

**UNIVERSIDADE ESTADUAL PAULISTA JULIO DE MESQUITA FILHO**

**Faculdade de Ciências Farmacêuticas**

**JOYCE APARECIDA MARTINS**

**ESTUDO DE FREQUÊNCIAS ALÉLICAS DE STRs DO  
CROMOSSOMO X NA POPULAÇÃO BRASILEIRA DE  
ARARAQUARA-SP**

**ARARAQUARA-SP**

**2008**

**UNIVERSIDADE ESTADUAL PAULISTA JULIO DE MESQUITA FILHO**

**Faculdade de Ciências Farmacêuticas**

**JOYCE APARECIDA MARTINS**

**ESTUDO DE FREQUÊNCIAS ALÉLICAS DE STRs DO  
CROMOSSOMO X NA POPULAÇÃO BRASILEIRA DE  
ARARAQUARA-SP**

Dissertação apresentada ao Programa de Pós-Graduação em Análises Clínicas (Área de Concentração em Biologia Molecular), da Faculdade de Ciências Farmacêuticas da Universidade Estadual Paulista, como requisito parcial para a obtenção do título de Mestre.

Orientador: Profa. Dra. Regina Maria Barretto Cicarelli.

**ARARAQUARA-SP**

**2008**

**DEDICATÓRIA**

Aos

Meus pais Gilberto Martins e Zulmira Furlan Fraga Martins por compartilharem os meus ideais e os alimentarem, incentivando-me a prosseguir nesta jornada, pelos grandiosos ensinamentos que fizeram de mim a pessoa de hoje e pelo laço de amor que nos une intensamente, fortalecendo-me a cada dia;

Meus Irmãos Jaqueline Martins, Gilberto Martins Filho e Juliana Lara Martins pela verdadeira amizade que cultivamos, pelas conversas enriquecedoras, pelo amparo nos momentos de fragilidade e por estarmos sempre juntos nessa longa caminhada;

Ao meu noivo José Roberto Lopes Ferraz Filho pelo imenso carinho e companheirismo, por respeitar profundamente a minha maneira de ser, apoiando-me em meus ideais mesmo que para isso fiquemos separados e pelos ótimos momentos que temos passados juntos;

A todos vocês, que amo muito, a minha gratidão e o meu reconhecimento, pois nos méritos desta conquista há muito de vossas presenças!

## **AGRADECIMENTOS**

À Professora Regina Maria Barretto Cicarelli por ter me mostrado, com muita garra e ética, o caminho da ciência, pela orientação e confiança, pela contribuição imensurável a minha formação acadêmica e pessoal e, principalmente, pela amizade e o respeito que temos uma pela outra.

À Professora Vera Lucy de Santi Alvarenga por sempre me apoiar na tomada de decisões, pelas palavras amigas e encorajadoras nos momentos difíceis, pela sua paz que me contagia!

Ao Professor Rogério Nogueira de Oliveira pelos trabalhos que temos realizado em conjunto, pela confiança, amizade e pela sugestão desse tema para o desenvolvimento do Mestrado.

A todos do Laboratório Biocod, especialmente à Cristiane Lommez de Oliveira, Edina, Gisele, Laélia, Gabriela, Elaine (s), Francisco, Débora, Carla, Roberto e Rosane por terem aberto as portas do laboratório para o desenvolvimento desse trabalho, pela disponibilidade e contribuição na realização das metodologias e, principalmente, pela amizade.

À Maria Jenny Mitraud Lima pela grande amizade, pelos maravilhosos ensinamentos e discussões que me proporciona na área de estatística, pela paciência e assistência em Belo Horizonte para que eu pudesse desenvolver parte do projeto.

À Greiciane pela serenidade e companheirismo em vários momentos da minha vida, pela troca de idéias tanto na área profissional como na pessoal, pelo excelente exemplo acadêmico que é!

Às “Meninas da Paternidade”, Gabriella, Adriana, Larissa, Maria Angélica e Jeane, pela verdadeira amizade e pelo auxílio imensurável na rotina laboratorial.

Ao Marco Túlio e à Daniela pela verdadeira amizade, pela disposição em ajudar e pelos ensinamentos na área de biologia molecular.

À Érica e Gabriela pela amizade e por me ajudarem até mesmo à distância (por e-mail, telefone).

Aos funcionários do Departamento de Ciências Biológicas (em especial, à Bernadete e Valéria), do Laboratório de Bioquímica (em especial, ao Tarcísio e Eduardo) e ao pessoal da Portaria da Farmácia (em especial, à Olívia, Sônia e Tiana) pela constante disponibilidade e apoio.

A todos que, voluntariamente, contribuíram com a principal “matéria-prima” desse projeto: a amostra biológica.

Às companheiras de graduação e de república, Ana Carolina, Beatriz, Camila, Fabiana, Julhiany, Kamila, Maira, Maísa e Mariana pela verdadeira amizade, pelos momentos de alegria, medo e incertezas que passamos juntas e por terem sempre me apoiado e incentivado na busca deste sonho.

Às companheiras atuais de república, Angélica, Sônia, Ana e Tatiane pela verdadeira amizade e pelos momentos de distração e de muitas risadas que me proporcionam.

À Faculdade de Ciências Farmacêuticas da Universidade Estadual Paulista pelos ensinamentos acadêmicos, pela alma científica e pelo núcleo profissional de excelência!

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

Ao PADC–UNESP (Programa de Apoio ao Desenvolvimento Científico) pelo auxílio financeiro para a realização deste trabalho.

Aos demais, amigos e familiares, que mesmo distante torceram por mim e fazem parte desta conquista!

*"É muito importante que o homem tenha ideais. Sem eles, não se vai a parte alguma. No entanto, é irrelevante alcançá-los ou não. É apenas necessário mantê-los vivos e procurar atingi-los."*

*Dalai - Lama*

## SUMÁRIO

### CAPÍTULO I

Resumo

Abstract

Lista de Figuras

Lista de Tabelas

Lista de Abreviaturas, Siglas e Símbolos

<b>1. Introdução.....</b>	<b>21</b>
<b>2. Revisão Bibliográfica.....</b>	<b>23</b>
2.1 Identificação Humana.....	24
2.2 Marcadores biológicos para identificação humana: panorama histórico.....	25
2.3 Polimorfismos genéticos e a análise do DNA.....	27
2.4 Bancos de dados de DNA.....	29
2.5 Cromossomo X.....	32
2.5.1 Evolução do cromossomo X.....	32
2.5.2 Características do cromossomo X.....	33
2.5.3 Marcadores do cromossomo X e desequilíbrio de ligação.....	37
2.5.4 Aplicações dos marcadores genéticos do cromossomo X.....	40
2.5.5 Estudos populacionais com marcadores genéticos do cromossomo X.....	43
<b>3. Objetivos.....</b>	<b>47</b>
<b>4. Material e Métodos.....</b>	<b>49</b>
4.1 Aspectos éticos.....	50
4.2 Casuística.....	50
4.3 Extração do DNA.....	51
4.4 Reação de PCR.....	51

4.4.1 Método Monoplex.....	54
4.4.2 Método Multiplex.....	55
4.5 Obtenção dos padrões internos da corrida em seqüenciador automático.....	56
4.6 Detecção e análise dos produtos da PCR.....	58
4.7 Análise estatística dos dados.....	59
<b>5. Resultados e Discussão.....</b>	<b>61</b>
5.1 Protocolo de extração.....	62
5.2 Avaliação dos métodos de PCR: Monoplex e Multiplex.....	63
5.3 Dado genético populacional.....	68
<b>6. Conclusões.....</b>	<b>84</b>
<b>7. Referências Bibliográficas.....</b>	<b>86</b>
<b>8. Apêndices.....</b>	<b>97</b>
8.1 Soluções e géis utilizados.....	98
8.2 Distribuição haplotípica do cromossomo X em indivíduos masculinos da população de Araraquara-SP.....	100
 <b>CAPÍTULO II</b>	
Genetic analysis of 5 X-STR loci in the population of São Paulo State (Southeast of Brazil).....	103
Genetic analysis of X and Y Short Tandem Repeat Loci in Araraquara population, Brazil.....	115

## **CAPÍTULO I**

A identificação humana através da análise do ácido desoxirribonucléico (DNA) é realizada pelo estudo de regiões polimórficas do DNA, transferidas dos pais para os filhos pelo mecanismo de herança genética. Os marcadores mais utilizados na rotina forense são regiões microssatélites ou STRs (Short Tandem Repeats) e se classificam em STRs autossômicos (AS STRs), do cromossomo Y (Y-STRs) e do cromossomo X (X-STRs). Estes últimos são de utilização recente nos testes de identificação humana, sendo aplicados com a finalidade de complementar os dados obtidos com os demais marcadores.

Tendo em vista a necessidade de ampliação dos dados da população brasileira em relação aos marcadores genéticos, este projeto teve o objetivo de determinar as frequências alélicas e os parâmetros estatísticos de interesse na prática forense e em testes de paternidade para 5 X-STRs. Para isto foram analisados 120 indivíduos não aparentados, classificados segundo o grupo de cor de pele (30 brancos, 30 pretos, 30 pardos e 30 amarelos) e residentes em Araraquara.

Através dos resultados obtidos neste estudo, verificou-se que a população brasileira de Araraquara apresenta particularidades na sua distribuição alélica, havendo diferenças entre esta e demais populações de outros países, sendo que uma maior distância genética foi obtida com a população asiática. Em relação aos marcadores X-STRs analisados, o DXS101 foi o mais polimórfico, seguido por DXS7424, DXS6854, DXS7132 e DXS6808. O poder de discriminação obtido foi de 0,9999928 e 0,9990702 em mulheres e homens, respectivamente, constituindo tal sistema em uma poderosa ferramenta para a prática forense e testes de paternidade.

Palavras-chave: Identificação Humana, STR, Cromossomo X, Forense, Araraquara.

Human identification by examining the desoxyribonucleic acid (DNA) is held by the study of polymorphic regions of the DNA, transferred from the parents to their children by mechanism of genetic inheritance. The more used markers in forensic routine are microsatellites regions or STRs (Short Tandem Repeats) and they can be classified into STRs autosomal (AS STRs), Y chromosome (Y-STRs) and X chromosome (X-STRs). These last two are for use in the recent identification of human testing, being implemented with the aim to complement the data obtained with other markers.

Because the needs of expansion of the Brazilian population data in relation to genetic markers, this project aimed to determine the allelic frequencies and the statistical parameters of interest in forensic and paternity tests to 5 X-STRs. To do this were analyzed 120 individuals not related, classified according to its skin color group (30 European Brazilians, 30 Pardo Brazilians, 30 African Brazilians and 30 Oriental Brazilians) and residents in Araraquara.

Through the results obtained in this study, it was found that the Brazilian population of Araraquara presents features in its allelic distribution, and there are differences between it and other populations from other countries, and that greater genetic distance was obtained with the Asian population. For X-STRs markers analyzed, DXS101 was the most polymorphic, followed by DXS7424, DXS6854, DXS7132 and DXS6808. The power of discrimination obtained was 0.9999928 and 0.9990702 in women and men, respectively, constituting such a system into a powerful tool for forensic and paternity tests.

**Keywords:** Human Identification, STR, X Chromosome, Forensic, Araraquara.

- Figura 1: Representação esquemática do cromossomo X e das regiões de homologia (coloridas) entre os cromossomos sexuais (X -Y).....36
- Figura 2: Perfis alélicos obtidos em um caso de investigação de paternidade do tipo “reconstrução”. STRs de 1 a 3 correspondem aos marcadores analisados e, SP, ao suposto pai falecido.....38
- Figura 3: Ideograma do cromossomo X com a localização dos STRs analisados neste trabalho.....52
- Figura 4: Gel de agarose 0,8% em TAE [1x], corado com BrEt. Produto da reação de amplificação dos STRs DXS6854, DXS7424, DXS101, DXS6808 e DXS7132 pelo método Monoplex, correspondentes as bandas obtidas nas canaletas de 1 a 5, respectivamente. LML: *Low DNA Mass Ladder*.....63
- Figura 5: Gel de agarose 4% em TAE [1x], corado com BrEt. Produto da reação de amplificação dos STRs DXS6854, DXS7424, DXS6808 e DXS7132 pelo método Multiplex e do STR DXS101 pelo método Monoplex. Canaleta 1 e 3: Produto da reação de PCR multiplex em uma amostra masculina e feminina (homozigota para 3 marcadores), respectivamente. Canaleta 2 e 4: Produto da reação de PCR monoplex em uma amostra masculina e feminina (heterozigota para o marcador), respectivamente. LML: *Low DNA Mass Ladder*.....65

Figura 6: Eletroferograma da escada alélica (Perfil 16) e do controle K562 (Perfil 17) para os 5 X-STRs, incluindo os padrões internos de corrida (SRY255 e D9S938).....	66
Figura 7: Eletroferograma das amostras nas quais os 5 X-STRs foram amplificados pelo método Monoplex. Perfil 16: escada alélica, Perfis 18 a 21: amostras de indivíduos masculinos, Perfil 20: controle K562.....	67
Figura 8: Eletroferograma das amostras nas quais os 4 STRs DXS6854, DXS7424, DXS6808 e DXS7132 foram amplificados pelo método Multiplex e, o DXS101, pelo Monoplex. Perfil 12: escada alélica, Perfis 15, 16, 19 e 22: amostras de indivíduos masculinos, Perfil 11: controle K562.....	67
Figura 9: Contribuição do DXS6808 para a identificação humana e o exame de paternidade na população de Araraquara-SP. MECD: Chance de Exclusão Significativa para marcadores do cromossomo X na ausência da mãe (dupla pai-filha); MECT: Chance de Exclusão Significativa para marcadores do cromossomo X em trios envolvendo filhas; PDM: Poder de Discriminação Masculino; PDM: Poder de Discriminação Feminino.....	82

Tabela 1. Características dos cinco X-STRs.....	53
Tabela 2. Seqüência dos primers utilizados para os cinco X-STRs.....	53
Tabela 3. Características dos Padrões Internos da corrida.....	57
Tabela 4. Seqüência dos primers utilizados para os Padrões Internos da corrida.....	57
Tabela 5. Freqüência alélica dos cinco X-STRs na população de Araraquara-SP.....	71
Tabela 6. Análise comparativa entre a população de Araraquara (n = 120) e outras populações para o marcador DXS7424.....	73
Tabela 7. Análise comparativa entre a população de Araraquara (n = 120) e outras populações para o marcador DXS101.....	74
Tabela 8. Análise comparativa entre a população de Araraquara (n = 120) e outras populações para o marcador DXS7132.....	76
Tabela 9. Parâmetros estatísticos dos cinco X-STRs na população de Araraquara-SP.....	78

## LISTA DE ABREVIATURAS, SIGLAS E SÍMBOLOS

AS STRs	<i>Autosomal Short Tandem Repeats</i>
BrEt	Brometo de Etídio
BSA	Albumina Sérica Bovina
C	Base nitrogenada Citosina
cM	Centimorgan
CODIS	<i>COmbined DNA Index System</i>
Cy5	Cianina 5
DH	Diversidade Haplotípica
DNA	Ácido Desoxirribonucléico
dNTP	Dinucleotídeotrifosfato
DTT	Dithiothreitol
EDTA	Ácido Etilenodiaminotetracético
EDUCAFRO	Educação e Cidadania de Afrodescendentes e Carentes
EUA	Estados Unidos da América
F	Seqüência do <i>Primer Forward</i>
FAPEAL	Fundação de Amparo à Pesquisa do Estado de Alagoas
FAPERJ	Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro
FBI	<i>Federal Bureau of Investigation</i>
FCF-UNESP	Faculdade de Ciências Farmacêuticas da Universidade Estadual Paulista
FO-USP	Faculdade de Odontologia da Universidade de São Paulo
G	Base Nitrogenada Guanina
GDB	<i>Genome Database</i>
HCL	Ácido Clorídrico

HETesp	Heterozigose Esperada
HETobs	Heterozigose Observada
HmtDB	<i>Human Mitochondrial Genomica Database</i>
HLA	<i>Histocompatibility Leucocyte Antigen</i>
HWE	<i>Hardy-Weinberg Equilibrium</i>
IDT	<i>Integrated DNA Technologies</i>
IBGE	Instituto Brasileiro de Geografia e Estatística
KCl	Cloreto de Potássio
LD	Desequilíbrio de Ligação
LDML	<i>Low DNA Mass Ladder</i>
LINES	<i>Long Interspersed Nuclear Elements</i>
LTRs	<i>Long Tandem Repeats</i>
M	Molar
mA	<i>Miliampers</i>
MEC	Chance de Exclusão Significativa
MEC <sub>D</sub>	Chance de Exclusão Significativa para Marcadores do cromossomo X na Ausência da Mãe (dupla pai-filha)
MEC <sub>T</sub>	Chance de Exclusão Significativa para Marcadores do cromossomo X em trios envolvendo filhas
MgCl <sub>2</sub>	Cloreto de Magnésio
Min	Minuto
mL	Mililitro
mM	Milimolar
Mpb	Mega Pares de Base
m/v	Massa/Volume

NCBI	<i>National Center for Biotechnology Information</i>
ng	Nanograma
NDIS	<i>Nacional DNA Index System</i>
OMIM	<i>Online Mendelian Inheritance in Man</i>
ONG	Organização Não Governamental
PAR1	<i>Pseudoautosomal Region 1</i>
PAR2	<i>Pseudoautosomal Region 2</i>
pb	Pares de Base
PCR	<i>Polymerase Chain Reaction</i>
PI	Padrão Interno
R	Seqüência do <i>Primer Reverse</i>
RFLP	<i>Restriction Fragment Length Polymorphism</i>
SENASP	Secretaria Nacional de Segurança Pública
SINEs	<i>Short Interspersed Nuclear Elements</i>
SNPs	<i>Single Nucleotide Polymorphism</i>
STRs	<i>Short Tandem Repeats</i>
TAE	Tris Acetato EDTA
TBE	Tris Borato EDTA
TEMED	N, N, N', N' - Tetrametiletilenodiamina
U	Unidade
UFRJ	Universidade Federal do Rio de Janeiro
V	<i>Volts</i>
VNTRs	<i>Variable Number of Tandem Repeats</i>
v/v	Volume/Volume
W	<i>Watts</i>

XAR	<i>X-Added Region</i>
Xcen	Região Centromérica do Cromossomo X
XCR	<i>X-Conserved Region</i>
X-STRs	<i>X- Short Tandem Repeats</i>
XTR	<i>X- Transposed Region</i>
YHRD	<i>Y-STR Haplotype Reference Database</i>
Y-STRs	<i>Y- Short Tandem Repeats</i>
°C	Graus Celcius
μL	Microlitro
∞	Infinito

“A Pesquisa pela  
identificação e validação  
de mais e melhores  
sistemas marcadores para  
análise forense deve  
continuar com a visão de  
tornar cada perfil único”

NRC II Report

## **1. INTRODUÇÃO**

A identificação humana por meio da análise do DNA constitui, atualmente, uma ferramenta indispensável na Genética Moderna, auxiliando de forma imensurável tanto nas questões de identificação civil como nas de identificação criminal (FRANÇA, 2001). Isto se deve ao alto poder de discriminação do DNA, o qual permite obter a identificação de um indivíduo e/ou determinar o vínculo biológico entre indivíduos com a menor probabilidade de erro.

O perfil genético de um indivíduo, utilizado para identificação humana, é constituído por diversos marcadores que são herdados dos seus progenitores. Estes marcadores correspondem a seqüências de DNA que variam de um indivíduo para outro, denominadas regiões polimórficas. Na prática forense, normalmente são analisados marcadores autossômicos e do cromossomo Y, sendo a utilização dos marcadores do cromossomo X ainda muito recente.

Entretanto, em alguns casos, como na exumação, em que pouco material biológico está disponível e, em casos complexos de vínculo biológico, como os exames de DNA realizados com o suposto pai falecido ou ausente, a análise dos marcadores tradicionais (autossômicos) pode não solucionar o caso. Nestas situações, faz-se necessária a análise de outros marcadores como as regiões polimórficas do cromossomo X (SZIBOR et al., 2003), as quais são capazes de complementar a análise dos autossômicos e do cromossomo Y de forma muito eficaz.

“Um livro é  
um mudo que fala,  
um surdo que responde, um  
cego que guia,  
um morto que vive”

Padre António Vieira

## **2. REVISÃO BIBLIOGRÁFICA**

## 2.1 Identificação Humana

A identificação humana é o processo pelo qual se determina a identidade de uma pessoa ou coisa, através do estabelecimento de um conjunto de caracteres que a individualize, fazendo-a igual apenas a si mesma (FRANÇA, 2001). De acordo com Sweet (2001), devido ao fato de todos os seres humanos possuírem uma identidade em vida, a sociedade requer que esta identidade seja reconhecida após a morte, seja para o conforto dos familiares da vítima ou para a resolução de questões jurídicas.

Existem dois tipos de processos de identificação humana: o comparativo e o reconstrutivo. O primeiro é baseado em registros anteriores ao óbito e permite a identificação personalista ou individual, podendo ser realizado através da utilização de registros médicos e prontuários odontológicos; já no processo reconstrutivo, não se têm dados anteriores à morte do indivíduo e procura-se realizar a identificação geral definindo-se, por exemplo, o sexo, a idade e a etnia (SASSOUNI, 1963).

Desta forma, nos casos em que o corpo está bem preservado, a identificação humana pode ser feita pelo reconhecimento facial ou pelas características que individualizem a pessoa (como tatuagens ou cicatrizes) e, através da impressão digital ou dentição (WOOD, 2006), desde que existam registros prévios destas características. No entanto, quando o organismo se encontra reduzido ao esqueleto, costuma-se realizar a análise antropométrica e/ou a análise do ácido desoxirribonucléico (DNA) (PRIMORAC; SCHANFIELD, 2000).

Em relação à antropometria, esta se fundamenta numa criteriosa investigação da espécie, da raça, da idade, do sexo, da estatura e das características individuais do esqueleto, porém, não permite a identificação exata do indivíduo (PRETTY; SWEET, 2001). Neste sentido, a Medicina Legal tem-se beneficiado com o avanço indiscutível da análise do DNA,

empregando tal metodologia tanto nas questões de identificação civil como nas de identificação criminal (FRANÇA, 2001).

## **2.2 Marcadores biológicos para identificação humana: panorama histórico**

Teste de identidade humana forense é um campo da Ciência que envolve a identificação de amostras biológicas. Tradicionalmente, as análises de marcadores genéticos protéicos e os produtos dos genes foram tipados para detectar polimorfismos que poderiam ser potencialmente usados para diferenciar indivíduos.

