieve more accurate estimates of haplotype frequencies.

In this work, we estimated paternal and maternal locus specific mutation rates considering mendelian incompatibilities both in duos and trios. Except for haploid genetic transmission, it is known that estimates of mutation rates through the proportion of observed incompatibilities are conservative, due to the possibility of hidden mutations, and the likelihood of this event may decrease when considering trios instead of duos. In this study, 18 mutations (more than 10 % of the total) would not be detected if only duos were analysed. When simplified algebraic formulas are used, the possibility of mutation is only considered when an incompatibility is observed. In this case, depending if duos or trios are under analysis, the rate of incompatibilities estimated through duos or trios should be used, respectively. Notwithstanding, specific software is nowadays available to compute likelihood ratios in kinship analysis using X-chromosomal markers, including mutation, among others [28,58]. In this tool, the likelihood of a specific parental allele mutating to a specific filial one is considered for all the possible allelic transitions (even in the absence of mendelian incompatibilities). Therefore, in this case, the most accurate estimate of the average mutation rate – the one estimated through trios, along with the most biologically reliable mutation model (in the light of the state of the art), should be used in any case.

The results obtained revealed higher estimates for mutation rates in male than in female meiosis, with no intersection of the confidence intervals in three cases for both duos and trios ($p = 0.05$; DXS10074, DXS10135 and DXS10079). Also, the proportion of mendelian

incompatibilities explained by a paternal mutation tend to increase with the age of the father. The DXS10135 marker showed the highest mutation rate, with order of magnitude equal to -2 , which is above the desirable for kinship investigations. Most of the found mutations (96 %) can be explained by the gain or loss of a single repeat, and only one mutation between alleles belonging to different microvariant classes were observed.

Declaration of Competing Interest

None.

Acknowledgments

We thank the support of Qiagen, as reduced prices in the purchase of the Investigator Argus X-12 QS were provided for the participants in this study.

We deeply thank the fruitful comments of an anonymous reviewer.

NP was supported by a Short-Term Fellowship grant from the International Society for Forensic Genetics (ISFG).

This work was partially financed by FEDER—Fundo Europeu de Desenvolvimento Regional funds through the COMPETE 2020—Operacional Programme for Competitiveness and Internationalisation (POCI), Portugal 2020; Portuguese funds through FCT—Fundação para a Ciência e a Tecnologia/Ministério da Ciência, Tecnologia e Inovação in the framework of the Decreto-Lei nº 57/2016 de 29 de Agosto (NP work contract) and the projects “Institute for Research and Innovation in Health Sciences” (POCI-01-0145-FEDER-007274) and “Center of Mathematics of the University of Porto” (UID/MAT/00144/2019), and by *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior* – CAPES (Finance Code 001). LG was supported by *Conselho Nacional de Desenvolvimento Científico e Tecnológico* – CNPq (ref. 305330/2016-0) and *Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro* – FAPERJ (CNE-2018). RC was supported by *Conselho Nacional de Desenvolvimento Científico e Tecnológico* – CNPq (ref. 302198/2016-4), *Fundação de Amparo à Pesquisa do Estado de São Paulo* – FAPESP (2016/10138-8) and *PADCFAR* (2017/04-1). GB was supported by *Dirección General de Investigación* through project MED.GBF.18.06. EP participation was supported by *Dirección Nacional de Investigaciones de la Fiscalía General del Estado, Ecuador*.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.fsigen.2020.102258>.

