

UNIVERSIDADE ESTADUAL PAULISTA JULIO DE MESQUITA FILHO

Faculdade de Ciências Farmacêuticas

CYNTIA SUMIE YOSHIZAKI

**ANÁLISE DE 15 STRS AUTOSSÔMICOS NA POPULAÇÃO DE
ARARAQUARA (SÃO PAULO, BRASIL).**

ARARAQUARA-SP

2011

UNIVERSIDADE ESTADUAL PAULISTA JULIO DE MESQUITA FILHO

Faculdade de Ciências Farmacêuticas

CYNTIA SUMIE YOSHIZAKI

**ANÁLISE DE 15 STRS AUTOSSÔMICOS NA POPULAÇÃO DE
ARARAQUARA (SÃO PAULO, BRASIL).**

Trabalho de Conclusão de Curso
apresentado à Universidade Estadual
Paulista Júlio de Mesquita Filho como
requisito para obtenção do grau de
Farmacêutico-Bioquímico.

Orientadora: Profa. Dra. Regina Maria
Barretto Cicarelli.

Co-orientadora: Joyce Aparecida Martins
Lopes Ferraz.

ARARAQUARA-SP

2011

DEDICATÓRIA

À **Deus**, por tornar possível a conclusão de mais uma etapa, iluminando sempre o caminho;

Aos meus pais, **Yogi** e **Yvette**, por serem responsáveis por cada conquista, sendo sempre exemplo de coragem, força e superação. Pela educação e pelo amor incondicional;

Aos meus irmãos, **Ricieri** e **Vanessa**, pela amizade, pelo amor, por tudo que passamos juntos e pelo que aprendemos uns com os outros desde pequenos;

Aos meus sobrinhos, **João Guilherme**, **Maria Fernanda** e **Ana Júlia**, por serem a razão da minha felicidade.

AGRADECIMENTOS

A minha família, por tudo que sempre fizeram por mim, pelo exemplo, amizade e carinho, fundamentais na construção do meu caráter;

À república Raparigaz, que sempre foi o porto seguro e meu lar em Araraquara, um lugar de muita felicidade;

Às companheiras de república, Lígia Scarpa, Maíra, Anna Maria, Lígia Beltrame e Bruna, 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;

Aos grandes amigos, pelos momentos de dificuldade e de alegria que passamos juntos. Agradeço pela compreensão nas vezes em que estive ausente e por, mesmo assim, estarem dispostos a me ouvir e aconselhar. Obrigada por tornarem inesquecível cada momento ao longo desses anos;

Aos professores que passaram por minha vida, em especial à Prof^a. Regina M. B. Cicarelli, orientadora desse projeto, e à doutoranda e amiga Joyce Aparecida Martins Lopes Ferraz por ter acreditado que seria possível, pelos conselhos e orientação;

A todos do Laboratório de Investigação de Paternidade da Faculdade de Ciências Farmacêuticas da Unesp – Câmpus de Araraquara, pela amizade e companheirismo;

À Faculdade de Ciências Farmacêuticas – UNESP e a todos os funcionários, por todo o conhecimento e atenção;

Às pessoas que, direta ou indiretamente, contribuíram com esse trabalho - técnicos, professores, funcionários e colegas - meu muito obrigada!

SUMÁRIO

Resumo

Lista de Figuras

Lista de Tabelas

1. Introdução.....	10
1.1. A molécula de DNA e o genoma humano.....	10
1.2. Marcadores biológicos.....	12
1.3. Frequências alélicas e parâmetros estatísticos	16
2. Objetivos.....	20
3. Material e Métodos.....	21
4. Resultados e Discussão.....	25
5. Conclusões.....	34
6. Referências Bibliográficas	35

RESUMO

Os marcadores genéticos mais utilizados para fins de identificação humana são regiões autossômicas de repetições consecutivas curtas (*Short Tandem Repeats* - STRs) e para interpretar a análise de DNA, os resultados de um caso necessitam ser comparados com os dados de uma pertinente população. O objetivo do trabalho, portanto, foi caracterizar o perfil genético da população de Araraquara, (São Paulo, Brasil) pela análise de 15 STRs autossômicos (D3S1358, TH01, D21S11, D18S51, Penta E, D5S818, D13S317, D7S820, D16S539, CSF1PO, Penta D, VWA, D8S1179, TPOX and FGA) inclusos no kit PowerPlex 16 (Promega) e correlacionar esses dados com a história de formação da população. Não foram observados desvio do equilíbrio de Hardy - Weinberg após correção de Bonferroni. Parâmetros forenses apresentaram valores elevados, sendo os marcadores mais polimórficos o Penta E, D18S51 e FGA. A árvore UPGMA (*Unweighted Pair- Group Method with Arithmetic Mean*) baseada na medida de distância genética mostrou a população de Araraquara agrupada com as populações da região sudeste do Brasil, próxima do grupo europeu e distante das populações Africana e Ameríndia. Estimativas de mistura revelaram que a contribuição parental para a população de Araraquara foi 76% européia, 18% africana e 6% ameríndia.

Palavras-Chave: STR, Frequências Alélicas, PowerPlex 16, Araraquara, Brasil.