No século XX, a descoberta dos grupos sanguíneos ABO, marcou o início dos estudos modernos sobre a variação genética humana (JORDE; WATKINS; BAMSHAD, 2001). Nesta primeira fase, o exame forense de amostras biológicas era realizado com a aplicação deste clássico sistema, o qual foi utilizado por várias décadas em testes de identificação humana e em evidências relacionadas a crimes. Posteriormente, outros sistemas de proteínas marcadoras como os componentes grupo-específicos (transferrina, albumina, ceruloplasmina, haptoglobina, fosfoglicomutase-1, fosfatase ácida, esterase D, etc) mostraram-se variáveis entre os grupos populacionais e passaram, também, a ser utilizados.

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

O uso de sistemas protéicos para a aquisição de altos níveis de diferenciação entre os indivíduos apresenta, no entanto, alguns inconvenientes, tais como: 1. a relativa baixa estabilidade das proteínas em amostras biológicas expostas ao ambiente; 2. o poder

relativamente baixo de discriminação dos sistemas e, 3. os mesmos marcadores protéicos não existem em todos os tecidos de um indivíduo (WEEDN; SWARNEN, 1998). A tipagem de polimorfismos genéticos pelo DNA contornou muitas dessas limitações, dando início à terceira fase do desenvolvimento das ciências forenses voltadas à identificação humana.

A tipagem do material genético, atualmente uma ferramenta indispensável no campo da Criminalística e da Medicina Legal, surgiu em 1985 quando o geneticista inglês Alec Jeffreys e seu grupo de trabalho descobriram que, em um dos íntrons do gene codificador da proteína mioglobina, havia certas regiões do DNA que variavam de um indivíduo para o outro, produzindo perfis genéticos conhecidos como “impressão digital do DNA” ou “DNA *fingerprinting*” (JEFFREYS; WILSON; THEIN, 1985).

A análise do DNA foi utilizada por Jeffreys, oficialmente pela primeira vez, em 1985, na Inglaterra, para solucionar um problema de imigração no qual um menino de 13 anos, após uma longa estadia com seu pai em Gana, retorna sozinho para a Inglaterra e é detido pela imigração sob alegação de passaporte forjado (JEFFREYS; BROOKFIELD; SEMEONOFF, 1985).

Um ano depois, também na Inglaterra, esta técnica foi utilizada pela primeira vez em um caso criminal, no qual vários homens entre 13 e 30 anos foram tipados para que o verdadeiro estuprador e assassino de 2 mulheres fosse identificado. O caso ficou conhecido como *Enderby* (vila onde ocorreu os crimes) e, desde então, a análise do DNA passou a ser potencialmente utilizada na rotina forense (TANDE, 1989).

Em 2003, a SENASP (Secretaria Nacional de Segurança Pública) iniciou o Projeto de Implantação de Laboratórios Regionais de DNA Forense, com a finalidade de capacitar a polícia para a realização de investigação e elucidação de crimes com um padrão de qualidade internacional.

O projeto propôs a instalação de seis laboratórios regionais, localizados no Rio Grande do Sul (RS), Rio de Janeiro (RJ), Bahia (BA), Distrito Federal (DF), Amazonas (AM) e na sede do Departamento da Polícia Federal (Brasília), os quais funcionariam como referência para o restante do país. Além dos pólos regionais, os laboratórios existentes nas Universidades Federais de Alagoas e do Pará também seriam reaparelhados e funcionariam nos mesmos moldes dos regionais. Cada laboratório ficaria responsável pela demanda de determinados estados, cabendo ao pólo do Rio de Janeiro o atendimento dos Estados de São Paulo, Minas Gerais e Espírito Santo (SBMM, 2007).

Atualmente, da rede dos seis laboratórios propostos, cinco já estão funcionando e apenas o do Estado do Amazonas encontra-se em fase de implantação. Conforme os dados apresentados recentemente no I Encontro Internacional de Genética Forense, em Maceió-AL, em julho de 2007, o equivalente a 8,5 milhões já foi investido nesse projeto pelo Fundo Nacional de Segurança Pública em parceria com os Ministérios da Justiça e de Ciência e Tecnologia.

### **2.3 Polimorfismos genéticos e a análise do DNA**

Na prática forense, o principal objetivo da análise do DNA é distinguir os indivíduos entre si e auxiliar no estabelecimento de uma relação de vínculo biológico em testes de parentesco. Neste contexto, a tecnologia do DNA é altamente discriminativa, permitindo ao cientista forense a habilidade de excluir indivíduos que tenham sido falsamente associados a uma amostra biológica obtida em cena de crime e reduzir o número de contribuidores em potencial da amostra a poucos (se não um único) indivíduos.

Nos testes de identificação humana pelo material genético são estudadas regiões hipervariáveis do DNA, as quais apresentam dois tipos de polimorfismos: o de seqüência e o

de comprimento. O primeiro, é composto de diferentes nucleotídeos em uma determinada localização do genoma e, geralmente, se baseia em meras mutações pontuais (WEEDEN; SWARNEN, 1998). O segundo, corresponde a seqüências de nucleotídeos que se repetem consecutivamente, conhecidas como VNTRs, *variable number of tandem repeats* (MOLLER; BRINKMANN, 1995).

De acordo com o International Human Genome Sequencing Consortium (2001), a seqüência repetitiva (também conhecida como motivo) formada de 14 a 500 pb, é denominada região minissatélite e, a formada de 1 a 13 pb, é conhecida como região microsatélite ou região de repetições consecutivas curtas (STRs, *short tandem repeats*). O número de repetições do motivo varia entre os indivíduos (criando o polimorfismo de comprimento) e estas formas variantes são denominadas “alelos”.

O primeiro método de detecção de regiões polimórficas do DNA baseou-se no uso de sondas para um único locus e nas sondas multi-locus pela técnica de RFLP (*restriction fragment length polymorphism*), a qual permite a análise de regiões VNTRs minissatélites. As sondas multi-locus RFLP são altamente discriminativas, mas a técnica é trabalhosa, demorada, não pode ser facilmente automatizada, requer grandes quantidades de DNA íntegro para a obtenção de bons resultados e, em casos com misturas de amostras, é praticamente impossível a análise com este método. Assim, atualmente, poucas análises são realizadas com RFLP e, nestes casos, são utilizadas as sondas de único locus (BUTLER, 2005).

A tecnologia mais recente, utilizada pela comunidade que atua na área de identificação humana, consiste na amplificação de regiões STRs pela reação de PCR (*polymerase chain reaction*), um processo *in vitro* que aumenta a quantidade de um fragmento pequeno de DNA previamente selecionado (MULLIS; FALOONA, 1987). A característica mais importante da PCR é sua capacidade de obter, relativamente, grandes quantidades de um fragmento de DNA, partindo de quantidades muito pequenas, da ordem de nanogramas ou picogramas

(SWEET, 2001). Comparada com a técnica da RFLP, a PCR apresenta algumas vantagens que incluem o aumento da sensibilidade e especificidade, a diminuição no tempo do ensaio e de trabalho e a fácil automação.

Há estimativas de que 20% das investigações forenses que utilizam o RFLP em suas análises não chegam a um bom termo e essas porcentagens de insucesso são maiores quando o DNA é recolhido de tecido cadavérico (SMITH et al., 1997). Isso ocorre porque, geralmente, a tipagem por RFLP requer DNA relativamente intacto, enquanto algumas amostras de DNA degradadas podem ser amplificadas por PCR e subseqüentemente tipadas, porque os alelos STRs amplificados têm um tamanho muito menor quando comparados aos alelos minissatélites detectados por RFLP. Por isso, a tipagem por PCR é particularmente útil em análises forenses (BUDOWLE; BROWN, 2001).

Os marcadores STRs estão espalhados pelos cromossomos em regiões não codificadoras entre os genes ou dentre eles (introns), correspondendo a, aproximadamente, 3% do genoma humano (SUBRAMANIAN; MISHRA; SINGH, 2003). Estes são classificados de acordo com o número de nucleotídeos que formam o motivo, e os de maior valor para a identificação humana são aqueles que apresentam alto poder de discriminação (maior que 90%), maior frequência de heterozigotos (maior que 70%), baixa taxa de mutação, menor tamanho e baixa formação de *stutters*, fragmentos de DNA com uma unidade a mais ou a menos que o alelo verdadeiro (JOBIM; JOBIM; BRENNER, 1999).

## **2.4 Bancos de Dados de DNA**

Bancos de dados que contenham perfis de DNA de criminosos e/ou perfis de evidências de casos em que não se encontrou o suspeito são de grande ajuda e utilidade na

investigação de certos crimes, uma vez que os perfis de DNA são arquivados em microcomputadores e comparados eletronicamente.

O laboratório do FBI (*Federal Bureau of Investigation*), nos Estados Unidos da América (EUA), vem desenvolvendo o sistema CODIS (*COmbined DNA Index System*) ou Sistema Combinado de Índices de DNA, que combina a Ciência Forense e a Tecnologia Informática, proporcionando uma ferramenta efetiva para o desenvolvimento da investigação criminal. Tal sistema permite a todos os laboratórios americanos (federais, estaduais e locais) realizarem permutas e comparações dos perfis de DNA eletronicamente, além de possibilitar a interligação dos crimes entre si e dos suspeitos envolvidos.

O CODIS iniciou-se como um projeto-piloto nos EUA em 1990 e inicialmente serviu a 14 estados e laboratórios locais. A Lei Pública 103.322 de 1994, contendo o Ato de Identificação pelo DNA, formalizou a autoridade do FBI em estabelecer para os propósitos legais uma indexação do DNA em nível nacional. Então, em outubro de 1998, o Sistema de Índice Nacional de DNA do FBI (NDIS - *Nacional DNA Index System*) tornou-se operacional. Até maio de 2007, o CODIS teve um papel importante nas investigações criminais de 50.343 casos (FEDERAL BUREAU OF INVESTIGATION, 2007).

No Brasil, o estado de Alagoas, com o estímulo de Geraldo de Majella, ouvidor geral do Estado, e o apoio financeiro da FAPEAL – Fundação de Amparo à Pesquisa do Estado de Alagoas, implantou um Banco de Dados Civil de DNA. Este banco tem o objetivo de reunir informações sobre pessoas desaparecidas, que podem estar vivas ou mortas e que não foram identificadas por métodos tradicionais. Neste banco, são arquivadas informações genéticas obtidas a partir do DNA doado pelos parentes das pessoas desaparecidas e informações genéticas obtidas a partir do DNA extraído de vestígios; o cruzamento dessas informações permite a identificação de indivíduos desaparecidos (UNIVERSIDADE FEDERAL DE ALAGOAS, 2007).

Assim como Alagoas, o estado do Rio de Janeiro, por meio da Universidade Federal do Rio de Janeiro (UFRJ) e com o apoio da Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ), está investindo em um banco de dados de DNA para localização de desaparecidos. O centro de pesquisa da UFRJ está ligado ao Projeto Fênix, que foi criado por pesquisadores da Universidade de Granada, Espanha (FUNDAÇÃO DE AMPARO À PESQUISA DO RIO DE JANEIRO, 2003).

A Secretaria Nacional de Segurança Pública (SENASP), em seu projeto ligado ao Ministério da Justiça, também está trabalhando na criação de um banco de dados de DNA criminal (como o CODIS), o qual será constituído por informações genéticas de vítimas e criminosos, além de parentes de pessoas desaparecidas em todo o país.

No Estado de São Paulo, os dados fornecidos pela 2ª Delegacia de Pessoas Desaparecidas do Departamento de Homicídios e Proteção à Pessoa, indicam 8000 registros de desaparecimento de crianças e adolescentes ao ano. Recentemente, o Departamento de Medicina Legal, Ética Médica e Medicina Social e do Trabalho, da Faculdade de Medicina da Universidade de São Paulo, vem executando o projeto denominado “Caminho de Volta”, o qual atende a famílias que tiveram seus filhos desaparecidos com idade inferior a 18 anos. Neste programa, além do Banco de DNA, o qual armazena perfis dos supostos parentes e de crianças desaparecidas (vivas ou mortas), é fornecido um suporte psicológico aos familiares durante todo o processo de busca e são estudadas as causas dos desaparecimentos, visando o estabelecimento de medidas de prevenção e de políticas públicas (GATTÁS; FIGARO-GARCIA, 2007).

Além disso, vários bancos de dados contendo as frequências alélicas de diversas populações estão sendo desenvolvidos, como o Banco do Cromossomo Y (YHRD, *Y-STR Haplotype Reference Database*), o Banco de DNA Mitocondrial (HmtDB, *Human Mitochondrial Genomica Database*) e o Banco do Cromossomo X ([www.chrx-str.org](http://www.chrx-str.org)).

O Banco do cromossomo X foi anunciado em 2005 e contém dados dos X-STRs utilizados na prática forense, em estudos antropológicos e em outras pesquisas genéticas. Neste, pode-se encontrar vários dados de um marcador como: localização física e genética, estrutura de repetição, nomenclatura alélica, taxa de mutação e dados populacionais. Além disso, a página contém um software que realiza o cálculo de diversos parâmetros estatísticos de interesse na prática forense e os dados populacionais devem ser classificados de acordo com as sete metapopulações (Europa, Ásia, América Latina, América do Norte, África, Oceania-Austrália e Pólo Ártico), indicando-se o grupo étnico (SZIBOR; HERING; EDELMANN, 2005).

## **2.5 Cromossomo X**

### **2.5.1 Evolução do cromossomo X**

O cromossomo X possui muitas características que são únicas no genoma, como o fato da mulher herdar o cromossomo X de cada um dos pais, mas o homem herdar apenas um, o materno.

As propriedades originais deste cromossomo são uma consequência da evolução dos cromossomos sexuais nos mamíferos. Tais cromossomos evoluíram de um par de autossomos homólogos (denominados proto-cromossomos sexuais) dentre os últimos 300 milhões de anos (OHNO, 1967). Uma barreira para a recombinação foi desenvolvida entre eles, isolando regiões determinantes do sexo ao longo dos dois cromossomos.

Na ausência de recombinação, o acúmulo de eventos de mutação levou à degeneração do cromossomo Y, tornando os cromossomos sexuais heteromórficos. Neste processo, o cromossomo Y perdeu praticamente todos os traços do ancestral autossômico e tem

desenvolvido um único padrão de seqüência repetitiva (SKALETSKY et al., 2003; JOBLING; TYLER-SMITH, 2003). Por contraste, os elementos funcionais foram conservados no cromossomo X e este é, fisicamente, o mais estável cromossomo nuclear entre os mamíferos placentários (SCHAFFNER, 2004).

Uma comparação feita com cromossomos humanos, de roedores e de galinha, mostrou que o cromossomo X humano é idêntico ao seu ancestral mamífero, tendo permanecido estável em sua organização desde a formação dos mamíferos placentários (HILLIER et al., 2004).

Em um trabalho realizado por Ross et al. (2005), a origem autossômica dos cromossomos sexuais foi ilustrada através do alinhamento entre as seqüências do cromossomo X humano e as seqüências de dois cromossomos autossomos de galinha. O presente estudo mostrou que o braço longo do cromossomo X apresenta grande homologia com o braço curto do cromossomo 4 de galinha e é formado por uma região conservada (*X-conserved region* ou XCR), descendente de um proto-cromossomo X; já seu braço curto, apresenta grande homologia com o braço longo do cromossomo 1 de galinha, sendo constituído por uma pequena região (Xcen -p11.3) conservada e pela adição de um grande fragmento de DNA (*X-added region* ou XAR), decorrente da translocação com um segundo cromossomo autossomo há, aproximadamente, 105 milhões de anos.

### **2.5.2 Características do cromossomo X**

O cromossoma X é uma molécula de DNA linear com aproximadamente 155 Mpb (Mega pares de base), uma baixa densidade gênica (4% do total de genes humanos), genes de tamanho pequeno, um baixo índice de C-G (39% comparado com a média de 41% do genoma) e é altamente rico em regiões repetitivas (56% da eucromatina comparados à média

de 45% do genoma). Foram anotados 1.098 genes, 700 pseudogenes (644 processados e 56 não processados) e 153.146 SNPs, *single nucleotide polymorphisms* (LANDER et al. 2001, ROSS et al. 2005).

Dentre as regiões repetitivas, os elementos intercalares curtos (SINEs, *short interspersed nuclear elements*) foram encontrados abaixo da média e, os longos (LINEs, *long interspersed nuclear elements*) da família 1, apareceram acima da média, compreendendo 29% da seqüência do cromossomo X (média do genoma é de 17%).

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

A baixa densidade gênica do cromossomo X pode, simplesmente, refletir seu ancestral autossômico. No entanto, acredita-se que a seleção pode ter favorecido a transposição de certa classe de genes do cromossomo X para os autossomos durante a evolução dos mamíferos. Tal classe deveria abranger genes cujo produto protéico seria requerido em dose dupla em ambos os sexos ou genes cuja mutação em tecidos somáticos masculinos seria letal.

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

pseudoautossômicas 1 e 2 (PAR1 e PAR2), localizadas nas extremidades dos braços dos cromossomos X e Y.

A PAR1, localizada na extremidade do braço curto dos cromossomos sexuais, compreende 2,7 Mpb e os genes desta região estão presentes em duas cópias em homens e mulheres, não estando sujeitos ao processo de inativação do cromossomo X. Esta região apresenta uma frequência de recombinação elevada e é necessária para a normal segregação dos cromossomas sexuais durante a meiose (ELLIS & GOODFELLOW, 1989).

Em relação à PAR2, localizada na extremidade do braço longo dos cromossomos sexuais, esta possui 330 Kb e foi criada pela duplicação de uma região do cromossomo X para o Y, desde a divergência das linhagens de humanos e chimpanzés (FREIJE et al., 1992). Nesta região a recombinação não ocorre em todas as meioses e alguns genes estão sujeitos a inativação do cromossomo X, refletindo a condição deste cromossomo antes da duplicação.

Além das regiões pseudoautossômicas, os cromossomos sexuais apresentam homologia em regiões não recombinantes (Figura 1), localizadas na XAR e em uma região de transposição do cromossomo X (XTR, *X-transposed region*), situada na XCR, formada pela duplicação de material do X para o Y há 4,7 milhões de anos (SKALETSKY et al., 2003; PAGE et al., 1984).

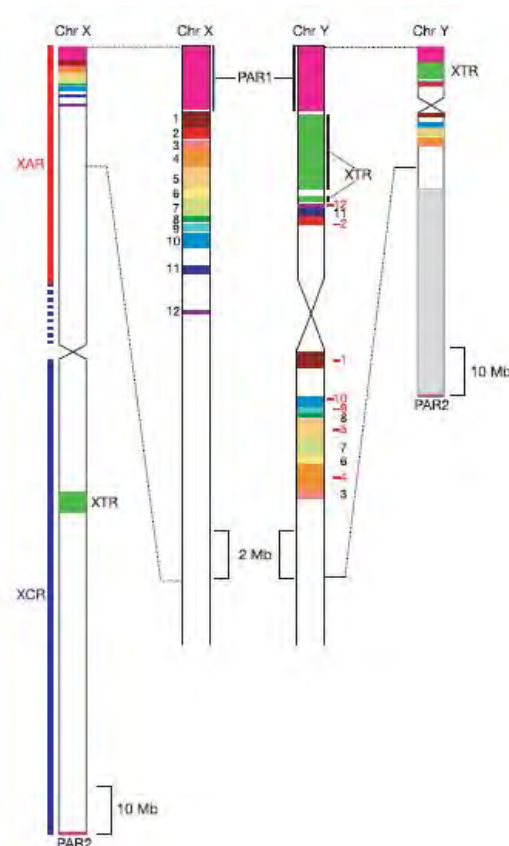


Figura 1: Representação esquemática do cromossomo X e das regiões de homologia (coloridas) entre os cromossomos sexuais (X-Y).

Os marcadores STRs do cromossomo X, aplicados em testes de identificação humana, estão localizados em regiões não recombinantes entre os cromossomos sexuais; mas, a recombinação ocorre entre os cromossomos X da mulher, garantido a variabilidade desses marcadores. Assim, da mesma forma que os autossomos, o cromossomo X se recombina e, similarmente ao cromossomo Y e DNA mitocondrial, tem um modelo que permite determinar diretamente o haplótipo.

Distintas características genéticas são, portanto, esperadas para o cromossomo X em relação aos autossomos (SCHAFFNER, 2004), principalmente: 1 – uma menor diversidade genética devido à menor taxa de mutação em mulheres do que em homens e um menor tamanho populacional efetivo (LANDER et al., 2001; LI; YI; MAKOVA, 2002), 2 – uma

estrutura genética mais pronunciada devido ao menor efeito do tamanho populacional e, conseqüentemente, uma deriva genética mais forte, 3 – um maior desequilíbrio de ligação, uma vez que o cromossomo X recombina apenas na mulher (PEREIRA et al., 2007).

### **2.5.3 Marcadores do cromossomo X e desequilíbrio de ligação**

No processo de recombinação, os dois alelos de um marcador são separados (segregados) para a formação do gameta, o qual irá apresentar metade da informação genética do indivíduo. Quando os marcadores estão em cromossomos separados, estes são ditos não ligados e a segregação dos alelos de diferentes marcadores ocorre de forma independente, apresentando, a respectiva prole, genótipos em proporções mendelianas previsíveis. No entanto, quando os marcadores estão presentes em um mesmo cromossomo, como os STRs do cromossomo X, a segregação não é independente e a proporção genotípica irá depender da taxa de recombinação entre os marcadores.

O processo de recombinação ocorre na meiose, de forma aleatória, alterando o perfil genético herdado em um determinado cromossomo. A probabilidade de tal evento ocorrer entre dois pontos do cromossomo é simbolizada por “ $r$ ”. Assim, se dois marcadores A e B estão ligados, a chance de recombinação entre eles em um indivíduo é “ $r$ ” e a chance de não ocorrer recombinação é “ $1-r$ ”. No caso dos marcadores não ligados,  $r = 0,5$  e, para marcadores ligados, o valor de  $r$  varia um pouco acima de 0, se os dois marcadores estão próximos um do outro (fortemente ligados), até 0,5, se os dois marcadores estão em extremidades opostas no cromossomo (frouxamente ligados) (MARK, 2006).

O desequilíbrio de ligação (LD) é um parâmetro estatístico que se refere à associação não randômica de alelos de diferentes STRs, medindo o desvio da frequência haplotípica em uma dada população (REICH; LANDER, 2001). Para marcadores com alto LD, a frequência

dos haplótipos não pode ser estimada pelas frequências alélicas, sendo necessário obter tal dado diretamente da população (HERING et al., 2006).