References

- [1] V. Pereira, L. Gusmão, Types of genomes, sequences and genetic markers, in: A. Amorim, B. Budowle (Eds.), *Handbook of Forensic Genetics. Biodiversity and Heredity in Civil and Criminal Investigation*, World Scientific, New Jersey, 2016.
- [2] T.M. Diegoli, Forensic typing of short tandem repeat markers on the X and Y chromosomes, *Forensic Sci. Int. Genet.* 18 (2015) 140–151.
- [3] N. Pinto, L. Gusmão, A. Amorim, X-chromosome markers in kinship testing: a generalisation of the IBD approach identifying situations where their contribution is crucial, *Forensic Sci. Int. Genet.* 5 (2011) 27–32.
- [4] D. Kling, Curiosities of X chromosomal markers and haplotypes, *Int. J. Legal Med.* 132 (2018) 361–371.
- [5] A.O. Tillmar, D. Kling, J.M. Butler, W. Parson, M. Prinz, P.M. Schneider, T. Egeland, L. Gusmão, DNA commission of the international society for forensic genetics (ISFG): guidelines on the use of X-STRs in kinship analysis, *Forensic Sci. Int. Genet.* 29 (2017) 269–275.
- [6] C. Tomas, J.J. Sanchez, A. Barbaro, C. Brandt-Casadevall, A. Hernandez, M.B. Dhiab, M. Ramon, N. Morling, X-chromosome SNP analyses in 11 human Mediterranean populations show a high overall genetic homogeneity except in North-west Africans (Moroccans), *BMC Evol. Biol.* 8 (2008) 75.
- [7] L. Gusmão, P. Sánchez-Diz, C. Alves, I. Gomes, M.T. Zarrabeitia, M. Abovich, 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.* 123 (2009) 227–234.
- [8] L. Li, C. Li, S. Zhang, S. Zhao, Y. Liu, Y. Lin, Analysis of 14 highly informative SNP markers on X chromosome by TaqMan SNP genotyping assay, *Forensic Sci. Int. Genet.* 4 (2010) e145–e148.
- [9] N.S. Freitas, R.L. Resque, E.M. Ribeiro-Rodrigues, J.F. Guerreiro, et al., X-linked insertion/deletion polymorphisms: forensic applications of a 33-markers panel, *Int. J. Legal Med.* 124 (2010) 589–593.
- [10] T.M. Diegoli, M.D. Coble, Development and characterization of two mini-X chromosomal short tandem repeat multiplexes, *Forensic Sci. Int. Genet.* 5 (2011) 415–421.
- [11] M. Nothnagel, R. Szibor, O. Vollrath, C. Augustin, J. Edelmann, M. Geppert, C. Alves, L. Gusmão, M. Vennenmann, Y. Hou, U.D. Immler, S. Inturri, H. Luo, S. Lutz-Bonengel, C. Robino, L. Roewer, B. Rolf, J. Sanft, K.J. Shin, J.E. Sim, P. Wiegand, C. Winkler, M. Krawczak, S. Hering, Collaborative genetic mapping of 12 forensic short tandem repeat (STR) loci on the human X chromosome, *Forensic Sci. Int. Genet.* 6 (2012) 778–784.
- [12] R. Pereira, V. Pereira, I. Gomes, C. Tomas, N. Morling, A. Amorim, M.J. Prata, A. Carracedo, L. Gusmão, A method for the analysis of 32 X chromosome insertion-deletion polymorphisms in a single PCR, *Int. J. Legal Med.* 126 (2012) 97–105.
- [13] T. Oki, T. Hayashi, M. Ota, H. Asamura, Development of multiplex assay with 16 SNPs on X chromosome for degraded samples, *Leg. Med. (Tokyo)* 14 (2012) 11–16.
- [14] E. Prieto-Fernández, M. Baeta, C. Núñez, M.T. Zarrabeitia, R.J. Herrera, J.J. Builes, et al., Development of a new highly efficient 17 X-STR multiplex for forensic purposes, *Electrophoresis* 37 (2016) 1651–1658.
- [15] C. Bini, L.N. Riccardi, S. Ceccardi, F. Carano, S. Sarno, D. Luiselli, et al., Expanding X-chromosomal forensic haplotype frequencies database: Italian population data of four linkage groups, *Forensic Sci. Int. Genet.* 15 (2015) 127–130.
- [16] J. Xing, A. Adnan, A. Rakha, K. Kasim, et al., Genetic analysis of 12 X-STRs for forensic purposes in Liaoning Manchu population from China, *Gene* 683 (2019) 153–158.
- [17] R. Tao, J. Zhang, Y. Bian, R. Dong, et al., Investigation of 12 X-STR loci in Mongolian and Eastern Han populations of China with comparison to other populations, *Sci. Rep.* 8 (2018) 4287.
- [18] F. Guo, Genetic polymorphism of 17 autosomal STR loci in the Lahu ethnic minority from Yunnan Province, Southwest China, *Forensic Sci. Int. Genet.* 31 (2017) e52–e53.