LISTA DE FIGURAS

Figura 1. Esquema do genoma humano constituído por 23 pares de cromossomos e pelo material genético encontrado no interior da organela celular mitocôndria (DNA mitocondrial). Fonte: BUTLER, J. M. Forensic DNA Typing. London: Elsevier Academic Press, 2005.....	11
Figura 2. Esquema de um marcador minissatélite (D1S80) e um marcador STR (TH01). Fonte: BUTLER, J. M. Forensic DNA Typing. London: Elsevier Academic Press, 2005.....	14
Figura 3. Árvore filogenética UPGMA baseada na distância genética de Reynolds.....	31

LISTA DE TABELAS

Tabela 1. Frequências alélicas de 15 STRs autossômicos na população de Araraquara - SP, Brasil	25
Tabela 2. Parâmetros estatísticos de 15 STRs autossômicos na população de Araraquara - SP, Brasil.....	27

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).

1.1. A molécula de DNA e o genoma humano

Em 1953, os pesquisadores James Watson e Francis Crick deduziram a maneira como a molécula de DNA é organizada. Esta consiste em duas cadeias de nucleotídeos complementares e dispostas de forma helicoidal, sendo que cada nucleotídeo apresenta três componentes: molécula de açúcar (desoxirribose), grupamento fosfato e base nitrogenada (adenina (A), guanina (G), citosina (C) ou timina (T) (SNUSTAD e SIMMONS, 2008).

Na formação de uma cadeia, o grupamento fosfato de um nucleotídeo se liga à molécula de açúcar do outro nucleotídeo através de ligações fosfodiéster; as bases nitrogenadas, por sua vez, estão ligadas a cada açúcar na cadeia e envolvidas no pareamento entre estas, através de ligações fracas, denominadas por ligações de hidrogênio. Nesta interação, adenina estabelece ligação com timina e guanina com citosina, sendo que a união de base pirimidina (T ou C) a sua purina (A ou G) complementar na cadeia oposta é o que se denomina de “par de bases”. O pareamento de bases complementares contribui para dar maior estabilidade à molécula de DNA e permite o conhecimento da sequência de nucleotídeos em uma fita através da sua complementar (SNUSTAD e SIMMONS, 2008).

O genoma humano nas células somáticas está contido em 46 cromossomos no núcleo (22 pares de cromossomos homólogos e 1 par de cromossomos sexuais X/Y), denominado de DNA nuclear, e em uma dupla fita circular presente no interior da organela celular mitocôndria, denominado DNA mitocondrial (Figura 1) (BUTLER, 2005).

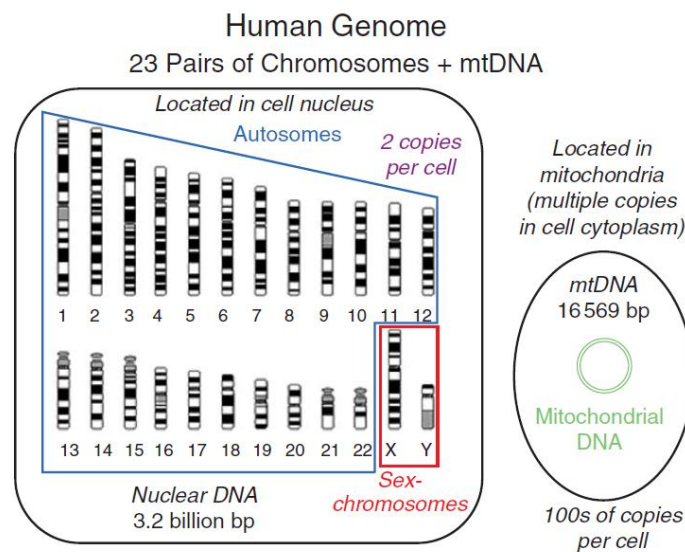


Figura 1. Esquema do genoma humano constituído por 23 pares de cromossomos e pelo material genético encontrado no interior da organela celular mitocôndria (DNA mitocondrial). Fonte: BUTLER, J. M. Forensic DNA Typing. London: Elsevier Academic Press, 2005.

O DNA nuclear é biparental (metade dos cromossomos herdada da mãe e a outra metade herdada do pai) e possui apenas uma cópia por célula; enquanto o DNA mitocondrial é uniparental (herdado apenas da mãe em todos os indivíduos) e possui muitas cópias por célula, pois uma única célula pode ter dezenas a centenas de mitocôndrias e em cada mitocôndria podem existir dezenas a centenas de cópias do DNA mitocondrial (PANETO, 2010).

Nos gametas, há diferença em relação ao DNA nuclear, o qual é haplóide, apresentando 23 cromossomos de origem materna (no caso do óvulo) ou de origem paterna (no caso do espermatozóide). Assim, uma pessoa possui bilhões de células somáticas em seu corpo e todas elas, com exceção de seus gametas, são, em termos genéticos, idênticas. São cópias fiéis oriundas de uma única célula diplóide (fusão do óvulo com o espermatozóide) que se multiplicou através do processo de divisão celular mitótico (BONACCORSO, 2004).

Todos os indivíduos, à exceção de gêmeos monozigóticos, possuem um patrimônio genético único, o qual é constituído pela combinação da informação genética de diversos marcadores herdados dos seus progenitores. A localização de um marcador genético no genoma é dito “locus” e as diferentes formas de um marcador genético são denominadas “alelos”. O indivíduo que carrega dois alelos iguais para um marcador genético é dito homozigoto; enquanto o heterozigoto carrega dois alelos diferentes para o mesmo marcador. Devido ao mecanismo de herança biológica, as relações entre pais e filhos, entre irmãos, tio-sobrinho e outras podem ser inferidas pelos marcadores genéticos.