Os marcadores STRs se caracterizam por ter uma maior taxa de mutação que os SNPs e, portanto, um menor LD (HERING et al., 2004). No entanto, devido à peculiar herança do cromossomo X, na qual apenas 2/3 destes se recombinam a cada geração, LD pode ocorrer entre os STRs muito próximos, sendo necessário que o grau de ligação entre os marcadores sejam avaliados para a posterior utilização destes na prática forense (ROBINO, 2006).

O desequilíbrio de ligação é interessante em exames de paternidade em que o suposto pai é falecido, uma vez que nestas situações são analisados seus parentes. Suponha um caso de investigação de paternidade do tipo “reconstrução” (Figura 2), em que 3 STRs são analisados na mãe e na criança questionada e, por parte do suposto pai falecido, estão disponíveis sua mãe e dois irmãos biológicos.

	<u>STR1</u>	<u>STR2</u>	<u>STR3</u>
Mãe:	10/13	9/11	7/9
Criança	10/14	11/15	7/8
Mãe do SP:	14/15	Não amplificado	8/9
Irmão1 do SP:	14	10	8
Irmão2 do SP:	15	10	9

Figura 2: Perfis alélicos obtidos em um caso de investigação de paternidade do tipo “reconstrução”. STRs de 1 a 3 correspondem aos marcadores analisados e, SP, ao suposto pai falecido.

Na criança, em vermelho estão assinalados os alelos recebidos pela mãe e, em azul, os alelos paternos obrigatórios (herdados obrigatoriamente do pai). Analisando os perfis dos parentes do suposto pai, observa-se a presença dos alelos paternos obrigatórios nos STRs 1 e 3, mas não no STR2. Se os STRs não estão ligados, pelo mecanismo de herança genética, pode-se dizer que a mãe do suposto pai tem o perfil 10/? para o STR2, sendo que o símbolo “?” pode representar um outro alelo 10 ou qualquer outro alelo da população (incluindo o alelo 15 paterno obrigatório). Assim, devido a não amplificação deste marcador, não se pode excluir a paternidade. No entanto, se os STRs estão ligados, estes são herdados na forma de haplótipo (em bloco) e pode-se deduzir o perfil completo da mãe do suposto pai, mesmo não havendo a amplificação do STR2, concluindo-se que esta apresenta em um dos seus cromossomos os alelos 14/10/8 e, no outro, os alelos 15/10/9 para os STRs 1, 2 e 3, respectivamente. Assim, observa-se que há exclusão de paternidade no STR2, pois a mãe (10/10) não poderia ter um filho com o alelo 15 paterno obrigatório.

Diante dessa característica haplotípica, dois tipos de mapas do cromossomo X têm sido estabelecidos em genética: o mapa físico e o mapa genético.

O mapa físico mostra a distância física (em mega pares de base - Mpb) entre os marcadores e é obtido através de um painel de células somáticas híbridas, criadas pela fusão entre uma célula humana de linfoblasto marcada com isótopo radioativo e uma célula de *hamster* não radioativa. O mapa genético, por sua vez, exhibe a distância genética (em centimorgan - cM), a qual reflete a taxa de recombinação entre os marcadores, e é obtido pela análise do heredograma. Assim, a distância genética de 1 cM equivale a 1 esperada recombinação por 100 meioses. É importante ressaltar que as distâncias genéticas e físicas não são equivalentes (NAGARAJA, 1998).

Segundo Szibor et al. (2003), quatro possíveis grupos de ligação foram estabelecidos no cromossomo X, considerando-se 16 STRs de grande interesse na prática forense. Em tais

grupos, marcadores muito próximos podem formar haplótipos que recombinaem durante a meiose com uma frequência correspondente a distância genética entre eles. Assim, para estes marcadores, se confirmada a ligação, o ideal é analisar o haplótipo como um todo em vez de seus constituintes alélicos.

Alguns desequilíbrios de ligação foram previamente reportados para os STRs DXS101 - DXS7424 (ELDEMAN et al., 2002; POETSCH et al., 2005); DXS7130 - DXS6803 (ELDEMAN; SZIBOR, 2003), DXS6801 - DXS6809 - DXS6789 (SZIBOR et al., 2005; GOMES et al., 2007a) e DXS10079 - DXS10074 - DXS10075 (HERING et al., 2006). No entanto, o cromossomo X parece estar mais propenso à recombinação e a ausência de ligação tem sido relatada em outros estudos, até mesmo para os marcadores citados acima (LEE et al., 2004; HERING et al., 2004; ROBINO et al., 2006).

#### **2.5.4 Aplicações dos marcadores genéticos do cromossomo X**

Em virtude do alto poder de individualização e praticabilidade, a análise dos STRs tornou-se uma rotina amplamente utilizada na prática forense. A grande maioria dos exames de vínculo biológico, aborda os STRs localizados em autossomos e cromossomo Y, sendo a análise de marcadores do cromossomo X algo ainda recente (ASAMURA et al., 2006a).

No entanto, os X-STRs são capazes de complementar a análise dos marcadores autossômicos e do cromossomo Y de forma muito eficiente, pois estes marcadores se caracterizam por apresentarem maiores valores de MECs (*mean exclusion chance*) que os autossômicos, ou seja, possuem uma maior capacidade de excluir de uma situação de investigação de paternidade, indivíduos não relacionados à criança testada e, em virtude disso, suficiente poder estatístico é obtido quando poucos X-STRs são analisados (SZIBOR et al., 2003).

A maior capacidade de exclusão dos X-STRs se dá pela diferença de alelos existente entre os tradicionais marcadores autossômicos e os do cromossomo X para os indivíduos do sexo masculino. Estes apresentam sempre dois alelos para os marcadores autossômicos e, apenas um, para os do cromossomo X. Assim, no mecanismo de herança genética, o pai transfere seu perfil haplotípico (100%) do cromossomo X para a filha, o que não ocorre com os autossômicos, nos quais apenas 50% da informação genética paterna é herdada.

Em virtude dessa propriedade do cromossomo X, a maior vantagem da sua aplicação reside nos casos deficientes de paternidade, nos quais o material biológico do suposto pai não está disponível e o DNA de seus parentes é analisado para a reconstrução do seu perfil genético (GOMES et al., 2007a). Com a análise dos tradicionais STRs autossômicos, muitas vezes estes casos permanecem inconclusivos, pois para se ter uma boa reconstrução do perfil do suposto pai, vários parentes deste precisam ser tipados. No entanto, por este fato encarecer o custo do exame e/ou por algumas pessoas se negarem a doar o material biológico, normalmente poucos parentes são analisados, gerando um baixo índice de vínculo biológico.

Nos casos de reconstrução com emprego de X-STRs, a mãe do suposto pai é bastante importante na análise, pois os alelos paternos podem ser facilmente determinados a partir dela, já que o pai herda apenas um cromossomo X, o materno (ROSS et al., 2005). Assim, quando a suposta avó paterna está presente, tais casos podem ser facilmente solucionados; o que não ocorre com os autossômicos, sendo necessária a tipagem de mais parentes do suposto pai. Além disso, se a suposta avó não está presente, seu perfil pode ser facilmente determinado através de suas filhas e, muito mais facilmente, a partir dos irmãos do suposto pai, o que aumenta as circunstâncias em que bons resultados podem ser obtidos com os X-STRs.

No caso do cromossomo X, por exemplo, quando dois indivíduos do sexo feminino possuem o mesmo pai, estas também apresentam o mesmo cromossomo X paterno. Assim,

em uma investigação em que duas irmãs ou meias-irmãs são tipadas, a presença de quatro diferentes alelos em um marcador do cromossomo X permite a exclusão da paternidade comum entre elas, até mesmo se nenhum dos pais biológicos ou seus parentes forem analisados. Através de marcadores autossômicos, ambas poderiam realmente ser irmãs, já que o pai pode ter doado um de seus alelos para uma e o seu outro alelo, para a outra. Turrina; Atzei; Leo (2007) utilizaram apenas 4 X-STRs para auxiliar na resolução de um caso de meia-irmãs, no qual a análise de 15 autossômicos não foi suficiente.

Tal cromossomo é, também, de grande utilidade em casos de paternidade em que os supostos pais são parentes. Por exemplo, se os supostos pais são pai e filho, no caso de STRs autossômicos, estes obrigatoriamente compartilham 50% dos alelos e o exame com um deles pode aparecer com apenas uma inconsistência genética, o que dificulta a diferenciação entre mutação paterna e relativa proximidade do suposto pai com o verdadeiro pai biológico da criança. No caso dos X-STRs, nenhum alelo entre pai e filho é idêntico por descendência, uma vez que estes herdam o cromossomo X de suas mães e ambas são diferentes, o que facilita a resolução desses casos. Já no caso dos supostos pais serem dois irmãos, ambos receberam o cromossomo X da mesma mãe e por isso há uma chance de 50% destes possuírem alelos idênticos por descendência, da mesma forma como ocorre com os STRs autossômicos, sendo ambos os sistemas equivalentes nesta situação.

Em um estudo realizado por Silveira et al. (2007), quatro casos complicados de paternidade, compreendendo dois trios (mãe/filha/suposto pai) e duas duplas (filha/suposto pai), obtiveram baixo índice de paternidade devido à presença de uma suposta mutação paterna, permanecendo inconclusivos. Ao se analisar 8 X-STRs, em todos os casos, o haplótipo paterno foi diferente do apresentado pela filha e o suposto pai foi excluído como pai biológico desta criança. Assim, a aplicação dos X-STRs mostrou ser uma poderosa ferramenta na resolução de casos que apresentaram poucas inconsistências após análise de autossômicos.

Os X-STRs podem também ser aplicados em casos de estupro com feto do sexo feminino. Após 6 a 8 semanas de gestação, quando ocorre o aborto, é muito difícil obter o material apenas do embrião, recuperando-se, geralmente, uma mistura de material embrionário com sangue materno. Neste caso, apenas autossômicos e X-STRs podem ser utilizados e a prova de paternidade recai sobre os alelos do embrião que não são compartilhados com a mãe. Como para os X-STRs o homem apresenta apenas um alelo, estes são melhores que os autossômicos, já que a probabilidade de se excluir um indivíduo falsamente acusado é maior. Em casos de incesto (pai-filha), no entanto, a aplicação dos X-STRs fica limitada, pois todos os alelos do embrião estão presentes na mãe, ao contrário dos STRs autossômicos, nos quais pode-se obter alelos do embrião que foram obrigatoriamente herdados do pai (SZIBOR et al., 2003).

Em virtude das altas taxas de ilegitimidade paterna vistas na sociedade moderna, a identificação de esqueletos através da análise de mãe-filho é mais requerida que a de pai-filho. Para os testes mães-filhas, os X-STRs são equivalentes aos autossômicos; no entanto, para mães-filhos, tais marcadores são extremamente eficientes, já que o filho possui apenas um alelo para cada X-STR e este deve obrigatoriamente estar presente na mãe. A chance de exclusão nesses casos é idêntica para os testes pai-filha. De acordo com Toni; Domenici; Presciuttini (2007), nestes casos, os MECs obtidos são quase duas vezes maiores que os valores conseguidos para STRs autossômicos com moderada heterozigose (0,6-0,75).

### **2.5.5 Estudos populacionais com marcadores genéticos do cromossomo X**

Populações que se encontram em regiões diferentes ou apresentam origem distinta podem apresentar alelos em frequências diferentes e algumas vezes únicos. Portanto, quando se deseja identificar um indivíduo proveniente de determinada população, é necessário o

estudo de diferentes marcadores naquela população para saber quais são os alelos presentes e em que frequência, a fim de se definir quais são os melhores marcadores a serem utilizados nestes casos.

Tal distribuição alélica será aplicada nos cálculos estatísticos realizados em testes de identificação humana e de paternidade a fim de dar força e credibilidade às informações obtidas pela análise da evidência genética (amostra biológica). A base para esses cálculos de frequências de alelos derivou dos estudos de Hardy-Weinberg e, uma vez determinadas as frequências individuais de cada marcador genético, estas são multiplicadas e o poder de combinação desses marcadores possibilita ao cientista adquirir altos valores discriminatórios.

Alguns estudos com marcadores do cromossomo X têm sido descritos na literatura, no entanto, estes compreendem poucos grupos populacionais: Espanha (ZARRABEITIA et al., 2002; 2005; 2006), Germânia (ELDEMAN; SZIBOR, 2001; SZIBOR et al., 2005; POETSCH et al., 2005), Itália (BINI et al., 2005; ROBINO et al., 2006; TURRINA; ATZEI; LEO, 2007); Ásia – Koreanos (LEE et al., 2004; SHIN KJ et al., 2004; SHIN SH et al., 2005); Ásia – Chineses (LIU & LI, 2005; GU et al., 2005; GAO et al., 2007); Estados Unidos da América (GOMES et al., 2007a), Portugal (PEREIRA et al., 2007) e outros.

Zarrabeitia et al. (2006) compararam cinco diferentes regiões da Espanha para o conjunto de STRs (DXS9895, DXS9898, DXS7130, DXS7132, DXS6789 e GATA172D05) e não observaram diferenças significativas nas distribuições alélicas. Igualmente, em uma pesquisa realizada por Robino et al. (2006), também não foi encontrada diferença significativa entre algumas populações italianas.

Shin SH et al. 2005 (27) realizaram estudo populacional referente a 18 STRs (DXS6807, DXS8378, DXS9895, DXS9902, DXS6810, DXS7132, DXS981, DXS6800, DXS9898, DXS6789, DXS101, DXS6797, GATA172D05, GATA165B12, HPRTB, GATA31E08, DXS8377 e DXS7423) na população da Coreia, observando-se que quatro

deles apresentaram mutações (GATA172D05, GATA31E08, DXS7132, HPRTB) e que a maioria dos marcadores apresentou diferença significativa na distribuição alélica quando comparados com as populações da Alemanha, China e Espanha.

Em outro estudo, Tabbada et al. (2005) estudaram os STRs DXS101, DXS8377, DXS981, HPRTB e DXS6789 nas populações da Alemanha, China, Japão, Tailândia e Filipinas, obtendo-se diferenças significativas entre a Alemanha e as demais populações. Da mesma forma, GOMES et al. (2007a), ao comparar 10 X-STRs em três populações dos Estados Unidos (africanos, hispânicos e asiáticos), concluiu que um banco de dados comum não pode ser utilizado para este sistema, uma vez que diferenças significativas foram observadas entre tais grupos.

Diante do exposto, o cromossomo X parece ter particularidades na sua distribuição alélica, o que torna o estudo populacional extremamente importante como mais uma ferramenta na atividade forense.

O potencial dos marcadores genéticos presentes nos autossomos, no cromossomo Y e no DNA mitocondrial tem sido amplamente explorado pela genética populacional e forense. No entanto, falta uma maior contribuição do cromossoma X, uma vez que apenas um número muito limitado de polimorfismos deste cromossomo foi descrito e bem caracterizado, e os dados populacionais existentes são escassos (SZIBOR et al. 2006).

O Brasil, com sua população em proporções continentais, possui apenas um trabalho até o momento, na população de Santa Catarina (CAINÉ et al., 2007), sendo necessário que mais estudos sejam realizados com tais marcadores, a fim de compilar informações, as quais, posteriormente, poderão alimentar um banco de dados da população brasileira, viabilizando sua aplicação no contexto forense.

Além de gerar dados populacionais úteis na prática forense, a idéia de se realizar este projeto teve a finalidade de aperfeiçoar o serviço oferecido pelo Laboratório de Investigação

de Paternidade da Faculdade de Ciências Farmacêuticas da Universidade Estadual Paulista (FCF-UNESP) de Araraquara-SP, o qual vem apresentando uma crescente solicitação de exames de investigação de paternidade com suposto pai é falecido, situações em que a aplicação de STRs do cromossomo X pode ser significativa para resolução do caso.

“Tenho em mim todos os  
sonhos do mundo”

Fernando Pessoa

### **3. OBJETIVOS**

Os objetivos deste trabalho foram:

- Determinar a frequência alélica de 5 STRs do cromossomo X (DXS6854, DXS7424, DXS101, DXS6808 e DXS7132) em indivíduos da população brasileira, residentes em Araraquara, os quais se autodeclararam como branco, preto, pardo ou amarelo;
  
- Estimar o valor estatísticos desses STRs na prática forense e em testes de paternidade.

“Quando ouvires os  
aplausos do triunfo, que  
ressoem também aos teus  
ouvidos os risos que  
provocaste com os teus  
fracassos”

Josemaría Escrivá

## **4. MATERIAL E MÉTODOS**

#### **4.1 Aspectos éticos**

Este trabalho se enquadra dentro de um projeto mais amplo intitulado “Estudo de frequência alélica de cinco loci STR do cromossomo X na população do Estado de São Paulo e sua contribuição na identificação humana”, coordenado pelo Prof. Dr. Rogério N. de Oliveira da Faculdade de Odontologia da Universidade de São Paulo (FOUSP). Portanto, foi aprovado pelo Comitê de Ética em Pesquisa desta Faculdade (Protocolo nº 84/06, 07 de julho de 2006).

#### **4.2 Casuística**

Neste projeto foram analisadas amostras sanguíneas de 120 indivíduos (60 homens e 60 mulheres) não aparentados e residentes em Araraquara, os quais, após serem informados a respeito dos objetivos deste estudo e dos procedimentos que seriam realizados, assinaram o respectivo Termo de Consentimento Livre e Esclarecido.

De acordo com os padrões do Instituto Brasileiro de Geografia e Estatística (IBGE), os indivíduos foram classificados, segundo a cor da pele, como branco, preto, pardo ou amarelo e para cada grupo foi coletado 15 homens e 15 mulheres. Os indivíduos pretos que participaram deste trabalho são descendentes de pais pretos e, os indivíduos pardos, apresentam um dos pais preto e, o outro, branco.

As amostras sanguíneas foram coletadas em cartões FTA *Classic* (Whatman-Bioscience), em Araraquara, no Laboratório de Investigação de Paternidade da Faculdade de Ciências Farmacêuticas da Universidade Estadual Paulista Júlio de Mesquita Filho (FCF-UNESP), na comunidade Nipo-brasileira e na ONG EDUCAFRO (Organização Não-Governamental de Educação e Cidadania de Afrodescendentes e Carentes).

Nos locais citados anteriormente, as amostras dos indivíduos do sexo masculino foram coletadas em 2005 para o Projeto intitulado “Estudo de frequências alélicas de 12 loci do cromossomo Y na população brasileira de Araraquara e da região da grande São Paulo” (Góis et al., 2008), sendo reutilizadas, após novo termo de consentimento, para o desenvolvimento deste projeto.

### **4.3 Extração do DNA**

Para a extração do DNA, 15 discos de 1,2 mm de papel de filtro FTA, contendo a amostra sanguínea do indivíduo, foram cortados com o auxílio do *Harris Micro Punche e Cutting Mat* (Whatman-Bioscience) e depositados em microtubos de 1,5 mL. Em seguida, procedeu-se à extração com o Kit DNA IQ (Promega), de acordo com as normas do fabricante.

Assim, promoveu-se a lise celular por meio de uma solução composta por 250  $\mu$ L de *Lysis Buffer* e 2,5  $\mu$ L de 1M Dithiothreitol (DTT). O DNA foi “capturado” por 7  $\mu$ L de uma resina composta por partículas magnéticas, lavado por três vezes com 300  $\mu$ L de solução *Wash Buffer* e eluído em 100  $\mu$ L de *Elution Buffer*.

### **4.4 Reação de PCR**

Neste trabalho foi realizada a amplificação de 5 STRs do cromossomo X (X-STRs), sendo três tetranucleotídeos (DXS6854, DXS6808 e DXS7132) e dois trinucleotídeos (DXS7424 e DXS101).

A distribuição de tais marcadores pode ser visualizada na Figura 3. Algumas informações referentes a estes, como a localização, o número de acesso no Banco do Genoma

(www.gdb.org) e a seqüência repetitiva (motivo), estão apresentadas na Tabela 1 e, a seqüência dos primers utilizados na amplificação, na Tabela 2, a seguir.

A seqüência dos primers foi retirada do Banco do Genoma e o primer de orientação forward foi marcado pela *Integrated DNA Technologies* (IDT) com o fluoróforo Cianina 5 (Cy5).

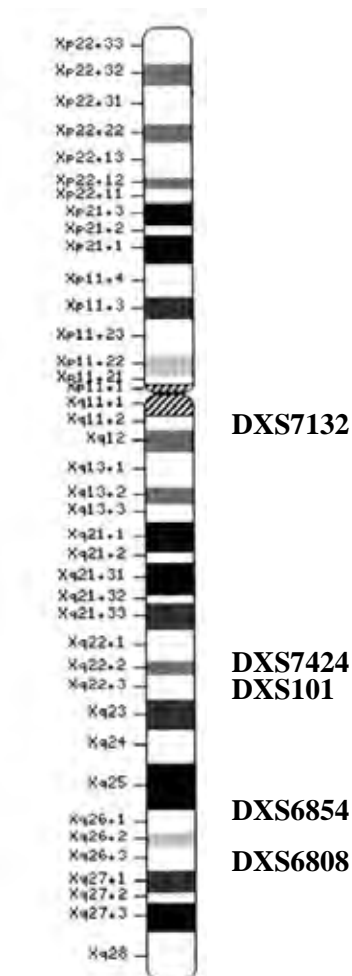


Figura 3: Ideograma do cromossomo X com a localização dos STRs analisados neste trabalho.

Tabela 1: Características dos cinco X- STRs.

<b>STR</b>	<b>Localização</b>	<b>Nº Acesso GDB</b>	<b>Seq. Repetitiva (5'-3')</b>
DXS6854	Xq25-q26.1	386910	(ATTT) <sub>n</sub>
DXS7424	Xq21.33-q22	577646	(TAA) <sub>n</sub>
DXS101	Xq21.33-q22.3	207667	(CTT) <sub>n</sub> (ATT) <sub>n</sub>
DXS6808	Xq26.3	365487	(TCTA) <sub>n</sub>
DXS7132	Xq11.1- q12	685605	(TCTA) <sub>n</sub>

Tabela 2: Sequência dos primers utilizados para os cinco X- STRs.