- [19] I. Cortés-Trujillo, F. Zuñiga-Chiquette, B. Ramos-González, M.L. Chávez-Briones, et al., Allele and haplotype frequencies of 12 X-STRs in Mexican population, *Forensic Sci. Int. Genet.* 38 (2019) e11–e13.
- [20] J.M. Salvador, D.L.T. Apaga, F.C. Delfin, G.C. Calacal, et al., Filipino DNA variation at 12 X-chromosome short tandem repeat markers, *Forensic Sci. Int. Genet.* 36 (2018) e8–e12.
- [21] G. Mršić, P. Ozretić, J. Crnjac, S. Merkaša, et al., Expanded Croatian 12 X-STR loci database with an overview of anomalous profiles, *Forensic Sci. Int. Genet.* 34 (2018) 249–256.
- [22] M.A. Almarri, R.A. Lootah, Allelic and haplotype diversity of 12 X-STRs in the United Arab Emirates, *Forensic Sci. Int. Genet.* 33 (2018) e4–e6.
- [23] I. Veselinović, D. Vapa, M. Djan, N. Veličković, et al., Genetic analysis of 12 X-STR loci in the Serbian population from Vojvodina Province, *Int. J. Legal Med.* 132 (2018) 405.
- [24] L. Cainé, S. Costa, M.F. Pinheiro, Population data of 12 X-STR loci in a North of Portugal sample, *Int. J. Legal Med.* 127 (2013) 63–64.
- [25] C. Tomas, V. Pereira, N. Morling, Analysis of 12 X-STRs in greenlanders, Danes and somalis using argus X-12, *Int. J. Legal Med.* 126 (2012) 121–128.
- [26] K. Rębala, S.A. Kotova, V.I. Rybakova, T.V. Zabauskaya, A.A. Shyla, A.A. Spivak, I.S. Tsybovsky, S. Zscherkwolska, Variation of X-chromosomal microsatellites in Belarus within the context of their genetic diversity in Europe, *Forensic Sci. Int. Genet.* 16 (2015) 105–111.
- [27] M.G. Garcia, C.I. Catanesi, G.A. Penacino, L. Gusmão, N. Pinto, X-chromosome data for 12 STRs: towards an Argentinian database of forensic haplotype frequencies, *Forensic Sci. Int. Genet.* 41 (2019) e8–e13.
- [28] D. Kling, B. Dell'Amico, A.O. Tillmar, FamLinkX – implementation of a general model for likelihood computations for X-chromosomal marker data, *Forensic Sci. Int. Genet.* 17 (2015) 1–7.
- [29] T.M. Diegoli, H. Rohde, S. Borowski, M. Krawczak, M.D. Coble, M. Nothnagel, Genetic mapping of 15 human X chromosomal forensic short tandem repeat (STR) loci by means of multi-core parallelization, *Forensic Sci. Int. Genet.* 25 (2016) 39–44.
- [30] P. Vicard, A.P. Dawid, A statistical treatment of biases affecting the estimation of mutation rates, *Mutat. Res.* 547 (2004) 19–33.
- [31] P. Vicard, A.P. Dawid, J. Mortera, S.L. Lauritzen, Estimating mutation rates from paternity casework, *Forensic Sci. Int. Genet.* 2 (2008) 9–18.
- [32] K. Slooten, F. Ricciardi, Estimation of mutation probabilities for autosomal STR markers, *Forensic Sci. Int. Genet.* 7 (3) (2013) 337–344.
- [33] A. Pico, A. Castillo, C. Vargas, A. Amorim, L. Gusmão, Genetic profile characterization and segregation analysis of 10 X-STRs in a sample from Santander, Colombia, *Int. J. Legal Med.* 122 (2008) 347–351.
- [34] K.J. Shin, B.K. Kwon, S.S. Lee, J.E. Yoo, M.J. Park, U. Chung, et al., Five highly informative X-chromosomal STRs in Koreans, *Int. J. Legal Med.* 118 (2004) 37–40.
- [35] J. Edelmann, R. Szibor, DXS101: a highly polymorphic X-linked STR, *Int. J. Legal Med.* 114 (2001) 301–304.
- [36] M. Poetsch, H. Petersmann, A. Repenning, E. Lignitz, Development of two pentaplex systems with X-chromosomal STR loci and their allele frequencies in a northeast German population, *Forensic Sci. Int.* 155 (2005) 71–76.
- [37] M. García, L. Gusmão, C. Catanesi, G. Penacino, N. Pinto, Mutation rate of 12 X-STRs from Investigator Argus X-12 kit in Argentine population, *Forensic Sci. Int. Genet. Suppl. Ser.* 6 (2017) e562–e564.
- [38] M. Scherer, M. König, M. Bussmann, A. Prochnow, R. Peist, Development and validation of the new investigator® argus X-12 QS kit, *Forensic Sci. Int. Genet. Suppl. Ser.* 5 (2015) e256–e257.
- [39] L. Excoffier, H.E.L. Lischer, Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows, *Mol. Ecol. Resour.* 10 (2010) 564–567.
- [40] D. Desmarais, Y. Zhong, R. Chakraborty, C. Perreault, L. Busque, Development of a highly polymorphic STR marker for identity testing purposes at the human androgen receptor gene (HUMARA), *J. Forensic Sci.* 43 (1998) 1046–1104.