1.2. Marcadores biológicos

Inicialmente, diversos marcadores protéicos foram tipados para detectar polimorfismos a fim de se diferenciar indivíduos. Dentre estes sistemas, podemos citar os grupos sanguíneos ABO, componentes grupo-específicos (transferrina, albumina, ceruloplasmina, haptoglobina, fosfoglicomutase-1, fosfatase ácida, esterase D, etc) e o sistema HLA (*histocompatibility leucocyte antigen*) (JORDE, WATKINS, BAMSHAD, 2001; CALABREZ, 1999). No entanto, o uso de proteínas

apresenta alguns inconvenientes, tais como: a baixa estabilidade em amostras biológicas expostas ao ambiente; o baixo poder de discriminação e a ausência em determinados materiais biológicos (WEEDN; SWARNEN, 1998). A análise do DNA contornou essas limitações, tendo início em 1985 com o geneticista inglês Alec Jeffreys e seu grupo de trabalho, os quais descobriram 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).

Nos testes de identificação humana, portanto, são estudadas regiões hipervariáveis, as quais apresentam dois tipos de polimorfismos:

- Polimorfismos de comprimento, originados por inserções ou deleções de um ou mais nucleotídeos.
- Polimorfismos de sequência, originados por substituição de um ou mais nucleotídeos numa sequência de DNA;

Dentre os polimorfismos de comprimento temos as Repetições de Sequência Simples (RSS) que incluem os marcadores genéticos VNTRs (*Variant Number of Tandem Repeats*) ou minissatélites (repetições consecutivas de unidades repetitivas constituídas por 14 a 500 pb) e STRs (*Short Tandem Repeats*) ou microssatélites (repetições consecutivas de unidades repetitivas constituídas por 1 a 13 pb) (Figura 2).

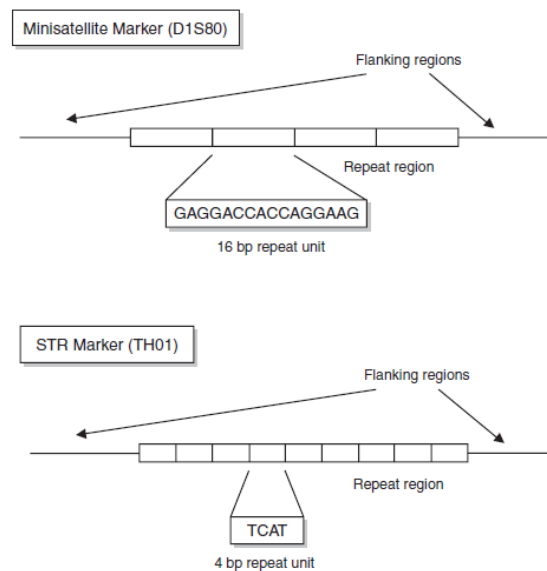


Figura 2. Esquema de um marcador minissatélite (D1S80) e um marcador STR (TH01). Fonte: BUTLER, J. M. Forensic DNA Typing. London: Elsevier Academic Press, 2005.

Os marcadores genéticos STR são classificados de acordo com o número de nucleotídeos que formam o motivo (dinucleotídeos, trinucleotídeos, tetranucleotídeos, etc.) e estão espalhados pelos cromossomos em regiões não codificadoras entre os genes ou dentre eles (íntrons), correspondendo a aproximadamente 3% do genoma humano (SUBRAMANIAN; MISHRA; SINGH, 2003).

Os STRs de maior valor para a identificação humana são aqueles que apresentam alto poder de discriminação (maior que 90%), frequência de heterozigotos maior que 70%, menor tamanho (90 a 500 pb) e baixa frequência de mutações (BUTLER, 2005).

Atualmente, as análises forenses empregam marcadores genéticos presentes tanto no DNA nuclear dos cromossomos autossômicos, bem como dos

cromossomos sexuais (X e Y), e no DNA mitocondrial, sendo que a seleção de um ou outro tipo de marcador irá depender das particularidades de cada caso.

Marcadores do cromossomo Y são ditos de linhagem paterna, pois são passados de pai para filho; enquanto o DNA mitocondrial é de linhagem materna, sendo transferido da mãe para seus filhos. No caso do cromossomo X, a mulher herda um cromossomo de cada um dos pais e o homem herda apenas o cromossomo materno. Diferentemente dos marcadores autossômicos, em que homens e mulheres herdam dois cromossomos: um de herança materna e, o outro, de herança paterna. Em virtude do tipo de herança (somente materna, paterna ou ambas), um maior poder de discriminação é obtido com os marcadores STR autossômicos e, portanto, estes são os de escolha nos exames de DNA.

No entanto, STRs autossômicos não serão de grande utilidade quando não houver DNA nuclear disponível, como em amostras de cabelo sem raiz (faz-se análise de DNA mitocondrial), em ocasiões em que se pretende avaliar o vínculo biológico entre indivíduos que pertençam a gerações muito distantes ou em situações com amostra biológica muito degradada, para as quais geralmente se obtém um perfil alélico autossômico incompleto, partindo-se para análise de polimorfismos de um único nucleotídeo - SNPs (*Single Nucleotide Polymorphisms*), por exemplo.