<b>STR</b>	<b>Orientação</b>	<b>Sequência (5' → 3')</b>
DXS6854	Forward	AGCACTTCTCCTACAACCCTC
	Reverse	CAGCCTGGGCAGTAGAGACT
DXS7424	Forward	CTGCTTGAGTCCAGGAATTCAA
	Reverse	GAACACGCACATTTGAGAACATA
DXS101	Forward	ACTCTAAATCAGTCCAAATATCT
	Reverse	AAATCACTCCATGGCACATGTAT
DXS6808	Forward	ATCCTGATATCTGCCTTTAAATG
	Reverse	GGTATATAACAATTGTTGGTGCA
DXS7132	Forward	AGCCCATTTTCATAATAAATCC
	Reverse	AATCAGTGCTTTCTGTACTATTGG

Inicialmente, 15 amostras masculinas foram submetidas à amplificação dos 5 X-STRs por um método denominado Monoplex e, posteriormente, as demais amostras foram

amplificadas por um método denominado Multiplex, descritos a seguir. Em todas as reações de PCR, foi amplificado o controle positivo K562 (Promega).

#### 4.4.1 Método Monoplex

Cada STR do cromossomo X foi amplificado separadamente em microtubos de 0,2 mL, utilizando-se:

7,12  $\mu$ L de Água Livre de Nuclease (Promega)

1  $\mu$ L de *Gold STAR 10X Buffer* (Promega)

0,3  $\mu$ L de Primer Mix (3  $\mu$ mols de cada primer - *Forward e Reverse*)

0,08  $\mu$ L (0,4U) de Platinum<sup>®</sup>*Taq* DNA Polymerase (Invitrogen)

1,5  $\mu$ L de DNA (1-2 ng)

10  $\mu$ L de volume total de reação

A amplificação foi realizada em termociclador gradiente MJ96G (Biocycler) conforme o ciclo abaixo:

94 °C – 5 min

94 °C – 1 min

55 °C – 1 min

72 °C – 2 min

} 35X

72 °C – 10 min

04 °C –  $\infty$

#### 4.4.2 Método Multiplex

Neste método, 4 STRs (DXS6854, DXS7424, DXS6808 e DXS7132) foram amplificados em sistema Multiplex, sendo apenas o STR DXS101 amplificado em sistema Monoplex, como descrito anteriormente.

As reações foram realizadas em microtubos de 0,2 mL, utilizando-se:

5,8 µL de Água livre de nuclease (Promega)

1 µL de Gold ST★R 10X Buffer (Promega)

0,35 µL de cada Primer Mix – DXS6854, DXS7424, DXS6808 e DXS7132 (3,5 pmols de cada primer – *Forward* e *Reverse*)

0,3 µL (1,5U) de Platinum<sup>®</sup> *Taq* DNA Polymerase (Invitrogen)

1,5 µL de DNA (1-2ng)

10 µL de volume total

A amplificação foi realizada em termociclador GeneAmp<sup>®</sup> PCR System 9700 (Perkin-Elmer) como segue.

95 °C – 5 min	95 °C – 1 min	} 2x
95 °C – 1 min	59 °C – 1 min	
67 °C – 1 min	72 °C – 1 min	} 30x
72 °C – 1 min	95 °C – 1 min	
95 °C – 1 min	55 °C – 1 min	} 30x
63 °C – 1 min	72 °C – 1 min	
72 °C – 1 min	72 °C – 10 min	
	04 °C – ∞	

Após a reação de PCR, os fragmentos amplificados foram avaliados e quantificados visualmente em gel de agarose 1% para o Método Monoplex e, gel de agarose 4% para o Método Multiplex, utilizando-se a Cuba *Horizon* 58, Modelo 200 (Gibco) por 20 e 50 minutos respectivamente, a 105 volts (V). Na preparação das amostras para aplicação no gel, 5,0 µL do produto da PCR foram adicionados a 1,5 µL de Tampão de Amostra I.

A intensidade das bandas foi comparada com as do marcador *Low DNA Mass Ladder* (Invitrogen) após coloração com 2,0 µL de Brometo de Etídio (Gibco) e a imagem do gel foi visualizada e fotografada em aparelho Flúor-S Multimager (Bio-Rad).

Os produtos da amplificação foram mantidos à temperatura ambiente até o momento da corrida eletroforética em seqüenciador automático ALF Express (GE Healthcare).

#### **4.5 Obtenção dos padrões internos da corrida em seqüenciador automático**

Além dos cinco STRs do cromossomo X, foram também amplificados os STRs D9S938 de um indivíduo homozigoto para o alelo de tamanho de 405 pb e o STR SY255 de um indivíduo masculino portador do alelo de tamanho de 126 pb. Estes dois fragmentos de DNA foram utilizados como padrão interno (PI) da corrida, sendo, portanto, adicionados em todas as amostras preparadas para a eletroforese em seqüenciador automático.

Algumas informações referentes a estes STRs, como a localização, o número de acesso no Banco do Genoma e o tamanho do fragmento amplificado, estão apresentadas na Tabela 3 e, a seqüência dos primers utilizados na amplificação, na Tabela 4.

Da mesma forma que os X-STRs, a seqüência dos primers foi retirada do Banco do Genoma e o primer de orientação *Forward* foi marcado pela *Integrated DNA Technologies* (IDT) com o fluoróforo Cianina 5 (Cy5).

Tabela 3: Características dos Padrões Internos da corrida

<b>STR</b>	<b>Localização</b>	<b>Nº Acesso GeneBank®</b>	<b>Tamanho do fragmento (bp)</b>
D9S938	9pter - 9qter	686037	405 pb
SY255	Yq11.223	6014112	126 pb

Tabela 4: Sequência dos primers utilizados para os Padrões Internos da corrida

<b>STR</b>	<b>Orientação</b>	<b>Sequência (5' → 3')</b>
D9S938	Forward	GTAAGGGGTTGAGGTTTTGC
	Reverse	CACCACATTTCTGACATCCA
SRY255	Forward	GTTACAGGATTCGGCGTGAT
	Reverse	CTCGTCATGTGCAGCCAC

A amplificação destes STRs foi realizada, separadamente, em microtubos de 0,2 mL, utilizando-se:

41,1 µL de Água livre de nuclease (Promega)

5 µL de Gold STR★10X Buffer (Promega)

1 µL de Primer Mix (10 pmols de cada primer - *Forward* e *Reverse*)

0,4 µL (2U) de Platinum® *Taq* DNA Polymerase (Invitrogen)

2,5 µL de DNA

50 µL de volume total

Ambos STRs foram amplificação em termociclador gradiente MJ96G (Biocycler) conforme o ciclo a seguir:

94 °C – 5 mim  
 94 °C – 1 mim }  
 55 °C – 1 mim } 35X  
 72 °C – 2 mim }  
 72 °C – 5 mim  
 04 °C – ∞

#### 4.6 Detecção e análise dos produtos da PCR

Para a obtenção dos perfis alélicos, os produtos da PCR foram submetidos à eletroforese em gel de poliacrilamida desnaturante a 6% no Seqüenciador Automático ALF Express (GE Healthcare).

As amostras foram preparadas de formas diferentes dependendo do método utilizado na amplificação. No caso do Método Monoplex, 1 µL do produto de PCR de cada STR do cromossomo X (total 5 µL) foi misturado com 7,5 µL do Tampão de Amostra II. Em relação ao Método Multiplex, 2,5 µL do produto de PCR Multiplex e 1,5 µL do produto de PCR do DXS101 (em sistema Monoplex) foram misturados com 6 µL do Tampão de Amostra II.

Antes da aplicação no gel, as amostras foram desnaturadas a 95°C por três minutos e a amostra controle K562 (Promega) e o Ladder (Preparado “in House” pelo Laboratório Biocod – Belo Horizonte, Minas Gerais) foram colocados em todas as corridas.

A corrida eletroforética foi realizada durante 2 horas e 40 minutos nas seguintes condições: Voltagem - 600 V, Corrente - 40 mA, Potência - 15 W, Temperatura - 45 °C.

A análise dos produtos de PCR para obtenção dos respectivos perfis alélicos foi feita com o auxílio do programa *Allelelocator* v. 1.03 (GE Healthcare), sendo os resultados confirmados visualmente em cada eletroferograma.

Pelo fato da escada alélica (Ladder) utilizada no processo estar padronizada em pares de base (pb), os perfis dos indivíduos foram obtidos nesta mesma forma e, após a leitura dos eletroferogramas, os fragmentos em pb foram convertidos para seus respectivos alelos, de acordo com os dados da literatura.

#### **4.7 Análise estatística dos dados**

Após obtenção dos perfis, as frequências alélicas dos 5 X-STRs foram determinadas pelo método da frequência relativa, ou seja, através do número de repetições do alelo observado nas amostras. Primeiramente, estabeleceu-se a frequência alélica para os homens e as mulheres, de acordo com os grupos branco, preto, pardo e amarelo e, posteriormente, esta foi determinada para a população total analisada.

No caso do sexo masculino, também pelo método da contagem direta, determinou-se a frequência dos haplótipos e, de acordo com a fórmula descrita por Nei (1987), calculou-se a diversidade haplotípica.

Diversos parâmetros estatísticos de interesse forense foram cuidadosamente determinados, tais como: Heterozigose observada (HETobs), pelo método de contagem direta; Heterozigose Esperada (HETesp), com auxílio do programa Arlequin v.3,1 (EXCOFFIER; LAVAL; SCHENEIDER, 2005); Poder de Discriminação Feminino (PD<sub>F</sub>) e Masculino (PD<sub>M</sub>), como proposto por Desmarais et al. (1998) e Chance de Exclusão Significativa (MEC) para duas diferentes situações:

- MEC<sub>T</sub>: Chance de Exclusão Significativa para marcadores do cromossomo X em trios envolvendo filhas, como descrito por Desmarais et al. (1998);

-  $MEC_D$ : Chance de Exclusão Significativa para marcadores do cromossomo X na ausência da mãe (dupla pai-filha), como proposto por Desmarais et al. (1998).

Para avaliar a divergência do equilíbrio de Hardy-Weinberg (HWE), foi realizado o teste exato de acordo com Guo e Thompson (1992), utilizando-se os dados de origem feminina e o programa Arlequin v.3,1.

Além disso, utilizando-se frequências alélicas e o programa Arlequin v. 3,1, os dados obtidos neste trabalho foram comparados entre os grupos de cor de pele e entre outras populações através do teste exato de diferenciação de populações e da determinação da medida de distância genética ( $R_{ST}$ ).

“Não há nada de nobre em sermos superiores ao próximo. A verdadeira nobreza consiste em sermos superiores ao que éramos antes”

Autor desconhecido

## **5. RESULTADOS E DISCUSSÃO**

## 5.1 Protocolo de extração

Optou-se por realizar a coleta da amostra sanguínea em cartão FTA *Classic* (Whatman-Bioscience), pois este é impregnado com uma mistura de substâncias químicas responsáveis por lisar a membrana celular, imobilizar e estabilizar os ácidos nucleicos e inibir o crescimento de bactérias e vírus, permitindo a conservação da amostra à temperatura ambiente por no mínimo 4 anos (SMITH; BURGOYNE, 2004). Assim, as amostras biológicas podem ser transportadas e armazenadas de forma prática, facilitando a rotina laboratorial.

Como as amostras masculinas estavam armazenadas desde 2005 e a reação de PCR seria realizada no Laboratório Biocod (Belo-Horizonte-MG), para que não houvesse problemas em relação à qualidade do DNA extraído, optou-se por utilizar o Kit DNA IQ (Promega), pois este é altamente eficiente na extração de DNA de material biológico antigo e promove a obtenção de DNA de qualidade e em quantidade adequadas para a reação de amplificação. Tal capacidade se deve, principalmente, a presença de partículas magnéticas que possuem afinidade pela molécula de DNA.

Esta técnica consiste na utilização de uma solução composta por Resina, *Lysis Buffer* e Dithiothreitol (DTT); a resina possui partículas magnéticas que capturam o DNA, fixando-o à parede do microtubo quando este é colocado na estante magnética, o *Lysis Buffer*, lisa a membrana celular das células sanguíneas, liberando o DNA e, o DTT, responsável pela redução de pontes dissulfeto, promove a desnaturação protéica. Durante o processo de extração também é utilizada a solução *Wash Buffer* para retirada de substâncias não aderidas à resina e, a solução *Elution Buffer* (Tris-EDTA), para a eluição do DNA da resina, após aquecimento a 56°C. O EDTA inibe a atividade de desoxirribonucleases (DNAses) devido ao fato de atuar como agente quelante de íons essenciais ( $Mg^{2+}$ ) para a atividade destas enzimas.

Para amostras sanguíneas coletadas em papel FTA, a quantidade de resina utilizada (7 $\mu$ L por amostra) captura, no máximo, 100 ng de DNA (1ng DNA/ $\mu$ L), não sendo necessária a diluição das amostras para posterior utilização na reação de amplificação.

## 5.2 Avaliação dos métodos de PCR: Monoplex e Multiplex

Inicialmente, cada STR foi testado pelo método Monoplex em 15 amostras masculinas apenas para avaliar a “performance” da reação de PCR mediante as condições de reagentes e ciclos propostos nesta metodologia, uma vez que a utilização desse sistema é inviável atualmente. Os produtos de PCR obtidos por este método podem ser visualizados na Figura 4, a seguir.

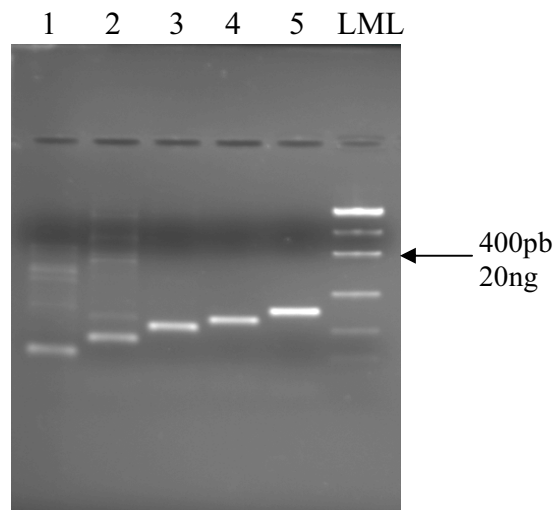


Figura 4: Gel de agarose 0,8% em TAE [1x], corado com BrEt. Produto da reação de amplificação dos STRs DXS6854, DXS7424, DXS101, DXS6808 e DXS7132 pelo método Monoplex, correspondentes as bandas obtidas nas canaletas de 1 a 5, respectivamente. LML: *Low DNA Mass Ladder*.

Através da Figura 4, observa-se a amplificação de todos os marcadores; esta ocorreu de forma extremamente específica para os marcadores DXS101, DXS6808 e DXS7132, mas não para os marcadores DXS6854 e DXS7424. Como a finalidade era facilitar a rotina laboratorial, neste método todos os STRs foram amplificados a uma temperatura de anelamento de 55°C, no entanto, segundo trabalhos na literatura, para os marcadores DXS6854 e DXS7424, o ideal é 61°C (QUIULIN et al., 2002) e 60°C (EDELMAAN et al., 2002), respectivamente, o que explica tal fato.

Assim, com o objetivo de diminuir o tempo e o custo do processamento das amostras, facilitando a prática laboratorial, foi testado o sistema Multiplex com os 5 X-STRs, utilizando-se a técnica de *Touchdown* (TD) PCR. Este processo consiste em minimizar a necessidade de otimização da temperatura de anelamento dos primers e as condições do tampão para se obter grandes quantidades dos amplicons de interesse. Inicialmente, o ciclo começa com uma temperatura de anelamento elevada e esta diminui gradativamente durante os ciclos até uma temperatura abaixo da ideal de anelamento, o que permite um maior rendimento na reação, mantendo-se a especificidade desta. (HECKER; ROUX, 1996).

Por motivos não determinados, não foi possível obter um bom resultado com o marcador DXS101 no sistema Multiplex, permanecendo a amplificação deste pelo método Monoplex. Segundo estudos realizados por Zarrabeitia et al. (2002), mesmo otimizando as condições da reação de PCR no sistema multiplex, alguns marcadores possuem uma melhor eficiência do que outros e o DXS101 parece ser mais sensível ao descongelamento dos primers. Em outro estudo, Shin et al. (2004), também teve dificuldade com a amplificação simultânea deste marcador e do marcador HumARA.

Os produtos da PCR para o sistema Multiplex (DXS6854, DXS7424, DXS6808 e DXS7132) e monoplex (DXS101) podem ser visualizados na Figura 5, a seguir.

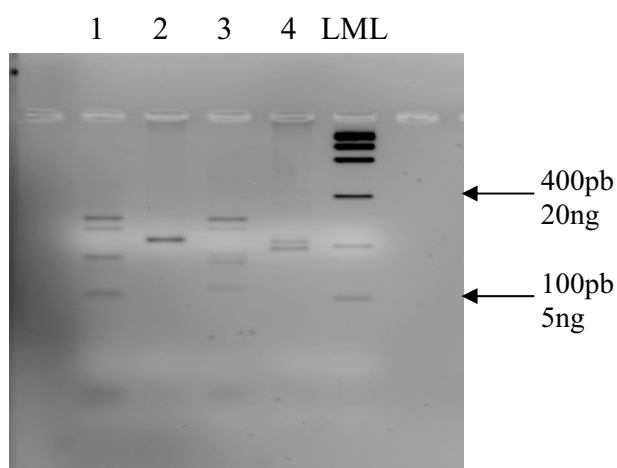


Figura 5: Gel de agarose 4% em TAE [1x], corado com BrEt. Produto da reação de amplificação dos STRs DXS6854, DXS7424, DXS6808 e DXS7132 pelo método Multiplex e do STR DXS101 pelo método Monoplex. Canaleta 1 e 3: Produto da reação de PCR multiplex em uma amostra masculina e feminina (homozigota para 3 marcadores), respectivamente. Canaleta 2 e 4: Produto da reação de PCR monoplex em uma amostra masculina e feminina (heterozigota para o marcador), respectivamente. LML: *Low DNA Mass Ladder*.

Pela análise da Figura 5, pode-se concluir que a PCR foi realizada com sucesso no sistema multiplex e monoplex, uma vez que as bandas obtidas apresentam aproximadamente 100-300 pb (valor correspondente ao tamanho esperado para os STRs) e não há a presença de bandas inespecíficas.

Após a reação de PCR, todas as amostras foram submetidas à corrida eletroforética no Sequenciador Automático ALF Express (GE Healthcare); a determinação alélica foi realizada com o auxílio do programa *Allelelocator* v. 1.03 (GE Healthcare), utilizando-se a escada alélica e o controle K562, cujos eletroferogramas estão apresentados na Figura 6.

A amostra K562, além de funcionar como controle da reação de amplificação, foi utilizada para certificar o correto alinhamento da escada alélica e, na ausência do Ladder, é utilizada diretamente para a designação alélica.

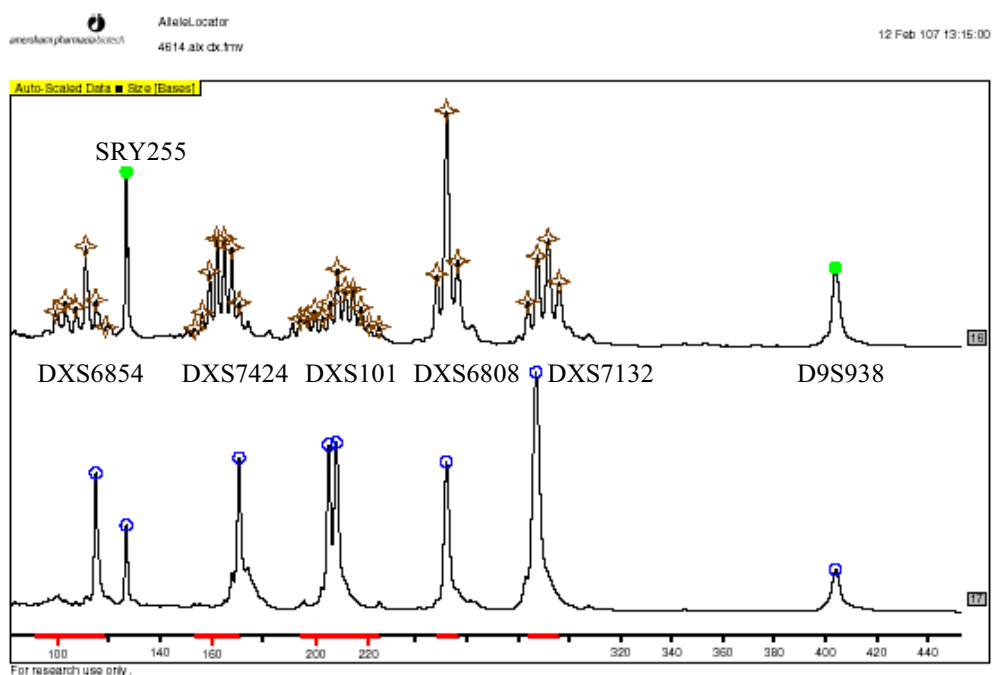


Figura 6: Eletroferograma da escada alélica (Perfil 16) e do controle K562 (Perfil 17) para os 5 X-STRs, incluindo os padrões internos de corrida (SRY255 e D9S938).

O eletroferograma das amostras em que os 5 X-STRs foram amplificados pelo método Monoplex está apresentado na Figura 7 e o eletroferograma das amostras nas quais uma parte dos STRs foi amplificada pelo método Multiplex (DXS6854, DXS7424, DXS6808 e DXS7132) e, a outra parte, pelo Monoplex (DXS101), está demonstrado na Figura 8.

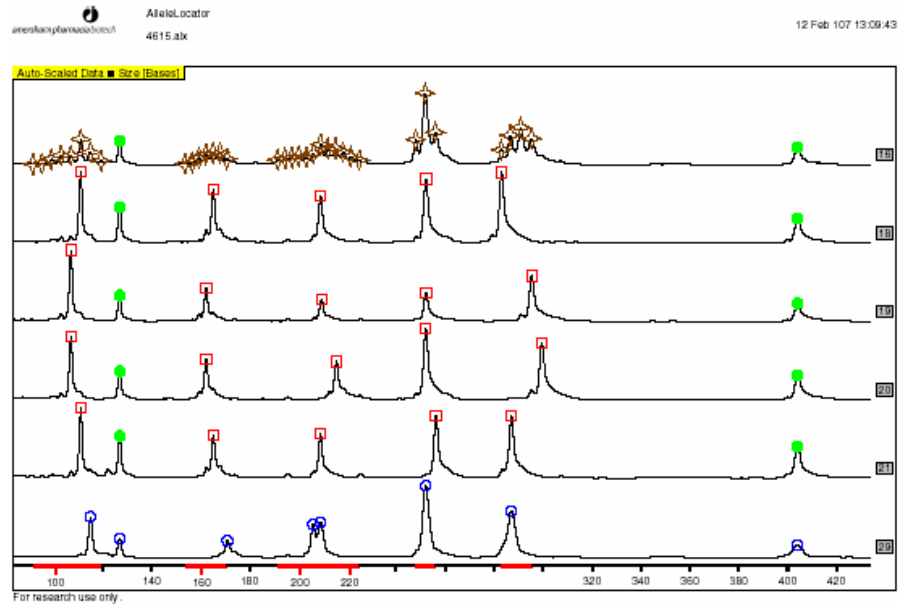


Figura 7: Eletroferograma das amostras nas quais os 5 X-STRs foram amplificados pelo método Monoplex. Perfil 16: escada alélica, Perfis 18 a 21: amostras de indivíduos masculinos, Perfil 20: controle K562.