- [41] I. Gomes, P.J.P. Pereira, S. Harms, A.M. Oliveira, P.M. Schneider, A. Brehm, Genetic characterization of Guinea-Bissau using a 12 X-chromosomal STR system: inferences from a multiethnic population, *Forensic Sci. Int. Genet.* 31 (2017) 89–94.
- [42] M. Chen, H. Ren, Z. Liu, J. Zhao, C. Chen, Y. Shi, et al., Genetic polymorphisms and mutation rates of 16 X-STRs in a Han Chinese population of Beijing and application examples in second-degree kinship cases, *Int. J. Legal Med.* (2019), <https://doi.org/10.1007/s00414-019-02047-8> [Epub ahead of print].
- [43] T.M. Diegoli, A. Linacre, M.S. Schanfield, M.D. Coble, Mutation rates of 15 X chromosomal short tandem repeat markers, *Int. J. Legal Med.* 128 (4) (2014) 579–587.
- [44] T. Nishi, A. Kurosu, Y. Sugano, J. Kaminiwa, Y. Sekine, S. Yano, et al., Application of a novel multiplex polymerase chain reaction system for 12 X-chromosomal short tandem repeats to a Japanese population study, *Leg. Med. (Tokyo)* 15 (1) (2013) 43–46.
- [45] Q.L. Liu, D.J. Lu, L. Quan, Y.F. Chen, M. Shen, H. Zhao, et al., Development of multiplex PCR system with 15 X-STR loci and genetic analysis in three nationality populations from China, *Electrophoresis* 33 (8) (2012) 1299–1305.
- [46] S. Tetzlaff, R. Wegener, I. Lindner, Population genetic investigation of eight X-chromosomal short tandem repeat loci from a northeast German sample, *Forensic Sci. Int. Genet.* 6 (6) (2012) e155–e156.
- [47] D. Becker, H. Rodig, C. Augustin, J. Edelmann, F. Götz, S. Hering, et al., Population genetic evaluation of eight X-chromosomal short tandem repeat loci using Mentype Argus X-8 PCR amplification kit, *Forensic Sci. Int. Genet.* 2 (1) (2008) 69–74.
- [48] Q.L. Liu, D.J. Lv, X.L. Wu, H.Y. Sun, X.Y. Wu, H.L. Lu, Development of a five ChX STRs loci typing system, *Int. J. Legal Med.* 122 (3) (2008) 261–265.
- [49] M.A. Tariq, O. Ullah, S.A. Riazuddin, S. Riazuddin, Allele frequency distribution of 13 X-chromosomal STR loci in Pakistani population, *Int. J. Legal Med.* 122 (6) (2008) 525–528.
- [50] S. Turrina, R. Atzei, G. Filippini, D. De Leo, Development and forensic validation of a new multiplex PCR assay with 12 X-chromosomal short tandem repeats, *Forensic Sci. Int. Genet.* 1 (2) (2007) 201–204.
- [51] K. Tabbada, M.C. De Ungria, L.P. Faustino, D. Athanasiadou, B. Stradmann-Bellinghausen, P.M. Schneider, Development of a pentaplex X-chromosomal short tandem repeat typing system and population genetic studies, *Forensic Sci. Int.* 154 (2–3) (2005) 173–180.
- [52] S. Turrina, D. De Leo, Population genetic comparisons of three X-chromosomal STRs (DXS7132, DXS7133 and GATA172D05) in North and South Italy, *Int. Congr. Ser.* 1261 (2004) 302–304.
- [53] D. Athanasiadou, B. Stradmann-Bellinghausen, C. Rittner, K.W. Alt, P.M. Schneider, Development of a quadruplex PCR system for the genetic analysis of X-chromosomal STR loci, *Int. Congr. Ser.* 1239 (2003) 311–314.
- [54] S. Antão-Sousa, E. Conde-Sousa, L. Gusmão, A. Amorim, N. Pinto, Underestimation and misclassification of mutations at X chromosome STRs depend on population's allelic profile, *Forensic Sci. Int. Genet. Suppl. Ser.* 7 (1) (2019) 718–720, <https://doi.org/10.1016/j.fsigs.2019.10.150>.
- [55] M. Ohno, Spontaneous de novo germline mutations in humans and mice: rates, spectra, causes and consequences, *Genes Genet. Syst.* 94 (1) (2019) 13–22.
- [56] L. Séguirel, M.J. Wyman, M. Przeworski, Determinants of mutation rate variation in the human germline, *Annu. Rev. Genomics Hum. Genet.* 15 (2014) 47–70.
- [57] A. Kong, M.L. Frigge, G. Masson, S. Besenbacher, P. Sulem, G. Magnusson, et al., Rate of de novo mutations and the importance of father's age to disease risk, *Nature* 488 (7412) (2012) 471–475.
- [58] D. Kling, A. Tillmar, T. Egeland, P. Mostad, A general model for likelihood computations of genetic marker data accounting for linkage, linkage disequilibrium, and mutations, *Int. J. Legal Med.* 129 (5) (2015) 943–954.