Com intuito de tornar úteis os bancos de dados dos perfis de DNA, foi essencial uma padronização dos marcadores genéticos usados pelos laboratórios. Inicialmente, o FBI (*Federal Bureau of Investigation*), nos Estados Unidos da América (EUA), iniciou o desenvolvimento de um banco de DNA, conhecido como CODIS (*COmbined DNA Index System*) e recomendou que os laboratórios, responsáveis pela obtenção dos perfis de DNA a serem adicionados no Banco,

analisassem um conjunto de 13 STRs (CSF1PO, FGA, TH01, TPOX, VWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51 e D21S11), os quais ficaram conhecidos como 13 CODIS. Esta recomendação foi aceita pelos laboratórios forenses de todo mundo e, assim, a tipagem de DNA pode ser comparada entre diferentes laboratórios (BUTLER, 2005).

Muitos kits foram desenvolvidos ao longo dos anos e os mais utilizados atualmente são:

- PowerPlex 16 System (Promega): CSF1PO, FGA, TPOX, TH01, VWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, Penta D, Penta E e Amelogenina.
- AmpF ℓ STR Identifier PCR Amplification Kit (Applied Biosystems): CSF1PO, FGA, TPOX, TH01, VWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, D2S1338, D19S433 e Amelogenina.

Como pode-se observar, ambos os kits analisam 15 STRs mais a amelogenina (marcador de gênero), sendo que 13 destes são comuns entre eles. Neste estudo, analisou-se os marcadores do kit PowerPlex 16 System (Promega) na população de Araraquara.

1.3. Frequências alélicas e parâmetros estatísticos

Populações que se encontram em regiões diferentes ou apresentam origem distinta, podem apresentar alelos em frequências significativamente diferentes e algumas vezes únicos, alelos esses que não são compartilhados com outros grupos. Portanto, para utilização de um marcador STR na prática forense, primeiramente,

deve-se realizar a caracterização dos polimorfismos em diferentes populações a fim de se estabelecer bancos de dados referência.

Diante do exposto, muitos estudos estão sendo realizados com marcadores autossômicos em diferentes populações e diferenças significativas têm sido reportadas. No trabalho realizado por Poiares et al. (2010), as principais diferenças foram obtidas entre as regiões Sul e Norte, bem como entre os Estados Centro-Oeste e Nordeste, observando-se uma maior similaridade entre regiões geograficamente próximas.

As bases de dados populacionais contêm a distribuição alélica dos marcadores, a qual será utilizada nos cálculos estatísticos realizados em testes de identificação humana a fim de dar força e credibilidade às informações genéticas obtidas pela análise das amostras biológicas.

No exame de DNA com mãe, criança e suposto pai (trio), por exemplo, os perfis alélicos são comparados entre si para avaliar a presença ou não de vínculo biológico. Se em todos os marcadores STR criança e suposto pai compartilham um alelo (aquele que não foi herdado da mãe pela criança), para valorar tal evidência genética, realiza-se o cálculo do Índice de Paternidade, o qual consiste em uma razão de verossimilhança ou *Likelihood Ratio* (LR), definido pelo contraste de hipóteses (X e Y) mutuamente exclusivas. No caso trio, o numerador (X) é a probabilidade de se ter os genótipos observados em mãe, criança e suposto pai, dado a hipótese de que este indivíduo é o pai biológico da criança e, o denominador (Y), é a probabilidade de se ter os genótipos observados, dado a hipótese de que outro homem, escolhido ao acaso na população, é o pai biológico da criança em questão (NETO, 2010). Assim, quanto maior for o valor obtido, maior é a evidência em favor da paternidade.

Além da caracterização dos polimorfismos, diversos parâmetros estatísticos são calculados para avaliar a eficiência dos marcadores genéticos, como descrito brevemente abaixo.

Qualitativamente, um marcador é considerado polimórfico quando possui pelo menos dois alelos e seu alelo mais freqüente encontra-se em freqüência inferior a 99%. Quantitativamente, o grau de polimorfismo é comumente medido por dois parâmetros distintos, conhecidos como heterozigose e conteúdo de informação polimórfica (PIC) (SHETE; TIWARI; ELSTON, 2000).

A heterozigose de um marcador é a probabilidade de um indivíduo ser heterozigoto para o mesmo e depende do número de alelos e de sua freqüência na população. O PIC é de maior interesse para a genética médica em estudos de ligação no qual os alelos de um determinado marcador (como STR) co-segregam com alelos de uma determinada característica (geralmente uma doença). O PIC foi originalmente desenvolvido por Botstein et al. (1980) e definido como a probabilidade de que o genótipo de um marcador em uma criança permite deduzir, na ausência de *crossing-over*, quais dos dois alelos do pai afetado ela herdou. Segundo a classificação de Botstein et al. (1980), marcadores com valores de PIC superiores a 0,5 são considerados muito informativos, com valores entre 0,25 e 0,50 mediamente informativos, e com valores inferiores a 0,25, pouco informativos.

Outra propriedade avaliada é o equilíbrio de Hardy-Weinberg. Se uma população está em equilíbrio, 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 dos genótipos obtidos é possível determinar as freqüências alélicas na população (BEIGUELMAN, 1996).

O Poder de Discriminação (PD) corresponde à probabilidade de se escolher dois indivíduos ao acaso na população e estes apresentarem genótipos diferentes para um determinado marcador. Já a Probabilidade de Identidade (PI) ou *Matching Probability* (MP) é o contrário do PD, ou seja, é a probabilidade de se pegar duas pessoas e elas serem idênticas para um mesmo marcador. Assim, PD+MP somam 1 (BUTLER, 2005).