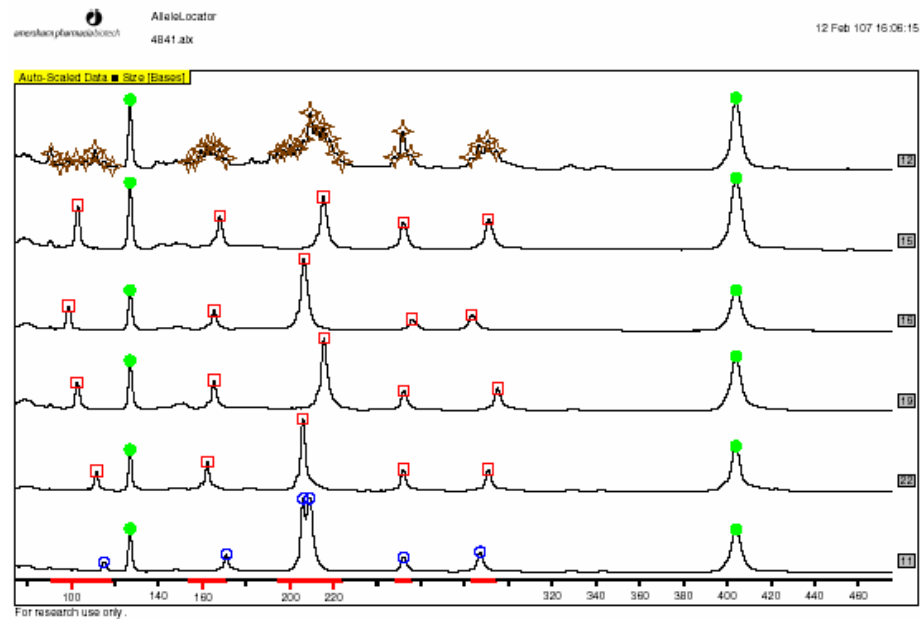


Figura 8: Eletroferograma das amostras nas quais os 4 STRs DXS6854, DXS7424, DXS6808 e DXS7132 foram amplificados pelo método Multiplex e, o DXS101, pelo Monoplex. Perfil 12: escada alélica, Perfis 15, 16, 19 e 22: amostras de indivíduos masculinos, Perfil 11: controle K562.

Analisando-se as figuras de 6 a 8, observa-se uma amplificação de extrema qualidade, uma vez que os picos estão bem definidos, com uma linha de base muito baixa e sem a presença de artefatos (*stutter*, *shoulder peak* e *split peak*<sup>1</sup>), demonstrando a eficácia de ambos os protocolos de amplificação, bem como o de extração. Embora ambos os sistemas (Multiplex e Monoplex) tenham apresentado resultados extremamente satisfatórios, uma melhor amplificação é obtida quando os STRs são amplificados separadamente (Figura 7), uma vez que há maior disponibilidade de nucleotídeos e maior facilidade de contato entre os componentes da PCR e a região de interesse do DNA.

### 5.3 Dado Genético Populacional

Após determinar as frequências alélicas dos 5 marcadores, em homens e mulheres, de acordo com os grupos classificados pela cor da pele (branco, preto, pardo e amarelo), foi realizado o “exact test” para verificar a diferenciação das populações masculina e feminina dentro cada grupo.

Em todos os grupos, não houve diferença significativa ( $p \geq 0,05$ ) entre os gêneros, permitindo a elaboração de tabelas alélicas com os dados femininos e masculinos combinados. Este resultado já era esperado, uma vez que quando o pai passa seu perfil haplotípico (em bloco) para a filha, estes alelos paternos herdados são recombinados com os alelos herdados pela mãe e repassados aleatoriamente aos descendentes. Vários estudos também não encontraram diferenças significativas entre os gêneros masculinos e femininos, como Robino et al. (2006) em sua pesquisa com a população italiana, Zarrabeitia et al. (2006)

---

<sup>1</sup> *Stutter*: pico de DNA com uma unidade de repetição a mais ou a menos que o verdadeiro pico de DNA, obtido pelo “deslizamento” (*slippage*) da DNA polimerase durante a amplificação.

*Shoulder peak* e *split peak*: picos de DNA com um par de base a mais que o verdadeiro pico de DNA, obtidos por uma incompleta adenilação do DNA durante a amplificação.

em seu trabalho com cinco regiões da Espanha e muitos outros autores (POETSCH et al., 2005; SHIN SH et al., 2005; EDELMANN et al., 2002; HERING et al., 2006).

Uma nova comparação pelo “Exact Test” foi efetuada para avaliar diferenças na distribuição alélica entre os grupos branco, preto, pardo e amarelo. Para este cálculo, utilizaram-se os dados de frequências alélicas obtidas para a amostra total (feminina e masculina) de cada grupo. No caso do marcador DXS7132, nenhuma diferença significativa foi obtida entre os grupos; para o DXS6854, o grupo amarelo se diferenciou ( $p \leq 0,05$ ) de todos os outros; para o DXS7424 e DXS101, houve diferença significativa ( $p \leq 0,05$ ) entre os grupos amarelo e preto e, para o DXS6808, entre brancos e pardos.

Os dados apresentados anteriormente sugerem que há algumas diferenças entre os grupos estudados, principalmente em relação aos amarelos. No entanto, tal fato deve ser considerado com cuidado, visto que o número amostral estudado não foi elevado e a classificação realizada levou em conta apenas as características fenotípicas, as quais não representam a origem ancestral do indivíduo.

Segundo os dados do Instituto Brasileiro de Geografia e Estatística (IBGE; Kodama K, 2000), a população asiática, marcada por uma grande diferença física e cultural, sofreu muita resistência para entrar no Brasil. Assim, a imigração iniciou-se somente em 1908, intensificando-se na primeira década do século XX, quando o governo norte-americano proibiu a entrada destes em seu país. Os asiáticos migraram principalmente para São Paulo e suas diferenças étnicas favoreceram os casamentos intra-raça e, conseqüentemente, uma menor miscigenação, o que pode ser um dos motivos pelo qual este grupo apresenta uma maior diferença em relação aos demais.

Asamura et al. (2006a e 2006b) compararam populações asiáticas entre si e não observaram diferenças significativas para diversos marcadores. No entanto, diferenças significativas de distribuição alélica foram encontradas entre a população asiática e a

germânica para os marcadores DXS101, HPRTB, DXS6789 e DXS981 (Tabbada et al., 2006); DXS101 (SHIN KJ et al., 2004); DXS9898, DXS6809 e DXS7424 (LEE et al., 2004); DXS6807, DXS8378, DXS9902, DXS6800, DXS9898, DXS101, HPRTB, DXS7424, DXS7423 (SHIN SH et al., 2005). Outras diferenças também foram encontradas entre a população asiática e a população da Itália (TURRINA et al., 2007) e da Espanha (SHIN SH et al., 2005).

Na população de Minas Gerais, após análise de marcadores informativos de ancestralidade em indivíduos classificados como brancos, pretos e pardos, foi demonstrado que existe uma alta variabilidade do índice de ancestralidade africana (IAA) dentro de cada uma destes grupos, no entanto, pouca variação entre eles, tendo sido encontrado o maior valor para um indivíduo pardo e não negro, enquanto um dos menores valores foi encontrado em um negro. Tais dados demonstraram que a cor, determinada pela avaliação física, é um pobre preditor de ancestralidade africana (PARRA et al., 2003). Em outro estudo, 12 STR autossômicos foram analisados em indivíduos de São Paulo, classificados segundo a cor de pele em brancos, pretos e pardos. Estes marcadores apresentaram uma boa capacidade de discriminação entre europeus e africanos, porém, na população brasileira, não foi possível estabelecer uma relação entre a cor de pele e tais marcadores, sugerindo que não há diferenças significativas entre estes três grupos de cor de pele (PIMENTA et al., 2006).

Além da grande miscigenação da população brasileira, trabalhos com a avaliação filogeográfica do cromossomo X ainda são poucos e muito recentes e nenhuma pesquisa, até o momento, comparou a população brasileira para os seus quatro grupos de cor de pele. Assim, a elaboração de uma base de dados segundo as características fenotípicas dos indivíduos é extremamente complicada e, portanto, as frequências alélicas (Tabela 5) e os parâmetros estatísticos (Tabela 9) para os cinco X-STRs na população de Araraquara, foram calculados considerando-se a população total.

Tabela 5: Frequência alélica dos cinco X-STRs na população de Araraquara-SP.

Alelo	DXS6854	DXS7424	DXS101	DXS6808	DXS7132
7					
8	0,0056				
9	0,1111			0,1167	
10	0,2222	0,0111		0,7278	0,0056
11	0,1333	0,0278		0,1500	0,0056
12	0,3722	0,0167			0,0944
13	0,0944	0,1111			0,2833
14	0,0611	0,2111		0,0056	0,3333
15		0,3000			0,2556
16		0,2389	0,0111		0,0222
17		0,0667	0,0056		
18		0,0167	0,0167		
19			0,0333		
20			0,0389		
21			0,0667		
22			0,0444		
23			0,0500		
24			0,1222		
25			0,1111		
26			0,2667		
27			0,1111		
28			0,0722		
29			0,0222		
30			0,0222		
31			0,0056		

As frequências alélicas apresentadas acima foram comparadas às obtidas por outras populações. É importante ressaltar que em tal análise foi considerada apenas a frequência alélica da população total de Araraquara, não havendo a subdivisão em grupos de cor de pele. No caso da comparação realizada com povos asiáticos, o nosso dado total (compreendendo todos os grupos de cor de pele juntos) foi comparado com diversas populações, subdivididas em coreanas, japonesas e chinesas.

Em relação ao DXS6854, os alelos variaram do 8 ao 14, sendo que o 12 foi o mais freqüente (37%), seguido pelo 10, 11 e 9. Tal distribuição se diferiu um pouco da população chinesa (QUIULIN et al., 2002), na qual os alelos mais freqüentes foram o 10 (42%), 9(28%), 11 (13%) e 12 (11%). Além disso, o alelo 8, o menos freqüente (0,6%) na população brasileira estudada, não foi encontrado na população chinesa (QUIULIN et al., 2002) e os alelos 7 e 15, apresentados em 0,26% dos indivíduos desta, estão ausentes neste trabalho.

Para o DXS7424, os alelos variaram do 10 ao 18, sendo que o alelo 15 (30%) foi o mais freqüente e, o alelo 10 (1%), o menos. Com exceção da China (YU et al., 2005), cujos alelos mais freqüentes foram 17 (36%), 16 (35%) e 15 (13%), as demais populações estudadas apresentaram os alelos 14, 15 e 16. Como neste trabalho, na população italiana (TURRINA et al., 2007), em ordem decrescente de freqüência, estão os alelos 15, 16 e 14; porém, na população germânica (Poetsch et al., 2005), coreana (LEE et al., 2004) e japonesa (ASAMURA et al., 2006b), tem-se 16, 15 e 14 (Tabela 6).

Ainda para este locus, o alelo 10 não foi encontrado nas três populações asiáticas e, o alelo 11, não esteve presente nos chineses e coreanos. Por outro lado, os alelos 19 e 20 apresentados por germânicos, abaixo de 0,03%, não foram tipados em nossa e nem nas demais populações.

A análise estatística dessas populações mostrou que não há diferenças significativas ( $p \geq 0,01$ ) entre a população estudada neste trabalho e a população italiana ( $R_{ST} = 0,0037$ ) e germânica ( $R_{ST} = 0,0066$ ), mas há diferenças em relação às populações asiáticas (Koréia,  $R_{ST} = 0,0176$ ; Japão,  $R_{ST} = 0,0245$ ; China,  $R_{ST} = 0,0928$ ).

Tabela 6: Análise comparativa entre a população de Araraquara (n = 120) e outras populações para o marcador DXS7424.

	ALELOS NÃO ENCONTRADOS		ALELOS MAIS FREQUENTES	
	Literatura	Nosso estudo	Literatura	Nosso estudo
<b>Germânia (n = 205)</b>				
Poetsch et al., 2005	-	19 - 20	16 > 15 > 14	
<b>Itália (n = 200)</b>				
Turrina et al., 2007	-	-	15 > 16 > 14	
<b>Asiáticos:</b>				
<b>Koréia (n = 300)</b>				15 > 16 > 14
Lee et al., 2004	10 - 11	-	16 > 15 > 14	
<b>China (n = 204)</b>				
Yu et al., 2005	10 - 11	-	17 > 16 > 15	
<b>Japão (n = 333)</b>				
Asamura et al., 2006b	10	-	16 > 15 > 14	

No caso do DXS101 (Tabela 7), os alelos 14 e 15 não foram descritos nesse trabalho, nem na população africana (Gomes et al., 2007b) e asiática (ASAMURA et al., 2006b; YU et al., 2005). No entanto, o alelo 14 foi relatado com frequência aproximada de 0,3% em Portugal (PEREIRA et al., 2007), Itália (BINI et al., 2005) e Germânia (POETSCH et al., 2005). Já o alelo 15, apresentou-se de 0,02 a 0,05% na população portuguesa, italiana e germânica (citadas acima) e espanhola (ÁLER et al., 2007).

Um dado interessante neste marcador é que o alelo 31 (0,6%) foi relatado apenas em germânicos (WIEGAND et al., 2003) e os alelos 16, 17, 18, 19, 20, 30 (0,6% a 0,4%) não foram encontrados nos asiáticos (ASAMURA et al., 2006b; YU et al., 2005; TABBADA et al., 2005).

Em geral, nas populações mencionadas anteriormente, os alelos mais frequentes foram o 24, 25 e 26, com uma pequena diferença em relação à população africana de Angola (GOMES et al., 2007b) na qual os alelos 26, 27 e 21 foram os mais comuns.

Para este marcador, a análise estatística mostrou que há diferenças significativas ( $p \leq 0,01$ ), com exceção da África, entre a população estudada neste trabalho e as demais populações. Em ordem decrescente de proximidade genética está a população africana ( $R_{ST} = 0,0043$ ), portuguesa ( $R_{ST} = 0,0116$ ), germânica ( $R_{ST} = 0,0127$ ), italiana ( $R_{ST} = 0,0184$ ), espanhola ( $R_{ST} = 0,0233$ ), Japonesa ( $R_{ST} = 0,0351$ ) e Chinesa ( $R_{ST} = 0,045$ ).

Tabela 7: Análise comparativa entre a população de Araraquara (n = 120) e outras populações para o marcador DXS101.

	ALELOS NÃO ENCONTRADOS		ALELOS MAIS FREQUENTES	
	Literatura	Nosso estudo	Literatura	Nosso estudo
<b>Italiana (n = 556)</b>				
Bini et al., 2005	31	14 - 15	24 > 25 > 26	
<b>Germânica (n = 205)</b>				
Poetsch et al., 2005	31	14 - 15	24 > 25 > 26	
<b>Portuguesa (n = 347)</b>				
Pereira et al., 2007	31	14 - 15	24 > 25 > 26 > 27	
<b>Espanhola (n = 145)</b>				
Áler et al., 2007	31	15	25 > 24 > 26	26 > 24 > 25=27
<b>Africana (n = 237)</b>				
Gomes et al., (2007b)	16 - 17 - 31	-	26 > 27 > 21	
<b>Asiáticos :</b>				
<b>Chinesa (n = 204)</b>				
YU et al, 2005	16 - 18 - 19 29 - 30 - 31	-	24 > 23 > 22	
<b>Japonesa (n = 333)</b>				
Asamura et al., 2006b	16 - 17 - 18 19 - 20 - 31	-	24 > 25 > 26	

O DXS6808 apresentou-se muito pouco polimórfico, com apenas 4 alelos encontrados, sendo que o alelo 10 ocorreu em 72% da população. Nenhuma comparação foi realizada devido à ausência de dados na literatura.

Para o DXS7132, os alelos 10 e 17 foram os que mais se diferenciaram entre as populações analisadas. O alelo 10, presente em 0,6% da população brasileira estudada, foi

relatado apenas em coreanos (Shin SH et al., 2005) e americanos (GOMES et al., 2007a). No entanto, o alelo 17 não foi encontrado em nossos dados, mas ocorreu entre 0,3 e 3% na população espanhola (ÁLER et al., 2007), portuguesa (PEREIRA et al., 2007), Asiática (SHIN SH et al., 2005; YU et al., 2005; ASAMURA et al., 2006a), americana (GOMES et al., 2007a) e brasileira de Santa Catarina (CAINÉ et al., 2007). Em todas as populações citadas anteriormente, os alelos mais freqüentes foram o 13, 14 e 15, perfazendo de 72 % (GOMES et al., 2007a) a 87 % (Araraquara) dos indivíduos (Tabela 8).

A análise estatística dessas populações mostrou que não há diferenças significativas ( $p \geq 0,01$ ), com exceção do Japão ( $p = 0,009$ ), entre a população estudada neste trabalho e as demais populações, sendo que uma maior proximidade genética foi obtida com a população africana de Angola ( $R_{ST} = -0,0046$ ), portuguesa ( $R_{ST} = -0,0016$ ), brasileira de Santa Catarina ( $R_{ST} = -0,0014$ ) e espanhola ( $R_{ST} = -0,00092$ ), estando um pouco mais distantes as populações asiáticas (Koréia,  $R_{ST} = 0,00015$ ; China,  $R_{ST} = 0,0046$ ; Japão,  $R_{ST} = 0,007$ ).

Tabela 8: Análise comparativa entre a população de Araraquara (n = 120) e outras populações para o marcador DXS7132.

	ALELOS NÃO ENCONTRADOS		ALELOS MAIS FREQUENTES	
	Literatura	Nosso estudo	Literatura	Nosso estudo
<b>Brasileira (n = 184)</b>				
Cainé et al., 2007	10	17	14 > 13 > 15	
<b>Americana (n = 377)</b>				
Gomes et al., 2007a	-	8 - 17	14 > 15 > 13	
<b>Portuguesa (n = 347)</b>				
Pereira et al., 2007	10	17	14 > 13 > 15	
<b>Espanhola (n = 145)</b>				
Áler et al., 2007	10	9 - 17	14 > 13 > 15	
<b>Africana (n = 237)</b>				
Gomes et al., 2007b	10	9	14 > 13 = 15	14 > 13 > 15
<b>Asiáticos:</b>				
<b>Chinesa (n = 204)</b>				
YU et al., 2005	10 - 11	17	14 > 15 > 13	
<b>Koreana (n = 401)</b>				
Shin et al., 2005	-	17 - 18	13 > 14 > 15	
<b>Japonesa (n = 401)</b>				
Asamura et al., 2006a	10	17 - 18	14 > 15 > 13	

Os dados apresentados anteriormente demonstram que a população brasileira estudada apresenta algumas particularidades na sua distribuição alélica, reforçando a necessidade em se desenvolver um banco de dados brasileiro. Uma dificuldade encontrada foi a diferença de marcadores X-STRs analisados na literatura, o que não permitiu que as mesmas populações fossem comparadas entre si para cada marcador analisado neste trabalho e ocasionou a comparação com um número limitado de populações. No entanto, pode-se notar que em todas as análises, a população mais próxima da estudada neste trabalho foi a africana, seguida pelas européias e, por fim, as asiáticas. A maior distância em relação às populações asiáticas também foi observado por Góis et al. (2008) e, como já relatado anteriormente, as diferenças físicas e culturais da população asiática pode ter favorecido tal fato.

Segundo os dados históricos, a população brasileira é tri-híbrida, caracterizada pela contribuição europeia, africana e ameríndia. No período de 1500 a 1808, praticamente, os únicos imigrantes europeus que vieram para o Brasil foram os portugueses. No entanto, de 1882 a 1930, maior período de imigração no Brasil, São Paulo recebeu 2.223.000 imigrantes, sendo principalmente portugueses, italianos e espanhóis, seguidos em menor escala por germânicos, sírios, japoneses, e outros [RIBEIRO, 1995; LOVE, 1982]. Em relação à população Africana, segundo o IBGE, uma grande imigração negra ocorreu no período de 1531 a 1853 e mais de 4 milhões de africanos desembarcaram no Brasil, sendo que a África Centro-Occidental (hoje ocupada por Angola) forneceu a maior parte dos escravos utilizados na colonização portuguesa.

Diante do exposto, os dados obtidos neste trabalho estão coerentes com a história de formação da população brasileira.

Prévios estudos com a população brasileira mostram que a linhagem do cromossomo Y tem uma maior contribuição europeia e, a linhagem mitocondrial, apresenta similar contribuição europeia, africana e ameríndia (CARVALHO-SILVA et al., 2001; ALVES-SILVA et al., 2000). Um estudo realizado com marcadores autossômicos na população de São Paulo mostrou uma maior contribuição europeia (79%), seguida por africana (14%) e ameríndia (7%) (FERREIRA et al., 2006).

A grande contribuição europeia e africana na população brasileira é bastante conhecida; no entanto, o fato de a população africana estar mais próxima que as europeias em relação à população estudada neste trabalho não contradiz os trabalhos descritos acima, pois nestes, apenas indivíduos brancos foram analisados. De qualquer forma, tal fato deve ser avaliado com cuidado, pois apenas uma população africana (Angola) foi testada.

Em relação aos parâmetros estatísticos de interesse forense, estes foram determinados e estão apresentados na Tabela 9, a seguir.

Tabela 9: Parâmetros estatísticos dos cinco X-STRs na população de Araraquara-SP.

Parametros	DXS6854	DXS7424	DXS101	DXS6808	DXS7132
Het obs	0,817	0,783	0,817	0,417	0,783
Het esp	0,799	0,804	0,864	0,429	0,747
HWE	0,832	0,650	0,002	0,590	0,818
MEC <sub>T</sub>	0,738	0,760	0,860	0,395	0,690
MEC <sub>D</sub>	0,606	0,632	0,767	0,259	0,550
PD <sub>F</sub>	0,916	0,925	0,972	0,641	0,884
PD <sub>M</sub>	0,769	0,790	0,871	0,434	0,737

Het obs.: heterozigose observada; Het esp.: heterozigose esperada; HWE: Hardy-Weinberg equilíbrio; MEC<sub>T</sub>: chance de exclusão significativa em trios envolvendo filha; MEC<sub>D</sub>: chance de exclusão significativa em duo pai/filha; PD<sub>F</sub>: poder de discriminação feminino; PD<sub>M</sub>: poder de discriminação masculino.