Analysis of the complete mtDNA in Brazilian samples: Haplotypes differentiation and identification of new variations in Native American haplogroups



Braganholi DF^{a,*}, Freitas JM^b, Andrekenas NC^a, Januário BB^a, Ambrosio IB^a, Polverari FS^a, Cicarelli RMB^a

^aLaboratório de Investigação de Paternidade, UNESP-FCEAr, Araraquara, SP, Brazil

^bPolícia Federal, Instituto Nacional de Criminalística, Brasília, DF, Brazil

ARTICLE INFO

Keywords:

mtDNA
Brazil
Center-west region
Human identification

ABSTRACT

Initial data about the population of the Center-West Brazil region for mtDNA CR were published in 2019. From this group, 74 samples were selected from individuals who shared haplotypes. The purpose was to analyze the complete mtDNA by Massively Parallel Sequencing and to evaluate if it is possible to differentiate these haplotypes. Libraries were created for all samples and to date 22 samples were sequenced by the Illumina system. Analyzing only the CR, 8 haplotypes were identified in these 22 samples. The complete mtDNA analysis allowed the identification of 16 haplotypes, 6 of which were shared by two samples each, that is, the diversity of haplotypes in this sample increased. In addition, we identified some genetic variants not reported in the EMPOP in mtDNAs classified in haplogroups A2, B2, L2a1f and L0a2a2a.

1. Introduction

With the advent of new sequencing technologies, the analysis of small genomes like the mtDNA becomes more practical, allowing the use of this complete genome to identify individuals.

The possible increase in the genetic diversity obtained when analyzing the complete mtDNA has been observed in different world populations [1] and can make the classification of haplogroups according to ethnic origin more efficient [2].

In a population study, initial data for mtDNA CR about the Center-West Brazil region were published in 2019 [3], being that 74 samples shared haplotypes, not allowing the differentiation for this sample group.

The general purpose is to analyze the complete mtDNA of these samples by Massively Parallel Sequencing (MPS) and to evaluate if is possible to differentiate these haplotypes. Initially we present here the results obtained for 22 samples that had the entire mtDNA control region analyzed by Sanger sequencing and presented 8 shared haplotypes.

2. Methodology

The complete mitochondrial genome was analyzed by MPS: amplification of the mtDNA by long PCR in two separate reactions using the

TaKaRa LA PCR Kit (TaKaRa) and library preparation using the Nextera XT DNA Sample Kit (Illumina). The libraries were analyzed on Miseq system (Illumina) and the obtained data were analyzed using the tools mtDNA Variant Processor and mtDNA Variant Analyzer in BaseSpace cloud (Illumina).