Poder de Exclusão é a probabilidade, dado o genótipo da mãe e da criança, de se excluir um homem como verdadeiro pai biológico da criança quando ele realmente não é o pai biológico desta.

Índice de Paternidade Típico (TPI) é a probabilidade de que os alelos obtidos em uma criança suportem a hipótese de que o homem testado é o verdadeiro pai biológico da criança e não outro indivíduo escolhido ao acaso na população (BUTLER, 2005).

2. OBJETIVOS

- Determinar as frequências alélicas para 15 STRs autossômicos na população de Araraquara - SP, Brasil;
- Calcular os parâmetros estatísticos para avaliação dos 15 STRs autossômicos na aplicação forense e análise de vínculo biológico;
- Correlacionar os dados genéticos obtidos com a história de formação da população de Araraquara- SP, Brasil.

3. MATERIAL E MÉTODOS

3.1. Obtenção das amostras biológicas:

Amostras de sangue de 403 (184 mulheres e 219 homens) indivíduos não aparentados, que vivem na região de Araraquara foram coletadas no Laboratório de Investigação de Paternidade da Faculdade de Ciências Farmacêuticas - UNESP - Campus de Araraquara, utilizando-se FTA MicroCard (Whatman), que é um papel absorvente a base de celulose que contém quatro substâncias químicas que protegem as moléculas de DNA da degradação da nuclease e preserva o papel da ação bacteriana (Burgoyne, 1996). Como resultado, o papel é estável à temperatura ambiente durante período de vários anos. As células são lisadas em contato com o papel e o DNA das células brancas do sangue, que é imobilizado dentro da matriz do papel, pode ser purificado por lavagem que remove inibidores de PCR e grupos hemes.

Todos os doadores participantes desta pesquisa foram esclarecidos sobre a finalidade do projeto e assinaram um Termo de Consentimento no qual declararam estar de acordo com os objetivos da pesquisa para participar como voluntário.

3.2. Extração de DNA

Nesta técnica, foram utilizados dois discos (medindo 1,2 mm cada um) de papel FTA, contendo amostra de sangue e, a estes, adicionou-se 200µL de solução MGM reagente (Promega), seguindo de três lavagens com 100µL desta solução. Posteriormente, os discos foram tratados, por duas vezes, com solução Tri-EDTA,

secos e armazenados à temperatura ambiente. Estes discos foram adicionados diretamente no mix preparado para a reação de amplificação.

3.3. Reação de amplificação

A reação de PCR foi realizada em todas as amostras de DNA oriundas da extração, com o emprego dos sistemas marcadores contidos no kit PowerPlex[®]16 (Promega) que avalia 15 loci STR (CSF1PO, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, FGA, THO1, TPOX, vWA, Penta E e Penta D). Utilizou-se para amplificação o termociclador GeneAmp PCR System 9700 (Applied Biosystems, Foster City, CA, USA).

3.4. Sequenciamento

A separação e detecção dos produtos da amplificação foram realizados pelo ABI PRISM 377 DNA Sequencer (Applied Biosystems, Foster City, CA, USA) e a genotipagem foi realizada por comparação com o Ladder incluso no kit, utilizando software 2.1 GeneScan (Applied Biosystems, Foster City, CA, USA).

3.5. Seleção de amostras

Os perfis genotipados na realização dos exames de investigação de vínculo biológico, compreendeu a análise de 15 marcadores STR autossômicos (CSF1PO, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, FGA, THO1, TPOX, vWA, Penta E e Penta D) incluídos no kit PowerPlex 16 System

(Promega). Para o desenvolvimento deste projeto, os perfis foram selecionados como segue: nos casos trio (com mãe, criança e suposto pai) utilizou-se o perfil da mãe e do suposto pai; para os casos duo (com ausência da mãe) foram selecionados o perfil da criança e suposto pai quando ambos não estavam relacionados ou apenas do suposto pai na presença de vínculo biológico entre eles; para os casos de reconstrução (ausência do suposto pai), sempre foram selecionados o perfil da mãe da criança questionada e de um dos supostos parentes de vínculo paterno.

3.6. Análise dos dados

Para os 15 STRs autossômicos, com auxílio do programa Arlequin 3.1. (EXCOFFIER; LAVAL; SCHNEIDER, 2005) foram determinadas as frequências alélicas pelo método de contagem direta, avaliado o Equilíbrio de Hardy – Weinberg pelo teste exato e determinada a Heterozigose esperada e observada. A frequência alélica mínima (FAM) foi obtida pela fórmula $5/2n$, onde n corresponde ao número de indivíduos analisados. Parâmetros estatísticos de interesse forense, como Poder de Discriminação (PD), Probabilidade de Correspondência (*Matching Probability* - MP) e Conteúdo de Informação Polimórfica (PIC), bem como de interesse em análises de vínculo biológico, como Poder de Exclusão e Índice de Paternidade Típico (TPI), foram calculados com o programa PowerStats 12 – Promega (TEREBA, 1999).