Segundo Szibor et al., (2003), uma das propriedades que devem ser avaliadas em um marcador, para sua posterior aplicação na prática forense, é o equilíbrio de Hardy-Weinberg. Se uma população está em equilíbrio de Hardy-Weinberg, presume-se que está é infinita, que não há eventos de mutação e seleção, que os cruzamentos são aleatórios e que o fluxo gênico existente não é capaz de alterar a composição alélica desta população. Assim, através das frequências alélicas obtidas, é possível determinar a proporção dos diferentes genótipos da população (BEIGUELMAN, 1996).

A maioria dos estudos realizados demonstra uma distribuição genotípica de acordo com o equilíbrio de Hardy-Weinberg, o que também foi observado em nosso estudo, com

exceção do DXS101, para o qual foi obtido um desvio significativo ( $p < 0,05$ ). A taxa de mutação para tal STR tem sido relatada como zero (SZIBOR et al., 2003; ELDEMANN et al., 2001) e apenas esses dados não são suficientes para se dizer que uma possível seleção possa estar ocorrendo. Como tal marcador apresenta uma grande quantidade de alelos, é mais provável que o tamanho amostral utilizado não tenha sido capaz de representar efetivamente a população estudada, constituindo-se um artefato. Por outro lado, tal desvio, juntamente com o excesso de 5% de homozigose, não permite a exclusão da possibilidade de alelos nulos. Tal resultado foi observado, também, em outras populações (BINI et al., 2005; ÁLER et al., 2007).

Em relação aos demais parâmetros estatísticos calculados, o Poder de Discriminação corresponde à probabilidade de se escolher dois indivíduos ao acaso na população e estes apresentarem genótipos diferentes entre si. Neste trabalho, ao avaliar a eficiência do sistema como um todo, o poder de discriminação obtido foi de 0,9999928 e 0,9990702 em mulheres e homens, respectivamente. Assim, pode-se deduzir que a capacidade de duas mulheres e dois homens serem semelhantes entre si para este conjunto de STRs é de 0,0000072 e 0,0009298, respectivamente, o que se caracteriza por um ótimo resultado, uma vez que apenas 5 marcadores foram analisados.

Os valores do poder de discriminação são menores em homens pelo fato destes apresentarem apenas um cromossomo X. Para o DXS7424, nossos dados foram comparados com o da população germânica (ELDEMANN et al., 2002; POETSCH et al., 2005), coreana (LEE et al., 2004), chinesa (LIU et al., 2006), japonesa (ASAMURA et al., 2006b) e italiana (TURRINA et al., 2007) e os valores de  $PD_F$  e  $PD_M$  variaram de 0,849-0,928 e 0,720-0,819, respectivamente, sendo que a Coreia apresentou o menor índice e a população brasileira estudada ficou entre os melhores resultados.

Em relação ao DXS7132, foram analisadas as populações portuguesa (PEREIRA et al., 2007), espanhola (ZARRABEITIA et al., 2006 e 2007; ÁLER et al., 2007); chinesa (LIU et al., 2006), italiana (ROBINO et al., 2006), americana (GOMES et al., 2007a), japonesa (ASAMURA et al., 2006a) e brasileira de Santa Catarina (CAINÉ et al., 2007). Os  $PD_F$  e  $PD_M$  variaram de 0,876-0,906 e 0,737-0,891, respectivamente, e a população brasileira deste trabalho obteve um dos menores resultados, devido à grande frequência dos alelos 13, 14 e 15, a qual fez 87% nesta população e cerca de 70-80% nas que obtiveram melhores índices. Em relação à população brasileira de Santa Catarina, está apresentou melhores resultados para o  $PDF$  (0,905) e  $PDM$  (0,891) em relação à Araraquara.

No caso do DXS101, este é um marcador altamente polimórfico e por isso uns dos mais estudados atualmente. Vários grupos foram comparados, compreendendo os italianos (BINI et al., 2005; ROBINO et al., 2006), espanhóis (ZARRABEITIA et al., 2002; ÁLER et al., 2007), coreanos (SHIN KJ et al., 2004; SHIN SH et al., 2005), japoneses (ASAMURA et al., 2006b), germânicos (POETSCH et al., 2005) e africanos (GOMES et al., 2007b). Os  $PD_F$  e  $PD_M$  variaram de 0,927-0,982 e 0,798-0,942, respectivamente e os dados obtidos neste trabalho mostraram-se intermediários entre os grupos.

Considerando-se uma situação de paternidade com o trio (mãe, filha e suposto pai), o  $MEC_T$  indica a probabilidade de um indivíduo qualquer da população e não relacionado à criança testada ser excluído da paternidade. O mesmo é válido para o  $MEC_D$ , porém nesta situação tem-se apenas pai e filha. As mesmas comparações realizadas para o poder de discriminação foram realizadas para os MECs e, no caso do DXS7424, os  $MEC_T$  e  $MEC_D$  variaram, respectivamente, de 0,639-0,760 e 0,492-0,652; para o DXS7132, de 0,689-0,723 e 0,549-0,598 e, para o DXS101, de 0,760-0,891 e 0,633-0,812.

Como no  $MEC_D$  trabalha-se na ausência da mãe, não é possível determinar quais alelos esta doou para a filha, portanto, o pai da criança poderá ser qualquer homem da

população que tenha o alelo A ou B presentes nesta. Assim, para os loci em que a filha é heterozigota, aumentam-se os pais possíveis e, portanto, diminui-se o MEC. Como tais parâmetros estão diretamente relacionados com a distribuição alélica dos STRs, o mesmo comportamento observado para o poder de discriminação foi obtido com os MECs.

Desta forma, através de uma análise relativa, podemos concluir que o DXS7424, ao contrário do DXS7132, mostrou-se um marcador melhor na população brasileira estudada do que nas demais analisadas, sendo que o DXS101 foi intermediário.

Em relação ao DXS6808, não foi realizada nenhuma comparação por não haver literatura disponível e, para o DXS6854, os dados foram avaliados com aqueles publicados por Quiulin et al. (2002), demonstrando que a população brasileira estudada apresentou melhores resultados.

Ainda em relação aos parâmetros determinados, em nossa população, observa-se que o marcador mais informativo e polimórfico foi o DXS101, seguido pelo DXS7424, DXS6854, DXS7132 e DXS6808, sendo que este último apresentou valores muito inferiores em relação aos demais marcadores.

Diante disto, foram calculados os MECs e PDs combinados na ausência e presença do marcador DXS6808 e os dados obtidos estão representados graficamente (Figura 9).

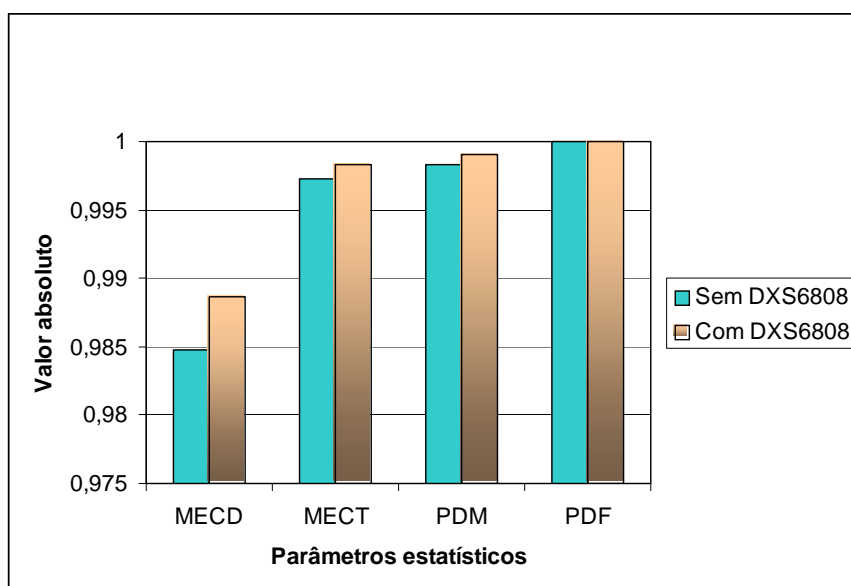


Figura 9: Contribuição do DXS6808 para a identificação humana e o exame de paternidade na população de Araraquara-SP. MECD: Chance de Exclusão Significativa para marcadores do cromossomo X na ausência da mãe (dupla pai-filha); MECT: Chance de Exclusão Significativa para marcadores do cromossomo X em trios envolvendo filhas; PDM: Poder de Discriminação Masculino; PDF: Poder de Discriminação Feminino.

Embora o marcador DXS6808 seja menos informativo que os demais, segundo o gráfico, pode-se concluir que este contribui para uma melhor capacidade de identificação humana, sendo tal contribuição pequena para os  $MEC_T$ ,  $PD_M$  e  $PD_F$  e, moderada, para o  $MEC_D$ . O fato deste último parâmetro ter sido mais favorecido com o DXS6808 ocorre em virtude de que na ausência da mãe a quantidade de pais possíveis aumenta de tal forma que ao adicionar um marcador STR, por mais que este não seja muito polimórfico, a pequena quantidade de homens que ele consegue excluir da paternidade já é significativa.

Em relação ao sexo masculino, foram analisados 60 indivíduos e todos os haplótipos obtidos foram únicos (Apêndice 8.2).

A diversidade haplotípica, um parâmetro estatístico que representa a probabilidade de se selecionar dois indivíduos ao acaso na população e estes apresentarem haplótipos diferentes para um determinado conjunto de marcadores, foi calculada na população em questão, obtendo-se o valor de 1,0. Assim, considerando os dados obtidos como representativos da população brasileira de Araraquara, podemos inferir que a probabilidade de dois indivíduos desta população, do sexo masculino e não aparentados, apresentarem o mesmo haplótipo é praticamente zero, demonstrando a alta capacidade discriminativa de tal sistema.

Uma vez que todos os X-STRs analisados encontram-se no braço longo do cromossomo X, foi avaliado o desequilíbrio de ligação entre estes. Embora os marcadores DXS7424-DXS101 e DXS6854-DXS6808 encontrem-se muito próximos no cromossomo X, não foi observada a ligação entre eles. Tal dado já foi relatado em outro estudo por Lee et al. (2004) e pode ser explicado pelo fato de que mesmo os STRs estando muito próximos, a pequena taxa de recombinação existente entre estes, ao longo das gerações, poderá promover uma variabilidade genética, reestabelecendo-se o equilíbrio de ligação. Além disso, o desequilíbrio de ligação não é decorrente apenas da distância entre os STRs, podendo ser ocasionado por vários outros fatores como: seleção, mutação, deriva aleatória ou não aleatória, miscigenação ou subdivisão populacional (CHAKRAVARTI, 1999; ZARRABEITIA et al., 2006)

Por outro lado, um resultado significativo ( $p < 0.05$ ) foi observado entre o DXS101 e o DXS6808; porém, como estes marcadores estão muito distantes um do outro, o desequilíbrio de ligação é improvável e tal dado pode ser melhor atribuído ao tamanho da amostra, a qual pode não ter sido capaz de representar todos os alelos possíveis do DXS101. A associação deste X-STR com outros marcadores distantes também foi relatada em asiáticos americanos (GOMES et al., 2007a).

“O prazer no  
trabalho  
aperfeiçoa a  
obra”

Aristóteles

## **6. CONCLUSÕES**

Os dados apresentados neste trabalho permitiram concluir que:

- O marcador DXS101 foi o mais polimórfico, seguido por DXS7424, DXS6854, DXS7132 e DXS6808;
  
- Embora o marcador DXS6808 seja pouco informativo, sua aplicação juntamente com os demais STRs, permitiu uma melhor discriminação, constituindo tal sistema em uma importante ferramenta para a identificação humana e testes de paternidade na população brasileira de Araraquara-SP.

“Aprender é a  
única coisa de que  
a mente nunca se  
cansa, nunca tem  
medo e nunca se  
arrepende”

Leonardo da Vinci

## **7. REFERÊNCIAS BIBLIOGRÁFICAS**

1. ALER, M.; SÁNCHEZ-DIZ P.; GOMES I.; GISBERT M. ; CARRACEDO A. ; AMORIM A.; GUSMÃO L. Genetic data of 10 X-STRs in a Spanish population sample. **Forensic Sci. Int.**, v.173, p.193-196, 2007.
2. ALVES-SILVA, J.; SANTOS, M.S.; GUIMARÃES, P.E.M.; FERREIRA, A.C.S.; BANDELT, H.; PENA, S.D.J.; PRADO, V.F. The ancestry of brazilian mtDNA lineages. **Am. J. Hum. Genet.**, v. 67, p.444-461, 2000.
3. ASAMURA, H.; SAKAI, H.; OTA, M.; FUKUSHIMA, H. Japanese population data for eight X-STR loci using two new quadruplex systems. **Int. J. Legal Med.**, v.120, p.303-309, 2006a.
4. 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, 2006b.
5. BEIGUELMAN, B. **Dinâmica dos genes nas famílias e nas populações**. 2. ed. Ribeirão Preto: Sociedade Brasileira de Genética, 1996.
6. BINI, C.; CECCARDI, S.; FERRI, G.; PELOTTI, S.; ALÙ, M.; ROCANGLIA, E.; BEDUSCHI, G.; CAENAZZO, L.; PONZANO, E.; TASINATO, P.; TURCHI, C.; BUSCEMI, L.; MAZZANTI, M.; TAGLIABRACCI, A.; TONI, C.; SPINETTI, I.; DOMENICI, R.; PRESCIUTTINI, S. Development of a heptaplex PCR system to analyse X-chromosome STR loci from five Italian population samples: a collaborative study. **Forensic Sci. Int.**, v.153, p.231-236, 2005.
7. BUDOWLE, B.; BROWN, B.L. The use of DNA typing for forensic identification. **Forensica**, v.1, p.23-37, 2001.
8. BUTLER, J.M. **Forensic DNA typing**. 2<sup>nd</sup> ed. San Diego: Academic Press, 2005. cap. 1, p.1-13.

9. CAINÉ, L.M.; PONTES, L.; ABRANTES, D.; LIMA, G.; PINHEIRO, F. Genetic Data of four X-Chromosomal STRs in a population sample of Santa Catarina, Brazil. **J. Forensic Sci.**, v.52, p.502-503, 2007.
10. CALABREZ, M.C.T. **Influência do calor na análise de DNA extraído de sangue e tecidos humanos**: importância para a identificação de corpos carbonizados. 1999. 92f. Tese (Doutorado) – Faculdade de Medicina da Universidade de São Paulo, 1999.
11. CARVALHO-SILVA, D.; SANTOS, F.R.; ROCHA, J.; PENA, S.D.J. The phylogeography of Brazilian Y-chromosome lineagens. **Am. J. Hum Genet.**, v.68, p.281-286, 2001.
12. CHAKRAVARTI, A. Population genetics – making sense out of sequence. **Nat. Genet.**, v.21, p.56-60, 1999.
13. CHOW, J.C.; YEN, Z.; ZIESCHE, S.M.; BROWN, C.J. Silencing of the mammalian X chromosome. **Annu. Rev. Genomics Hum. Genet.**, v.6, p.69–92, 2005.
14. 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.
15. EDELMANN, J.; SZIBOR, R. DXS101: a highly polymorphic X-linked STR. **Int. J. Legal Med.**, v.114, p.3101-3104, 2001.
16. EDELMANN, J.; SZIBOR, R. The X-linked STRs DXS7130 and DXS6803. **Forensic Sci Int.**, v.136, p.73-75, 2003.
17. EDELMANN, J.; HERING, S.; KUHLISCH, E.; SZIBOR, R. Validation of the STR DXS7424 and the linkage situation on the X-chromosome. **Forensic Sci. Int.**, v.125, p.217–222, 2002.

18. ELLIS, N.; GOODFELLOW, P.N. The mammalian pseudoautosomal region. **Trends Genet.**, v.5, p.406-410, 1989.
19. EXCOFFIER, L.; LAVAL, G.; SCHNEIDER, S. Arlequin version. 3.0: An integrated software package for population genetics data analysis. **Evolution. Bioinform. Online**, v.1, p. 47-50, 2005.
20. FEDERAL BUREAU OF INVESTIGATION. **CODIS (COmbined DNA Index System)**. Disponível em: <http://www.fbi.gov/hq/lab/codis/index1.htm>. Acesso em 10 set. 2007.
21. FERREIRA, L.B.; MENDES-JUNIOR, C.T.; WIEZEL, C.E.V.; LUIZON, M.R.; SIMÕES AL. Genomic ancestry of a sample population from the State of São Paulo, Brazil. **Am. J. Human Biol.**, v.18, p.702-705, 2006.
22. FRANÇA, G.V. Medicina Legal. 6. ed., Rio de Janeiro, Guanabara-Kogan, 2001. cap. 3, p.32-63.
23. FREIJE, D.; HELMS, C.; WATSON, M.S.; DONIS-KELLER, H. Identification of a second pseudoautosomal region near the Xq and Yq telomeres. **Science**, v.258, p.1784–1787, 1992.
24. FUNDAÇÃO DE AMPARO À PESQUISA DO RIO DE JANEIRO. O DNA da verdade, 2003. Disponível em: [http://www.faperj.br/interna.phtml?obj\\_id=892](http://www.faperj.br/interna.phtml?obj_id=892). Acesso em 10 set. 2007.
25. GAO, S.; QIAO, K.; RAKHA, A.; LAI, J.; LI, S. Allele frequencies for 10 X-STR loci in Nu population of Yunnan, China. **Leg. Med.**, Tokyo, v.9, p.284-286, 2007.
26. GATTÁS, G.F.; FIGARO-GARCIA, C. **Caminho de volta: tecnologia na busca de crianças e adolescentes desaparecidos no Estado de São Paulo**. São Paulo: Secretaria Especial de Direitos Humanos, 2007.

27. GÓIS, C.C.; MARTINS, J.A.; PEREIRA, G.A.; FRESCHI, A.; PANETO, G.G.; ALVARENGA, V.L.S.; CICARELLI, R.M.B.; HIRATA M.H.; OLIVEIRA, R.N. Genetic population data of 12 STR loci of the PowerPlexY system in the State of São Paulo population (Southeast of Brazil). **Forensic Sci. Int.**, v.174, p.80-85, 2008.
28. GOMES, I.; PRINZ, M.; PEREIRA, R.; MEYERS, C.; MIKULASOVICH, R.S.; AMORIM, A.; CARRACEDO, A.; GUSMÃO, L. Genetic analysis of three US population groups using a X-chromosomal STR decaplex. **Int J Legal Med**, v.121, p.198-203, 2007a.
29. 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.: Genetics**, v.1, p.208-211, 2007b.
30. GU, S.Z.; CHEN, T.; LIU, Q.B.; YU, B.; LI, S.B. Allele distribution of five X-chromosome short tandem repeat loci in Ewenke population of north China. **Chin. Med. Sci. J.**, v.20, p.237-241, 2005.
31. GUO, S.W.; THOMPSON, E.A. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. **Biometrics**, v.48, p.361-372, 1992.
32. HAMOSH, A.; SCOTT, A.F.; AMBERGER, J.S.; BOCCHINI, C.A.; MCKUSICK, V.A. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. **Nucleic Acids Res.**, v.33, p.514-517, 2005.
33. HECKER, K.H.; ROUX, K.H. High and low annealing temperatures increase both specificity and yield in touchdown and stepdown PCR. **Biotechniques**, v.20, p.478-85, 1996.
34. HERING, S.; AUGUSTIN, C.; EDELMANN, J.; HEIDEL, M.; DRESSLER, J.; RODIG, H.; KUHLSCH, E.; SZIBOR, R. DXS10079, DXS10074 and DXS10075 are STRs located within a 280-kb region of Xq12 and provide stable haplotypes useful for complex kinship cases. **Int. J. Legal Med.** v.120, p.337-345, 2006.

35. HERING, S.; BRUNDIRS, N.; KUHLISCH, E.; EDELMANN, J.; PLATE, I.; BENECKE, M.; VAN, P.H.; MICHAEL, M.; SZIBOR, R. DXS10011: studies on structure, allele distribution in three populations and genetic linkage to further q-telomeric chromosome X markers. **Int. J. Legal Med.**, v.118, p.313-319, 2004.
36. HILLIER, L.W. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. **Nature**, v.432, p.695-716, 2004.
37. JEFFREYS, A.J.; BROOKFIELD, J.F.Y. Semeonoff R, Positive identification of an immigration test-case using human DNA fingerprints. **Nature**, v.317, p.818-819, 1985.
38. JEFFREYS, A.J.; WILSON, V.; THEIN, S.L. Hipervariable “minisatellite” regions in human DNA. **Nature**, v.314, p.67-73, 1985.
39. JOBIM, L.F.; JOBIM, M.R.; BRENNER, C. Identificação humana pelo DNA: investigação de paternidade e análise de casos forense. In: Tochetto, D., **Identificação humana**. Porto Alegre: Sagra Luzzato, 1999. pte. 4, p.237-303.
40. JORDE, L.B.; WATKINS, W.S.; BAMSHAD, M.J. Population genomics: a bridge from evolutionary history to genetic medicine. **Human Mol. Genetics**, v.10, n.20, p.2199-2207, 2001.
41. KODAMA, K. **O sol nascente do Brasil**: um balanço da imigração japonesa. In: IBGE. Brasil: 500 anos de povoamento. Rio de Janeiro, 2000. Disponível em: <http://www.ibge.gov.br>. Acesso em 15 dez. 2007.
42. LANDER, E.S.; LINTON, L.M.; BIRREN, B.; et al. Initial sequencing and analysis of the human genome. **Nature**, v. 409, p.860–921, 2001.
43. LEE, H.Y.; PARK, M.J.; JEONG, C.K.; LEE, S.Y.; YOO, J.E.; CHOI, C.U.H.; KIM, C.Y.; SHIN, K.J. Genetic characteristics and population study of 4 X-chromosomal STRs in Koreans: evidence for a null allele at DXS9898. **Int J Legal Med.**, v.118, p.355-360, 2004.