3. Results and discussion

Analyzing only the CR, 8 haplotypes were identified in these 22 samples. The complete mtDNA analysis allowed the identification of 16 haplotypes, 6 of which were shared by two samples each, that is, the diversity of haplotypes in this sample increased (Fig. 1).

In addition, we identified some genetic variants not reported in the EMPOP: 16192.1 T, 14881A; 15382A, 15716A. Each one of these variations was identified separately in 4 different mtDNAs, 2 of which were classified in A2 and B2 haplogroups, which have Native American ancestral origin; the other 2 mtDNAs were classified into haplogroups with African ancestral origin (L2a1f and L0a2a2a) (Table 1).

4. Conclusion

The complete mtDNA analysis improved the discrimination between individuals in this studied sample but did not detect a unique mtDNA

* Corresponding author.

E-mail address: danielobraganholi@hotmail.com (B. DF).

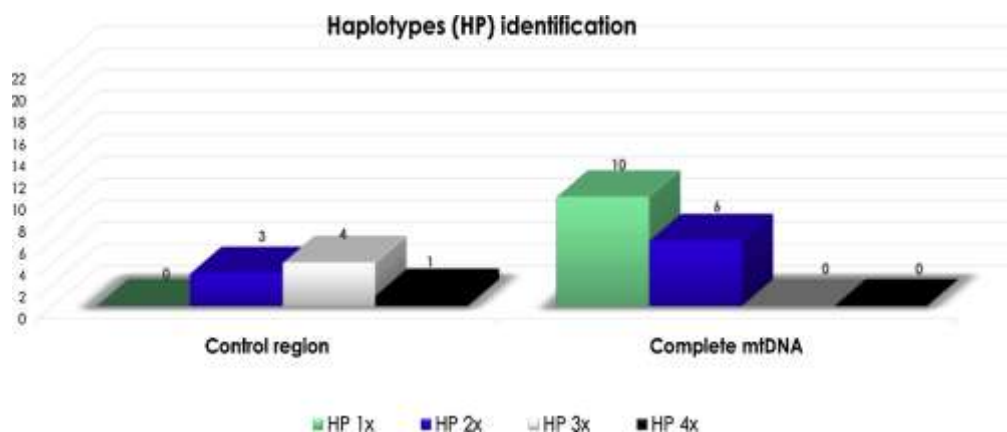


Fig. 1. Comparison of haplotype sharing when analyzing only the CR and the complete mtDNA. HP = Haplotype. 1X = unique haplotype. 2X, 3X and 4X = number of times the haplotype has been shared.

Table 1

Variants found in samples not yet reported in EMPOP and their classification in haplogroups. *The haplogroup L2a1f has been reported to occur in African Americans.

N	Unobserved Variation	Haplogroup	Ancestral Origin
1	C15382A	A2	NAM
1	C15716A	B2	NAM
1	C14881A	L20a2a2a	AFR
1	16192.1 T	L2a1f	AFR*

for all the individuals.

It is possible that the identified variations are related to Native American haplogroups or haplogroups that occur in the Americas, which should be confirmed by the analysis of more samples.

Declaration of Competing Interest

No one.

References

- [1] K. Ma, H. Li, Y. Cao, X. Zhao, W. Liu, X. Zhao, Haplotype diversity in mitochondrial genome in a Chinese Han population, *J. Hum. Genet.* (2016), <https://doi.org/10.1038/jhg.2016.74>.
- [2] S. Park, S. Cho, H.J. Seo, J.H. Lee, M.Y. Kim, S.D. Lee, Entire mitochondrial DNA sequencing on massively parallel sequencing for the Korean population, *J. Korean Med. Sci.* (2017), <https://doi.org/10.3346/jkms.2017.32.4.587>.
- [3] J.M. Freitas, L.H. Fassio, D.F. Braganholi, G. Chemale, Mitochondrial DNA control region haplotypes and haplogroup diversity in a sample from Brasilia, Federal District, Brazil. *Forensic Sci. Int. Genetics.* (2019), <https://doi.org/10.1016/j.fsigen.2019.02.006>.