As frequências alélicas obtidas para 13 marcadores STRs (CSF1PO, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, FGA, TH01, TPOX e vWA) na população de Araraquara-SP foram comparadas àquelas publicadas na literatura para outras populações, a saber: Brasil - região

sudeste (Minas Gerais (DEL CASTILHO et al., 2009), Rio de Janeiro (RODRIGUES et al., 2009), São Paulo (FRIDMAN et al., 2008) e região sudeste como um todo (POIARES et al., 2010)) e ameríndios (Xavante e Zoró (HUTZ et al., 2002); Guarani e Caingang (KOHLRAUSCH et al., 2005)); Europa – Itália (BRISIGHELLI et al., 2009), Espanha (CAMACHO et al., 2007), Norte de Portugal (AMORIM et al., 2006) e Portugal Central (LOPES et al., 2009); África - Moçambique (ALVES et al., 2004), Guiné Equatorial (ALVES et al., 2005), Cabinda - Angola (BELEZA et al., 2004), Angola (MELO et al., 2010) e África do Sul (KIDO et al., 2007). Para tanto, utilizou-se o método de distância genética de Reynolds com auxílio do pacote de programas PHYLIP 3.68 (FELSENSTEIN, 1989). A matrix de distância genética obtida foi visualizada em uma árvore baseada no método de agrupamento UPGMA (*Unweighted Pairgroup Method With Arithmetic Mean*) com auxílio do programa FigTree 1.3.1, acessível em <http://tree.bio.ed.ac.uk/software/figtree/>.

Estimativas de mistura genética foram determinadas para o mesmo conjunto de 13 loci, em comparação com as populações já mencionadas, através do método de identidade gênica fornecido no programa ADMIX95 (<http://www.genetica.fmed.edu.uy>).

Por fim, os perfis genotípicos dos indivíduos de Araraquara-SP, baseados nos 15 STRs autossômicos, foram inseridos no programa PopAffiliator para calcular a afiliação populacional individual (PEREIRA et al., 2011).

4. RESULTADOS E DISCUSSÃO

As frequências alélicas e os parâmetros de interesse forense e em análise de vínculo biológico determinados para 15 STRs autossômicos na população de Araraquara-SP estão apresentados nas Tabelas 1 e 2, respectivamente.

Tabela 2: Parâmetros estatísticos de 15 STRs autossômicos na população de Araraquara - SP, Brasil.

Parâmetros	D3S1358	TH01	D21S11	D18S51	PENTAE	D5S818	D13S317	D7S820	D16S539	CSF1PO	PENTAD	VWA	D8S1179	TPOX	FGA
Ho	0,7370	0,8213	0,8313	0,8685	0,9325	0,7469	0,7767	0,7916	0,8015	0,7246	0,8809	0,8313	0,8412	0,6743	0,8586
He	0,7812	0,7967	0,8429	0,8761	0,8961	0,7309	0,7854	0,8163	0,7822	0,7343	0,8563	0,8124	0,8204	0,6884	0,8750
P	0,2082	0,2662	0,1017	0,7607	0,9195	0,3651	0,0488	0,7269	0,0261	0,7105	0,6538	0,5581	0,4566	0,5978	0,0156
PD	0,9208	0,9231	0,9535	0,9699	0,9780	0,8783	0,9231	0,9399	0,9160	0,8827	0,9585	0,9347	0,9419	0,8621	0,9676
PE	0,4877	0,6392	0,6583	0,7316	0,8622	0,5045	0,5565	0,5835	0,6018	0,4713	0,7566	0,6535	0,6776	0,3896	0,7118
MP	0,0792	0,0769	0,0465	0,0301	0,0220	0,1217	0,0769	0,0601	0,0840	0,1173	0,0415	0,0653	0,0581	0,1379	0,0324
PIC	0,7458	0,7635	0,8233	0,8618	0,8863	0,6843	0,7548	0,7889	0,7480	0,6857	0,8382	0,7868	0,7956	0,6450	0,8603
TPI	1,9009	2,7986	2,9632	3,8019	7,4074	1,9755	2,2389	2,3988	2,5188	1,8318	4,1979	2,9203	3,1484	1,5352	3,5351

Ho, heterozigose observada; He, heterozigose esperada; P, valor de p para o teste de equilíbrio de Hardy-Weinberg; PD, poder de discriminação; PE, poder de exclusão; MP, probabilidade de correspondência; PIC, conteúdo de informação polimórfica; TPI, índice de paternidade típico.

Analisando-se a Tabela 1, pode-se observar uma grande variação alélica, com marcadores apresentando de 6 (TH01) a 19 alelos (D18S51 e FGA) e um total de 175 alelos para o sistema de 15 STRs. Em relação aos alelos reportados por Poiars et al. (2010) para a região sudeste em seu estudo realizado com 12.886 indivíduos das 5 regiões brasileiras (10,32% destes pertencentes à região sudeste), alguns alelos não foram encontrados em Araraquara: 11 (D3S1358); 12 (VWA); 25, 33, 34 e 37 (D21S11); 12.2, 14.2, 16.2, 18.2, 19.2, 21.2 e 27 (D18S51); 6 e 16 (D5S818); 7 (D13S317); 6 (D16S539); 5 e 11 (TH01); 5 (TPOX); 6 (CSF1PO); 16, 17, 21.2, 23.2, 30, 30.2, 31.2, 44.2 e 45.2 (FGA). Porém, outros foram relatados para Araraquara e não para a região sudeste, tais como: alelo 21 (D3S1358); 15 (CSF1PO); 18 (D8S1179); 19.2, 20.2, 24.3 e 25.2 (FGA). Apenas para o marcador D7S820 os mesmos alelos foram compartilhados em ambas as populações, reforçando a necessidade de se utilizar bancos de dados específicos nas interpretações de análises com DNA.