44. LI, W.H.; YI, S.; MAKOVA, K. Male-driven evolution. **Curr. Opin. Genet. Dev.**, v.12, p.650–656, 2002.
45. LIU, Q.; LI, S. Patterns of genetic polymorphism at the 10 X-chromosome STR loci in Mongol population. **Forensic Sci. Int.**, v.158, p.76-79, 2006.
46. LOVE J. **A locomotiva: São Paulo na Federação Brasileira 1889-1937**. São Paulo, Paz e Terra, 1982.
47. MOLLER, A.; BRINKMANN, B. PCR-VNTRs (PCR-variable number of tandem repeats) in forensic science. **Cell. Mol. Biol.**, v.41, n.5, p.715-724, 1995.
48. 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.
49. NAGARAJA, R.; MACMILLAN, S.; JONES, C.; MASISI, M.; PENGUE, G.; PORTA, G.; MIAO, S.; CASAMASSINI, A.; D'URSO, M.; BROWNSTEIN, B. Integrated YAC/STS physical and genetic map of 22.5 Mb of human Xq24–q26 at 56-kb inter-STS resolution. **Genomics**, v.52, p.247–266, 1998.
50. NEI, M. **Molecular evolutionary genetics**. New York: Columbia University Press, 1987.
51. OHNO, S. **Sex chromosomes and Sex-linked genes**. Berlin: Springer, 1967.
52. PAGE, D.C.; HARPER, M.E.; LOVE, J.; BOTSTEIN, D. Occurrence of a transposition from the X-chromosome long arm to the Y-chromosome short arm during human evolution. **Nature**, v.311, p.119–123, 1984.
53. PARRA, F.C.; AMADO, R.C.; LAMBERTUCCI, J.R.; ROCHA, J.; ANTUNES, C.M.; PENA, S.D.J. Color and genomic ancestry in Brazilians. **Proc. Natl. Acad. Sci., U.S.A.**, v.100, p.177–182, 2003.
54. PEREIRA, R.; GOMES, I.; AMORIM, A.; GUSMÃO, L. Genetic diversity of 10 X chromosome STRs in northern Portugal. **Int. J. Legal Med.**, v.121, p.192-197, 2007.

55. PIMENTA, J.R.; ZUCCHERATO, L.W.; DEBES, A.A.; MASELLI, L.; SOARES, R.P.; MOURA-NETO, R.S.; ROCHA, J.; BYDLOWSKI, S.P.; PENA, S.D.J. Color and genomic ancestry in Brazilians: a study with forensic microsatellites. **Hum. Hered.**, v.62, p.190-195, 2006.
56. PRETTY, I.A.; SWEET, D. A look at forensic dentistry – part 1: the role of teeth in the determination of human identity. **Br. Dent. J.**, 190-7, p.359-366, 2001.
57. PRIMORAC, D.; SCHANFIELD, M.S. Application of forensic DNA testing in the legal system. **Croatian Medical Journal**, 41-1, p.32-46, 2000.
58. POETSCH, M.; PETERSMANN, H.; REPENNING, A.; LIGNITZ, E. Development of two pentaplex systems with X-chromosomal STR loci and their allele frequencies in a northeast German population. **Forensic Sci. Int.**, v.155, p.71-76, 2005.
59. QUILING, M.U.; DEJIAN, L.U.; HANLING, L.U. Polymorphism of DXS6854 locus in Guandong Han population. **China Acad. J.**, v.17, p.22-23, 2002.
60. REICH, D.E.; LANDER, E.S. On the allelic spectrum of human disease. **TRENDS in Genetics**, v.17, n.9, p.502-510, 2001.
61. RIBEIRO, D. **O povo brasileiro: a formação e o sentido do Brasil**. São Paulo. Companhia das Letras, 1995.
62. RIDLEY M. **Evolução**. 3. ed. Porto Alegre: Artmed, 2006. cap. 2, p.45-65.
63. ROBINO, C.; GIOLITTI, A.; GINO, S.; TORRE, C. Development of two multiplex PCR systems for the analysis of 12 X-chromosomal STR loci in a northwestern Italian population sample. **Int. J. Legal Med.**, v.120, p.315-318, 2006.
64. ROSS, M.T.; GRAFHAN, D.V.; COFFEY, A.J.; SCHERER, S.; MCLAY, K.; MUZNY, D.; et al. The DNA sequence of the human X chromosome. **Nature**, v.434, p.325-327, 2005.

65. SASSOUNI, V. Deontofacial radiography in forensic dentistry. **J. Dent. Res.**, 42-1, p. 274-302, 1963.
66. Sociedade Brasileira de Microscopia e Microanálise, **Programa de Ciência e Tecnologia Aplicada na Segurança Pública**. Disponível em: [http://sbmm.org.br/arquivos/rede\\_polvora/ProgramaMJMCT2.pdf](http://sbmm.org.br/arquivos/rede_polvora/ProgramaMJMCT2.pdf). Acessado em 16 ago. 2007.
67. SCHAFFNER, S. The X chromosome in population genetics. **Nature Rev.**, v.5, p.43-51, 2004.
68. SHIN, S.H.; YU, J.S.; PARK, S.W.; MIN, G.S.; CHUNG, K.W. Genetic analysis of 18 X-linked short tandem repeat markers in Korean population. **Forensic Sci. Int.**, v.147, p.35-41, 2005.
69. SHIN, K.J.; KWON, B.K.; LEE, S.S.; YOO, J.E.; PARK, M.J.; CHUNG, U.; LEE, H.Y.; HAN, G.R.; CHOI, J.H.; KIM, C.Y. Five highly informative X-chromosomal STRs in Koreans. **Int. J. Legal Med.**, v.118, p.37-40, 2004.
70. SILVEIRA, D.; SILVA, F.F.; JESUS, P.R.; WHITTLE, M.R. Use of X-linked short tandem repeat loci in routine parentage casework. **Transfusion**, v.47, p.1050-1053, 2007.
71. SKALETSKY, H.; KURODA-KAWAGUCHI, T.; MINX, P.J.; CORDUM, H.S.; HILLIER, L.; BROWN, L.G.; R, TEPPING, S.; PYNTIKOVA, T.; ALI, J.; BIERI, T.; et al. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. **Nature**, v.423, p.825–837, 2003.
72. SMITH, L.M.; BURGOYNE, L.A.; Collecting, archiving and processing DNA from wildlife samples using FTA® databasing paper, **BMC Ecology**, v.4, 2004.
73. SMITH, B.C.; HOLLAND, M.M.; SWETT, D.L.; DIZINNO, J.A. DNA and the Forensic Odontologist. In: BOWERS, C.M.; BELL, G.L. **Manual of forensic odontology**. 3<sup>rd</sup> ed. Ontario: Manticore Publisher, cap.10, p.283-298, 1997.

74. SUBRAMANIAN, S.; MISHRA, R.K.; SINGH, L. Genome-wide analysis of microsatellite repeats in humans: their abundance and density in specific genomic regions. **Genome Biology**, v.4, R13, 2003.
75. Sweet, D. Why a dentist for identification?. **Dent. Clin. North Am.**, v.45, n.2, p.237-251, 2001.
76. SZIBOR, R.; HERING, S.; ELDEMANN, J. A new Web site compiling forensic chromosome X research is now online. **Int J Legal Med.**, v.120, p.252-254, 2006.
77. SZIBOR, R.; KRAWCZAK, M.; HERING, S.; EDELMANN, J.; KUHLISCH, E.; KRAUSE, D. Use of X-linked markers for forensic purposes. **Int. J. Legal Med.**, v.117, p.67-74, 2003.
78. SZIBOR, R.; HERING, S.; KUHLISCH, E.; PLATE, I.; DEMBERGER, S.; KRAWCZAK, M.; EDELMANN, J. Haplotyping of STR cluster DXS6801– DXS6809– DXS6789 on Xq21 provides a powerful tool for kinship testing. **Int. J. Legal Med.**, v.119, p.363–369, 2005.
79. TABBADA, K.A.; DE UNGRIA, M.C.; FAUSTINO, L.P.; ATHANASIADOU, D.; STRADMANN-BELLINGHAUSEN, B.; SCHNEIDER, P.M. Development of a pentaplex X-chromosomal short tandem repeat typing system and population genetic studies. **Forensic Sci. Int.**, v.154, p.173-180, 2005.
80. TANDE, C.M. DNA Typing: A New Investigatory Tool. **Duke Law Journal**, v.1989, n. 2, p.474-494, 1989.
81. TONI, C.; DOMENICI, R.; PRESCIUTTINI, S.; Genotype probabilities of pairs of individuals for X-chromosome markers. **Transfusion**, v.47, p.1276-1280, 2007.
82. TURRINA, S.; ATZEI, R.; DE LEO, D. Polymorphism of four X-chromosomal STRs: DXS7423, DXS7424, DXS8378 and DXS6809 in a North Italian population sample and their use in kinship testing. **Forensic Sci. Int.**, v.168, p.241-243, 2007.

83. UNIVERSIDADE FEDERAL DE ALAGOAS. **Laboratório de DNA Forense**. Disponível em: <http://www.labdnaforense.org/index2.php?endereco=bd>. Acessado em 16 ago. 2007.
84. WEEDN, V.W.; SWARNEN, S.L. Exames forenses de identificação por análises do DNA. In: Henry, J.B. **Diagnósticos clínicos e tratamento por métodos laboratoriais**. 19. ed. São Paulo:Manole, 1998. cap.63, p.1427-1438.
85. WOOD, R.E. Forensic aspects of maxillofacial radiology. **Forensic Sci. Int.**, v.159, p.47-55, 2006.
86. ZARRABEITIA, M.T.; ALONSO, A.; ZARRABEITIA, A.L.; BLANCO, L.; RIANCHO, J.A. Unlinked tetrameric microsatellites on the X chromosome: frequency data in males from Cantabria (northern Spain). **J Forensic Sci.**, v.50, p.1497-1498, 2005.
87. ZARRABEITIA, M.T.; AMIGO, T.; SAÑUDO, C.; ZARRABEITIA, A.; GONZÁLES-LAMUÑO, D.; RIANCHO, J.A. A new pentaplex system to study short tandem repeat markers of forensic interest on X chromosome. **Forensic Sci. Int.**, v.129, p.85-89, 2002.
88. ZARRABEITIA, M.T.; ALONSO, A.; MARTIN, J.; GONZALEZ-GAY, M.A.; MARTIN-ESCUADERO, J.C.; DE PANCORBO, M.M.; SANZ, P.; RUIZ-CABELLO, F.; RIANCHO, J.A. Study of six X-linked tetranucleotide microsatellites: population data from five Spanish regions. **Int J. Legal Med.**, v.120, p.147-150, 2006.

“ Quando vamos ao teatro,  
nós vemos luzes.  
Precisamos de luzes para  
que possamos enxergar.  
Mas a luz não é a peça, ela  
é simplesmente o que torna  
a peça possível. O genoma  
humano faz exatamente  
isso: ilumina a pesquisa  
biomédica. Ele não a dirige  
nem a controla, mas sim  
precisa que todos estejam  
livres para fazerem seus  
papéis ”

John Sulston

## **8. APÊNDICES**

## 8.1 Soluções e géis utilizados

- Acrilamida 30%

4,08 M Acrilamida (GE Healthcare, Chalfont St. Giles, UK), 0,64 M Bis-acrilamida (GE Healthcare, Chalfont St. Giles, UK) em água.

- Brometo de Etídio [5 mg/mL] – (Gibco-BRL, Gaithersburg, MD)

- DTT 1M (USB, Cleveland, Ohio, USA)

- *Elution Buffer* (Promega, Madison, WI, USA)

10 mM Tris-HCl (pH 8,0), 0,1mM EDTA.

- Gel de Agarose 1%

1% m/v Agarose (BioAgency, São Paulo, SP, BR) em TAE 1X.

- Gel de Agarose 4%

3,56% m/v de Agarose (BioAgency, São Paulo, SP, BR), 0,44% m/v de Agarose - *Low Melting* (USB, Cleveland, Ohio, USA) em TAE 1X.

- Gel de poliacrilamida desnaturante a 6%

3,46 M Uréia (USB, Cleveland, Ohio, USA), 20% v/v Acrilamida 30% (GE Healthcare, Chalfont St. Giles, UK), 10% v/v TBE 10X, 0,07% v/v TEMED (Gibco-BRL, Gaithersburg, MD) e 0,5% v/v APS 10% (Sigma-Aldrich, St. Louis, MO, USA).

- Gold STAR 10X Buffer (Promega, Madison, WI, USA)

500 mM KCl, 100 mM Tris-HCl pH 8,3, 15 mM MgCl<sub>2</sub>, 1% Triton<sup>®</sup> X-100, 2 mM de cada dNTP, 1,6 mg/mL de BSA.

- TAE (Tris Acetato EDTA) 1X

40 mM Tris-base (USB, Cleveland, Ohio, USA), 0,114% v/v de Ácido Acético Glacial (Synth), 1 mM EDTA (Merck, Darmstadt, Germany), pH 8,0.

- Tampão de Amostra I

0,25% m/v Azul de Bromofenol (Sigma-Aldrich, St. Louis, MO, USA), 0,25% m/v Xilenocianol (Sigma-Aldrich, St. Louis, MO, USA) e 30% v/v glicerol (Amresco, Solon, Ohio, USA).

- Tampão de Amostra II

8,33% do produto de PCR do D9S938, 8,33% do produto de PCR do SY258 e 83,34% de *Blue Dextran Loading Solution* (Promega, Madison, WI, USA)

- TBE (Tris Borato EDTA) 10X

20 mM EDTA (Merck, Darmstadt, Germany), 890 mM Tris-base (USB, Cleveland, Ohio, USA), 890 mM Ácido Bórico (USB, Cleveland, Ohio, USA), pH 8,3.

**8.2 Distribuição haplotípica do cromossomo X em indivíduos masculinos da população de Araraquara-SP.**

Haplótipo	DXS6854	DXS7424	DXS101	DXS6808	DXS7132	N
<b>h1</b>	9	13	28	10	12	1
<b>h2</b>	9	14	19	11	13	1
<b>h3</b>	9	14	26	10	12	1
<b>h4</b>	9	15	21	11	14	1
<b>h5</b>	9	16	22	10	15	1
<b>h6</b>	9	16	26	10	14	1
<b>h7</b>	9	17	25	10	15	1
<b>h8</b>	10	14	25	10	13	1
<b>h9</b>	10	14	27	9	13	1
<b>h10</b>	10	15	23	10	13	1
<b>h11</b>	10	15	24	10	15	1
<b>h12</b>	10	17	23	10	14	1
<b>h13</b>	10	16	16	10	13	1
<b>h14</b>	10	16	26	10	13	1
<b>h15</b>	10	16	27	9	13	1
<b>h16</b>	10	15	27	10	13	1
<b>h17</b>	10	17	26	9	14	1
<b>h18</b>	11	11	23	10	14	1
<b>h19</b>	11	13	25	10	13	1
<b>h20</b>	11	14	24	10	15	1
<b>h21</b>	11	14	26	10	16	1
<b>h22</b>	11	15	21	14	15	1
<b>h23</b>	11	15	23	10	15	1
<b>h24</b>	11	16	28	11	14	1
<b>h25</b>	12	10	27	9	14	1
<b>h26</b>	12	13	20	10	13	1
<b>h27</b>	12	13	30	10	14	1
<b>h28</b>	12	13	31	10	14	1
<b>h29</b>	12	14	21	11	15	1
<b>h30</b>	12	14	24	10	13	1
<b>h31</b>	12	14	25	11	13	1
<b>h32</b>	12	14	26	9	12	1
<b>h33</b>	12	14	26	10	16	1
<b>h34</b>	12	14	28	10	12	1
<b>h35</b>	12	15	20	11	14	1
<b>h36</b>	12	15	21	9	15	1

<b>h37</b>	12	15	24	10	12	1
<b>h38</b>	12	15	24	10	13	1
<b>h39</b>	12	15	24	11	13	1
<b>h40</b>	12	15	25	10	13	1
<b>h41</b>	12	15	25	10	14	1
<b>h42</b>	12	15	26	9	15	1
<b>h43</b>	12	15	26	10	13	1
<b>h44</b>	12	15	26	10	14	1
<b>h45</b>	12	15	26	10	15	1
<b>h46</b>	12	15	28	10	13	1
<b>h47</b>	12	15	28	10	14	1
<b>h48</b>	12	15	28	10	15	1
<b>h49</b>	12	15	29	11	13	1
<b>h50</b>	12	16	19	10	14	1
<b>h51</b>	12	16	23	10	15	1
<b>h52</b>	12	16	24	10	15	1
<b>h53</b>	12	16	28	9	15	1
<b>h54</b>	12	17	24	10	15	1
<b>h55</b>	13	14	28	10	14	1
<b>h56</b>	13	15	24	10	14	1
<b>h57</b>	13	16	26	10	14	1
<b>h58</b>	14	11	27	10	14	1
<b>h59</b>	14	14	27	10	14	1
<b>H60</b>	14	15	24	10	13	1

---

## **CAPÍTULO II**

Elsevier Editorial System(tm) for Forensic Science International

Manuscript Draft

Manuscript Number:

Title: Genetic analysis of 5 X-STR loci in the population of São Paulo State (Southeast of Brazil)

Article Type: Original Research Paper

Section/Category: Forensic Genetics

Keywords: X-STR; Genetic Population Data; São Paulo State; Brazil

Corresponding Author: M.D. Joyce Aparecida Martins, M.D

Corresponding Author's Institution: UNESP-School of Pharmaceutical Sciences, State University of São Paulo

First Author: Joyce Aparecida Martins, M.D.

Order of Authors: Joyce Aparecida Martins, M.D.; Ricardo Henrique A da Silva, Ph.D; Adriana Freschi; Greiciane G Paneto, M.D.; Regina Maria B Cicarelli, Ph.D; Rogério N de Oliveira, Ph.D

Abstract: Allele frequency distributions and population data for 5 X-STR (DXS6854, DXS7424, DXS101, DXS6808 and DXS7132) were obtained from sample of 243 unrelated males and female living in the population of São Paulo State (Southeast of Brazil). The forensic efficiency values demonstrate that especially DXS101 is highly informative marker, followed by DXS7424, DXS6854, DXS7132 and DXS6808. No deviations from Hardy-Weinberg equilibrium were observed, with the exception of DXS101. The combined powers of discrimination in men and women were 0.9992 and 0.999994, respectively. These high values show the potential of this system in human identification and paternity testing. Population comparisons were performed and revealed that there are some differences in allele distributions between our population and European, African and Asiatic populations.

## Genetic analysis of 5 X-STR loci in the population of São Paulo State (Southeast of Brazil)

J.A. Martins<sup>a</sup>, R.H.A. Silva<sup>b</sup>, A. Freschi<sup>a</sup>, G.G. Paneto<sup>a</sup>, R.N. Oliveira<sup>b</sup>, R.M.B. Cicarelli

<sup>a</sup>

<sup>a</sup>UNESP – School of Pharmaceutical Sciences, State University of São Paulo, Brazil.

<sup>b</sup>USP – School of Dentistry, University of São Paulo, Brazil.

### Abstract

Allele frequency distributions and population data for 5 X-STR (DXS6854, DXS7424, DXS101, DXS6808 and DXS7132) were obtained from sample of 243 unrelated males and female living in the population of São Paulo State (Southeast of Brazil). The forensic efficiency values demonstrate that especially DXS101 is highly informative marker, followed by DXS7424, DXS6854, DXS7132 and DXS6808. No deviations from Hardy–Weinberg equilibrium were observed, with the exception of DXS101. The combined powers of discrimination in men and women were 0.9992 and 0.999994, respectively. These high values show the potential of this system in human identification and paternity testing. Population comparisons were performed and revealed that there are some differences in allele distributions between our population and European, African and Asiatic populations.

*Keywords:* X-STR; Genetic Population data; São Paulo State; Brazil

### 1. Introduction

The short tandem repeat (STR) markers have been extensively studied for forensic purpose and applied in human identification and paternity testing. Most forensic interests

have been focused on autosomal (1-2) and Y-chromosomal STRs (3-4). However, the analysis of X-chromosome short tandem repeats (X-STRs) has become of increasingly interest in the forensic practice, specially in complex cases of kinship testing where the offspring is female and the alleged father is unavailable or when there is difficult to analyse template materials, such as DNA from exhumed skeletons. In such instances sufficient statistical power has to come from a small number of X-STRs, because these markers have high Mean Exclusion Chance (5).

As population genetic data of these markers are still little known in Brazilian population, the aim of the present study was to analyse 5 X-STRs (DXS6854, DXS7424, DXS101, DXS6808, DXS7132) in the population of São Paulo State and to compare our results with those from other populations studies.

## **2. Population**

Two hundred forty-three unrelated healthy individuals (109 females and 134 males) living in the population of São Paulo State. All participants provided written consent for the use of genetic data.

## **3. Extraction**

Blood or buccal samples were collected in FTA cards (Whatman, Clifton, NJ, USA) and genomic DNA was extracted using DNA IQ kit (Promega, Madison, WI, USA).

## **4. PCR**

Amplification was carried out in a multiplex system for DXS6854, DXS7424, DXS6808 and DXS7132 and, monoplex system for DXS101, using a GeneAmp PCR System 9700 thermocycler (Applied Biosystems, Foster City, CA, USA). The primer sequences were

obtained from Genome Database (<http://www.gdb.org>) and labeled at the 5'-end with the fluorescent dye Cy5.

The quadruplex PCR was performed in a 10  $\mu$ L final reaction volume containing 1-2 ng DNA, 1.0  $\mu$ L of Gold Star Buffer (Promega, Madison, WI, USA), 0.35  $\mu$ M each primer (IDT, Coralville, IA), 1.5U PlatinumTaq DNA Polymerase (Invitrogen, Carlsbad, CA, USA), brought to the volume with ddH<sub>2</sub>O. The amplification components for monoplex systems were practically the same of the quadruplex, but were utilized 0.30  $\mu$ M each primer (IDT, Coralville, IA) and 0.4U PlatinumTaq DNA Polymerase (Invitrogen, Carlsbad, CA, USA).

A touchdown protocol PCR was carried out for quadruplex systems as follows: initial denaturation for 5 min at 95°C, six cycles with 1 min at 95°C, 1 min at 67-59°C, decreasing the annealing temperatures by 4°C in each two cycles, and 1 min at 72 °C, then 30 cycles with constant annealing at 55°C, followed by a final extension at 72 °C for 10 min. For monoplex systems, the thermal cycling conditions were: initial denaturation for 5 min at 94°C, followed by 35 cycles of 1 min at 94°C, 1 min at 55°C and 2 min at 72°C with a final incubation for 10 min at 72°C.

## 5. Typing

PCR products were analyzed by electrophoresis using an ALF Express Automated Sequencer (GE Healthcare, Chalfont, St. Giles, UK) with 6% denaturing polyacrylamid gel and Allelelocator v. 1.03 software (GE Healthcare, Chalfont, St. Giles, UK) was used to determine fragment size.