Desvios do equilíbrio de Hardy-Weinberg foram observados para os marcadores D13S317, D16S539 e FGA (Tabela 2). Tal fato não pode ser justificado pela presença de alelos nulos, uma vez que o percentual de homozigotos nestes marcadores não é excessivo (inferior a 2%). Para o marcador FGA, tal fato pode ser explicado pela presença de alguns alelos raros, mas são necessários mais dados para fornecer uma melhor resposta. Ao aplicar a correção de Bonferroni para múltiplas análises (nível de significância de 0,0033), estes desvios não são considerados, em acordo com a maioria dos estudos realizados com STRs autossômicos em genética populacional (FRIDMAN et al., 2008; POIARES et al., 2010). Assim, se por um lado a correção de Bonferroni diminui o limiar do valor de p, tornando o teste menos sensível para detecção do desvio, não identificando a

ocorrência de possíveis eventos biologicamente importantes; por outro lado, ele evita que associações não significantes sejam consideradas.

Os parâmetros determinados apresentaram valores elevados (Tabela 2), sendo os marcadores mais polimórficos o Penta E, seguido por D18S51 e FGA. O conteúdo de informação polimórfica (PIC) variou de 0,6450 (TPOX) para 0,8863 (Penta E) e o poder de discriminação (PD) variou entre 0,8621 (TPOX) e 0,9780 (Penta E). O poder combinado de discriminação e exclusão para os 15 loci STR foram 0,999999999999999999 e 0,999999785, respectivamente. Estes resultados foram comparados com aqueles reportados para todos os Estados brasileiros (POIARES et al., 2010) e os marcadores D8S1179, D7S820, D3S1358, VWA, CSF1PO e TPOX apresentaram PD e valores PIC acima da média nacional, enquanto o TH01, D21S11, D18S51, D5S818, D13S317, D16S539 e FGA apresentaram valores inferiores a média nacional. Os altos índices obtidos revelam que este sistema de 15 STRs é uma poderosa ferramenta para identificação humana na população de Araraquara-SP.

A relação entre a distribuição alélica de 13 STRs em diversas populações pode ser visualizada na árvore filogenética (Figura 3).

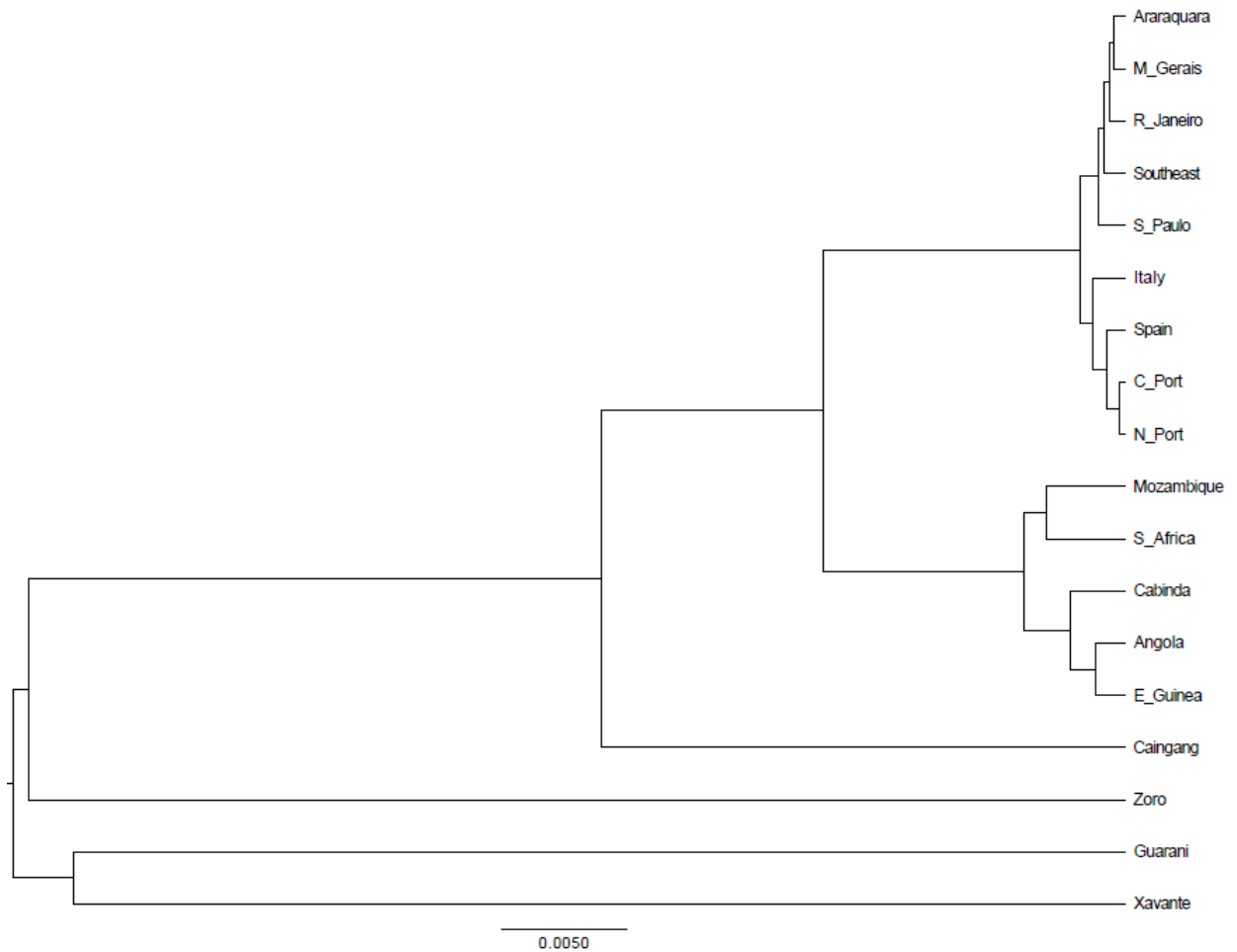


Figura 3: Árvore filogenética UPGMA baseada na distância genética de Reynolds.