Allele typing was carried out based on home-made allelic ladders and using the K562 cell line sample (Promega, Madison, WI, USA) which were used also as control DNA for

calibrating allelic ladders. Allele were assigned according to the recommendations of the International Society of Forensic Science (ISFG) Commission [6]

## 6. Statistical Analysis

The allele frequencies, exact test of population differentiation, population pairwise genetic distance ( $R_{ST}$ ), and exact test to verify Hardy-Weinberg equilibrium (HWE) and the possibility of linkage disequilibrium were calculated using software Arlequin ver. 3.1 [7].

Parameters of forensic interest, namely mean exclusion chance in trios involving daughters ( $MEC_T$ ) as well as in father/daughter duos ( $MEC_D$ ) and power of discrimination in females ( $PD_F$ ) and in males ( $PD_M$ ) were computed as proposed by Desmarais et al. [8], using pool data. The polymorphism information content (PIC) and expected heterozygosity value ( $HET_{exp}$ ) were calculated as suggested by Botstein et al. [9] and Nei and Roychoudhury [10], respectively, using female data.

## 7. Access to the data

The complete data are available upon request: joyce\_apa@hotmail.com

## 8. Results

Allele frequencies of the five markers were computed separately by gender and exact test of population differentiation revealed no significant differences between females and males, therefore only the combined frequencies are shown in Table 1.

Statistical parameters of forensic interest are shown in Table 2. STR DXS101 had the highest power of discrimination in male and female, as well as the highest polymorphic information content, probabilities of exclusion in trios involving daughters and father/daughter duos, followed by DXS7424, DXS6854, DXS7132 and DXS6808. When the

forensic efficiency of the 5 X-STR was calculated, the combined  $PD_M$  and  $PD_F$  were 0.9992 and 0.999994, respectively and the combined power of exclusion in trio and motherless cases were 0.998 and 0.98, respectively. These high values show the potential of this system in human identification and paternity testing.

No deviations from Hardy-Weinberg equilibrium were observed, with exception of the DXS101 locus (Table 2). Due to the very high number of alleles at DXS101, the relatively small population size might have been inadequate to produce reliable allele frequencies. Such results have been shown by some authors [11, 12].

The marker DXS7132 is included in Marshfield genetic linkage maps; for the other, the map position was determined by their localization on the physical map and by alignment with closely markers.

Since all markers are located on the q-arm of X-chromosome (Fig.1), intermarker linkage disequilibrium was analyzed. According to the genetic localization, the loci DXS7424 - DXS101 and DXS6854 – DXS6808 are very close, but were not detected linkage disequilibrium between them. However, significant result ( $p < 0.05$ ) was obtained between DXS101 and DXS6808; as these markers are quite distant on the chromosome, a linkage disequilibrium manifestation is unlikely and this data is best attributed to sampling affects. The association of the DXS101 with others distant markers was also observed in Asians American [13]. Moreover, among 134 men typed, 122 haplotypes were unique and 6 were found in 2 individuals.

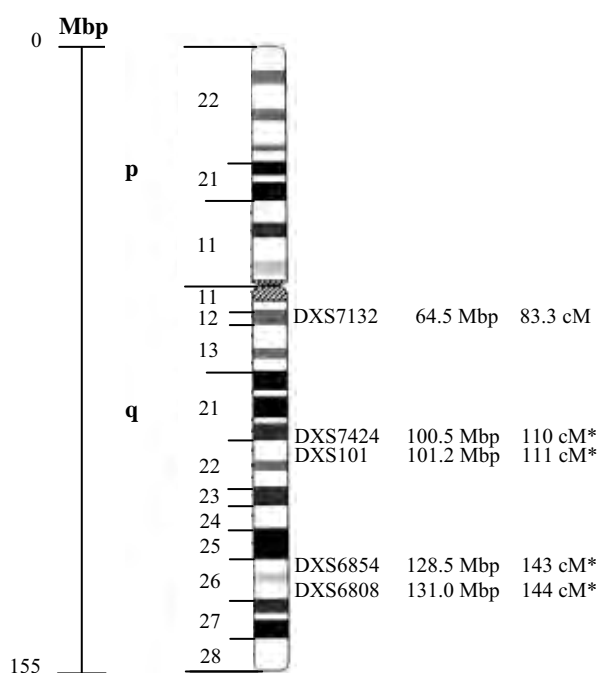
## **9. Other remarks**

Although the initial PCR was designed as a pentaplex reaction, the amplification simultaneous of DXS101 with others markers was difficult. Therefore, a quadruplex and a monoplex for DXS101 were carried out.

Allele frequencies at three of the five loci examined in this study (DXS7424, DXS101 and DXS7132) were compared with data of others populations. At DXS7132 exact test of population differentiation and genetic distance analysis revealed no significant results (exact  $p$ -values $>0.01$ ; RST  $p$ -values $>0.05$ ) between our population and Spain [11], Portugal [14], Angola [15] and Mozambique [15], but differed from Japan [16] population. Moreover, the allele 10 was found only in our population while the alleles 9, 17 and 18 were not found in our study. In the case of DXS7424, we performed comparisons between the populations of São Paulo State and of Germany [17] and the tests showed no significant differences (exact  $p$ -values $>0.01$ ; RST  $p$ -values $>0.05$ ), being that the allele 8 was unique in our population. For DXS101, the allele distribution was similar (exact  $p$ -values $>0.01$ ; RST  $p$ -values $>0.05$ ) only with Angola [15] and differed from Spain [11], Portugal [14], Mozambique [15] and Germany [17], being that the allele 31 was unique in our population.

In summery, the system studied would appear to be a valuable tool for forensic purposes in the population of São Paulo State.

This paper follows the guidelines for publication of population data request the journal [18].



**Fig.1** Positions of 5 X-STR used in this study on chromosome X ideogram. Detailed positions were obtained in mega base pairs (<http://www.ncbi.nlm.nih.gov>) and centimorgans (<http://www.marshfieldclinic.org/research/genetics/>). \*Value interpolated from nearby markers.

Table 1: Allele frequencies for five X-STRs in the population of São Paulo State

Alelo	DXS6854	DXS7424	DXS101	DXS6808	DXS7132
7	0.003				
8	0.006	0.003			
9	0.074			0.114	
10	0.210	0.009		0.713	0.006
11	0.111	0.017		0.165	0.011
12	0.375	0.026		0.006	0.102
13	0.153	0.094			0.264
14	0.068	0.188	0.003	0.003	0.355
15		0.298	0.006		0.230
16		0.267	0.009		0.031
17		0.082	0.011		
18		0.014	0.034		
19			0.045		
20		0.003	0.037		
21			0.065		
22			0.031		
23			0.063		
24			0.179		
25			0.119		
26			0.216		
27			0.088		
28			0.060		
29			0.020		
30			0.011		
31			0.003		

Table 2: Statistical parameters for 5 STRs in the population of São Paulo State

Parameters	DXS6854	DXS7424	DXS101	DXS6808	DXS7132
HETobs	0.789	0.780	0.771	0.395	0.716
HETexp	0.773	0.798	0.882	0.444	0.746
HWE	0.483	0.991	0.002	0.430	0.811
PIC	0.740	0.770	0.870	0.410	0.700
MEC <sub>T</sub>	0.738	0.757	0.871	0.410	0.695
MEC <sub>D</sub>	0.606	0.629	0.782	0.271	0.556
PD <sub>F</sub>	0.916	0.924	0.975	0.658	0.888
PD <sub>M</sub>	0.769	0.788	0.881	0.451	0.740

Observed heterozygosity (HETobs); expected heterozygosity (HETexp); Hardy-Weinberg equilibrium p-value (HWE); polymorphism information content (PIC); mean exclusion chance in trios (MEC<sub>T</sub>) and in father/daughter duos (MEC<sub>D</sub>); power of discrimination in female (PD<sub>F</sub>) and in male (PD<sub>M</sub>).

### Acknowledgements

The authors would like to thank BIOCOD Laboratory (Belo Horizonte, MG, Brazil), where part of this work was done. This study was partially granted by PADC/FCF-UNESP (proc. 2006/05-I) and JAM was the recipients of a fellowship from CAPES.

### References

- [1] A. Urquhart, C.P. Kimpton, T.J. Downes, P. Gill, Variation in short tandem repeat sequences: a survey of twelve microsatellite loci for use as forensic identification markers, *Int J Legal Med* 107 (1994) 13–20.
- [2] B. Brinkmann, Overview of PCR-based systems in identity testing, *Methods Mol Bio* 98 (1998) 105–119.
- [3] M.A. Jobling, A. Pandya, C. Tyler-Smith, The Y chromosome in forensic analysis and paternity testing, *Int J Legal Med* 110 (1997) 118–124.
- [4] M. Kayser, A. Cagliá, D. Corach et al. Evaluation of Ychromosomal STRs:a multicenter study, *Int J Legal Med* 110 (1997) 125–133.
- [5] R. Szibor, M. Krawczak, S. Hering, J. Edelmann, E. Kuhlisch, D. Krause, Use of X-linked markers for forensic purposes, *Int J Legal Med* 117 (2003) 67-74.

- [6] W. Bär, B. Brinkmann, B. Budowle, A. Carracedo, P. Gill, P. Lincoln, W. Mayr, B. Olaisen, DNA recommendations. Further report of the DNA Commission of the ISFH regarding the use of short tandem repeat systems, International Society for Forensic Haemogenetics, *Int J Legal Med* 110 (1997) 175–176.
- [7] Excoffier, L. G. Laval, S. Schneider, Arlequin version. 3.0: An integrated software package for population genetics data analysis, *Evol Bioinformatics Online* 1 (2005) 47–50.
- [8] 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) 1046–1049.
- [9] D. Botstein, R.I. White, M. Skolnich, R.W. Davis, Construction of a genetic linkage map in man using restriction fragment length polymorphisms, *Am J Hum Genet* 32 (1980) 314–331.
- [10] M. Nei, A.K. Roychoudhury, Sampling variances of heterozygosity and genetic distance, *Genetics* 76(1974) 379–390.
- [11] M. Aler, P. Sanchez-Diz, I. Gomes, M. Gisbert, A. Carracedo, A. Amorim, L. Gusmão, Genetic data of 10 X-STRs in a Spanish population sample, *Forensic Sci. Int.* (2007), doi:10.1016/j.forsciint.2006.10.023.
- [12] C. Bini, S. Ceccardi, G. Ferri, S. Pelotti, M. Alù, E. Roncaglia, G. Beduschi, L. Caenazzo, E. Ponzano, P. Tasinato, C. Turchi, L. Buscemi, M. Mazzanti, A. Tagliabracci, C. Toni, I. Spinetti, R. Domenici, S. Presciuttini, Development of a heptaplex PCR system to analyse X-chromosome STR loci from five Italian population samples: A collaborative study, *Forensic Sci. Int.* 153 (2005) 231–236.
- [13] I.Gomes, M. Prinz, R. Pereira, C. Meyers, R.S. Mikulasovich, A. Amorim, A. Carracedo, L. Gusmão, Genetic analysis of three US population groups using an X-chromosomal STR decaplex, *Forensic Sci. Int.* (2007), doi:10.1007/s00414-006-0146-2.
- [14] R. Pereira, I. Gomes, A. Amorim, L. Gusmão, Genetic diversity of 10 X chromosome STRs in northern Portugal, *Int J Legal Med* (2007), doi:10.1007/s00414-006-0144-4.
- [15] I. Gomes, C. Alves, K. Maxzud, R. Pereira, M.J. Prata, P. Sánchez-Diz, A. Carracedo, A. Amorim, L. Gusmão, Analysis of 10 X-STRs in three African populations, *Forensic Sci. Int.: Genetics* (2007) doi:10.1016/j.fsigen.2007.01.001.
- [16] H. Asamura, H. Sakai, M. Ota, H. Fukushima, Japanese population data for eight X-STR loci using two new quadruplex systems, *Int J Legal Med* 120 (2006) 303–309.

- [17] 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.
- [18] P. Lincoln, A Carracedo, Publication of population data of human polymorphisms, *Forensic Sci. Int.* 110(2000) 3-5.

**Journal of Forensic Science****Genetic analysis of X and Y Short Tandem Repeat Loci in Araraquara population, Brazil<sup>2</sup>**

**Joyce A. Martins,<sup>1</sup> M.S.; Ricardo Henrique A. Silva,<sup>2</sup> Ph.D.; Carolina C. Góis,<sup>2</sup> M.S.; Greiciane G. Paneto,<sup>1</sup> M.S.; Rogério N. Oliveira,<sup>2</sup> Ph.D.; Regina M. B. Cicarelli,<sup>1</sup> Ph.D.**

<sup>1</sup>UNESP - São Paulo State University, School of Pharmaceutical Sciences, Paternity Laboratory, Araraquara-Jaú Road, km1, 14801-902, Araraquara, São Paulo, Brazil.

<sup>2</sup>USP - University of São Paulo, School of Dentistry, Legal Dentistry Laboratory of Molecular Biology, Avenue Professor Lineu Prestes, 2227, 05508-000, São Paulo, Brazil.

**Corresponding author:** Joyce Aparecida Martins

UNESP - São Paulo State University, School of Pharmaceutical Sciences, Paternity Laboratory, Araraquara-Jaú Road, km1, 14801-902, Araraquara, São Paulo, Brazil.

E-mail: joyce\_apa@hotmail.com

---

<sup>2</sup> This study was partially supported by PADCF/FCF-UNESP (2006/05-I) and JAM was the recipient of fellowship from CAPES.

**POPULATION:** One hundred and twenty unrelated individuals from Araraquara (Southeast of Brazil).

**KEYWORDS:** forensic science, DNA typing, X-STR, Y-STR, population genetics, Araraquara, Brazil, DXS6854, DXS7424, DXS101, DXS6808, DXS7132, DYS391, DYS389I/II, DYS439, DYS438, DYS437, DYS19, DYS392, DYS393, DYS390, DYS385.

## Extraction

After informed written consent, blood samples from 120 unrelated individuals (60 men and 60 women) living in Araraquara, southeast of Brazil, were collected on FTA cards (Whatman, Clifton, NJ) and genomic DNA was extracted using DNA IQ kit (Promega, Madison, WI) (1).

Amplification was carried out in a multiplex system for DXS6854, DXS7424, DXS6808 and DXS7132 and monoplex system for DXS101 using a GeneAmp<sup>®</sup> PCR System 9700 thermocycler (Applied Biosystems, Foster City, CA). The quadruplex PCR was performed in a 10  $\mu$ L final reaction volume containing 1-2 ng DNA, 1X GoldStar Buffer (Promega, Madison, WI), 0.35  $\mu$ M each primer (IDT, Coralville, IA), 1.5 U PlatinumTaq DNA Polymerase (Invitrogen, Carlsbad, CA), made up to volume with ddH<sub>2</sub>O. The amplification components for the monoplex system were practically the same as the quadruplex, but 0.30  $\mu$ M of primer (IDT, Coralville, IA) and 0.4 U PlatinumTaq DNA Polymerase (Invitrogen, Carlsbad, CA) were used. A touchdown protocol was carried out for the quadruplex system as follows: initial denaturation for 5 min at 95°C, six cycles of 1 min at 95°C, 1 min at 67-59°C, decreasing the annealing temperatures by 4°C after each two cycles, and 1 min at 72°C, then 30 cycles of constant annealing at 55°C, followed by a final extension at 72°C for 10 min. For the monoplex system, the thermal cycling conditions were: initial denaturation for 5 min at 94°C, followed by 35 cycles of 1 min at 94°C, 1 min at 55°C and 2 min at 72°C with a final incubation for 10 min at 72°C.

Amplification of 12 Y-STR loci was performed using a PowerPlex<sup>®</sup>Y Systems (Promega, Madison, WI) in a GeneAmp PCR System 9700 thermocycler (Applied Biosystems, Foster City, CA), following the manufacturer's recommendations.

## Typing

The X-STR and Y-STR amplicons were detected by electrophoresis using ALF Express Automated Sequencer (GE Healthcare, Chalfont, St. Giles, UK) and ABI377 Automated Sequencer (Applied Biosystems, Foster City, CA), respectively. Fragment sizes were determined automatically using the Allelelocator, version 1.03 (GE Healthcare, Chalfont, St. Giles, UK) and GeneScan<sup>®</sup> Analysis 2.1 Software (Applied Biosystems, Foster City, CA), respectively.

## Results and Analysis

Population statistics data were obtained using the Arlequin, version 3.11 (2), mean exclusion chance in trios involving daughters ( $MEC_T$ ) as well as in father/daughter duos ( $MEC_D$ ) and power of discrimination in women ( $PD_F$ ) and in men ( $PD_M$ ) were calculated using formulae according to Desmarais et al. (3). Gene and haplotype diversities were analyzed according to Hou's method (4). Allele frequencies and statistical parameters of the 5 X-STR and 12 Y-STR are listed in Table 1 and 2, respectively.

For X-STR, no deviations from the Hardy-Weinberg equilibrium were observed with exception of the DXS101. Since all markers are located on the q-arm of the X-chromosome, intermarker linkage disequilibrium was analyzed and significant result ( $p=0.014$ ) was obtained between DXS101 and DXS6808; as these markers are quite distant on the chromosome, a linkage disequilibrium manifestation is unlikely and this data is best attributed to sampling effects. The association of the DXS101 with other distant markers was also observed in Asian Americans (5). The STR DXS101 had the highest power of discrimination in males and females, as well as the highest probabilities of exclusion in trios involving daughters and father/daughter duos, followed by DXS7424, DXS6854, DXS7132 and DXS6808. When the forensic efficiency of the 5 X-STR was calculated, the combined  $PD_F$

and  $PD_M$  were 0,9999928 and 0,9990702, respectively and the combined  $MEC_T$  and  $MEC_D$  were 0.9983 and 0.9887, respectively.

In addition, the average gene diversity of the 12 Y-STR was 0.6923, the loci DYS385 and DYS390 being the most polymorphic and loci DYS391, DYS393 and DYS437, the least polymorphic, below the 0.6 range. A total of 59 different haplotypes were obtained out of the 60 men studied and a high discrimination capacity of 0.9833 and haplotype diversity of 0.9994 was observed.

The data generated revealed that the systems studied are highly polymorphic and discriminative in Araraquara population, Brazil. The complete data are available upon request: [joyce\\_apa@hotmail.com](mailto:joyce_apa@hotmail.com)

## References

1. <http://www.promega.com/applications/hmnid/productprofiles/dnaiq/>
2. <http://cmpg.unibe.ch/software/arlequin3/>
3. 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* 1998;43:1046–1049.
4. Hou YP, Zhang J, Li YB, Wu J, Zhang S, Prinz M. Allele sequences of six new Y-STR loci and haplotypes in the Chinese Han population. *Forensic Sci Int* 2001;118:147–52.
5. Gomes I, Prinz M, Pereira R, Meyers C, Mikulasovich RS, Amorim A, Carracedo A, Gusmão L. Genetic analysis of three US population groups using an X-chromosomal STR decaplex, *Forensic Sci Int* 2007; 121(3):198-203.

Additional information and reprint request:

Joyce Aparecida Martins, M.S.

UNESP - São Paulo State University, School of Pharmaceutical Sciences, Paternity Laboratory, Araraquara-Jaú Road, km1, 14801-902, Araraquara, São Paulo, Brazil.

E-mail: [joyce\\_apa@hotmail.com](mailto:joyce_apa@hotmail.com)

Table 1: Allele frequencies and statistical parameters for 5 X-STRs in Araraquara population (n=120).

Allele	DXS6854	DXS7424	DXS101	DXS6808	DXS7132
7					
8	0.0056				
9	0.1111			0.1167	
10	0.2222	0.0111		0.7278	0.0056
11	0.1333	0.0278		0.1500	0.0056
12	0.3722	0.0167			0.0944
13	0.0944	0.1111			0.2833
14	0.0611	0.2111		0.0056	0.3333
15		0.3000			0.2556
16		0.2389	0.0111		0.0222
17		0.0667	0.0056		
18		0.0167	0.0167		
19			0.0333		
20			0.0389		
21			0.0667		
22			0.0444		
23			0.0500		
24			0.1222		
25			0.1111		
26			0.2667		
27			0.1111		
28			0.0722		
29			0.0222		
30			0.0222		
31			0.0056		
<i>Parameters</i>					
Ho	0.817	0.783	0.817	0.417	0.783
He	0.799	0.804	0.864	0.429	0.747
HWE	0.832	0.650	0.002	0.590	0.818
MEC <sub>T</sub>	0.738	0.760	0.860	0.395	0.690
MEC <sub>D</sub>	0.606	0.632	0.767	0.259	0.550
PD <sub>F</sub>	0.916	0.925	0.972	0.641	0.884
PD <sub>M</sub>	0.769	0.790	0.871	0.434	0.737

Ho, observed heterozygosity; He, expected heterozygosity; HWE, Hardy-Weinberg equilibrium p-value; MEC<sub>T</sub>, mean exclusion chance in trios; MEC<sub>D</sub>, mean exclusion chance in father/daughter duos; PD<sub>F</sub>, power of discrimination in female; PD<sub>M</sub>, power of discrimination in male.

Table 2: Allele/genotype frequencies and gene diversity values at 12 Y-STR in Araraquara population (n=60).

Allele	DYS 391	DYS 389I	DYS 439	DYS 389II	DYS 438	DYS 437	DYS 19	DYS 392	DYS 393	DYS 390	DYS Genotype	DYS 385	DYS Genotype	DYS 385
8										9-15	0.017	15-16	0.033	
9	0.083		0.017		0.067					10-18	0.017	16-17	0.067	
10	0.533		0.050		0.450			0.017		10-19	0.017	16-20	0.017	
11	0.367	0.017	0.267		0.200			0.483		11-12	0.033	17-17	0.033	
12	0.017	0.217	0.433		0.283			0.100	0.183	11-14	0.200	18-18	0.017	
13		0.517	0.217				0.167	0.283	0.583	11-15	0.017			
14		0.217	0.017			0.600	0.333	0.083	0.200	11-16	0.033			
15		0.033				0.350	0.300	0.017	0.033	12-13	0.017			
16						0.050	0.117	0.017		12-14	0.033			
17							0.083			12-15	0.033			
21										13-13	0.017			
22										13-14	0.133			
23										13-17	0.033			
24										13-18	0.017			
25										14-14	0.033			
27				0.050						14-15	0.017			
28				0.083						14-16	0.017			
29				0.433						14-17	0.050			
30				0.217						14-18	0.033			
31				0.183						14-19	0.017			
32				0.033						15-15	0.050			
GD	0.584	0.648	0.703	0.734	0.684	0.524	0.763	0.680	0.595	0.765				0.935

GD, gene diversity.