A análise da Figura 3 permite a identificação de quatro grupos distintos: ameríndios, africanos, europeus e as populações da região sudeste do Brasil. Araraquara está incorporada dentro deste último grupo, o qual está próximo ao grupo europeu, mas distante das populações africana e ameríndia.

As contribuições parentais na população de Araraquara foram estimadas e correspondem a $0,76 \pm 0,0068$ de origem europeia, $0,18 \pm 0,0041$ de origem africana e $0,06 \pm 0,0042$ de origem ameríndia. As proporções encontradas estão de acordo com a proximidade entre os grupos brasileiro e europeu observado na árvore UPGMA (Figura 3).

Por fim, o perfil genotípico de cada indivíduo foi inserido no programa PopAffiliator para identificar a filiação populacional de cada sujeito. Dos 403 perfis, 280 apresentaram maior probabilidade de filiação com o grupo eurásia (200 perfis acima de 95%; 239 perfis acima de 90% e 280 perfis acima de 80%), 24 com o grupo sub-saara africano (19 perfis acima de 95%; 21 perfis acima de 90% e 24 perfis acima de 80%) e 16 perfis com o grupo asiático (6 perfis acima de 95%, 11 perfis acima de 90% e 16 perfis acima de 80%). É possível que esses 16 perfis reflitam uma contribuição ameríndia, uma vez que os nativos americanos foram os primeiros habitantes das Américas, provavelmente oriundos da imigração asiática para a América, e tiveram uma contribuição em nossa população, como descrito por alguns autores (GODINHO et al., 2008; FERREIRA et al., 2006). No entanto, também podem refletir uma recente influência asiática, já que os japoneses, embora em menor escala, imigraram para Araraquara no século XX. Como poucos dados de populações nativas americanas estão adicionados em tais bancos, em comparação com outros grupos, devemos lidar com esses dados com cautela. Os perfis remanescentes obtiveram dúbia afiliação: em 57, 16 e 10 perfis, respectivamente o grupo eurásia, asiático e sub-saara africano predominou, mas a probabilidade em todos estes casos foi inferior a 80%. Isso é de se esperar, dada a miscigenação da população brasileira. Estes resultados mostram uma alta filiação da população de Araraquara com o grupo eurásia e são consistentes com a árvore UPGMA e as estimativas de mistura genética relatadas acima.

Os resultados acima mencionados são consistentes com a história de colonização de Araraquara. A região de Araraquara e distritos vizinhos, anteriormente conhecida por "Campos de Araraquara", foi originalmente habitada por indígenas, no entanto, os dados sobre estes grupos são escassos, pois a região

estava muito isolada e os contatos iniciais não foram registrados (CORREA, 2008). A partir da década de 1790, com a abertura de rotas para as minas de ouro de Cuiabá e Goiás, a região começou a ser ocupada por imigrantes. O primeiro grande grupo de colonos foram provenientes das antigas zonas açucareiras da província de São Paulo (sudeste do Brasil), em especial de Itú, Tietê, Porto Feliz, Campinas e Piracicaba. Mais tarde, a partir do início do século XIX, a região recebeu um segundo grande contingente de imigrantes de Minas Gerais (sudeste do Brasil), principalmente do Vale do Rio das Mortes (BENINCASA, 2003). Além disso, desde o início da colonização, os africanos foram trazidos para a região como mão de obra escrava. No entanto, com o crescimento em tamanho e número de plantações de café, este tipo de trabalho tornou-se insuficiente e isso, juntamente com as dificuldades envolvidas na obtenção e manutenção dos escravos, levou a um grande fluxo de imigrantes europeus (Espanhol, Português e, sobretudo, italianos), que começaram a chegar em grande número na década de 1880, constituindo a terceira grande onda de migração, a qual estabeleceu, em comparação aos ameríndios e africanos, uma maior proximidade genética com a população estudada.

5. CONCLUSÕES

- As frequências alélicas foram determinadas e pode-se observar uma grande variação alélica para os 15 STRs autossômicos na população de Araraquara-SP.
- Os elevados valores obtidos para os parâmetros estatísticos confirmam a poderosa aplicação deste sistema na prática forense e em análise de vínculo biológico na população de Araraquara-SP;
- A população de Araraquara-SP apresenta uma maior contribuição européia, seguida, em menor escala, pela contribuição africana e ameríndia.

6. REFERÊNCIAS BIBLIOGRÁFICAS

- ALVES C, GUSMÃO L, DAMASCENO A, SOARES B, AMORIM A (2004) Contribution for an African autosomic STR database (AmpF/STR Identifiler and Powerplex 16 System) and a report on genotypic variations. **Forensic Sci Int** 139:201-5.
- ALVES C, GUSMÃO L, LÓPEZ-PARRA AM, MESA MS, AMORIM A, ARROYO-PARDO E (2005) STR allelic frequencies for an African population sample (Equatorial Guinea) using AmpFISTR Identifiler and Powerplex 16 kits. **Forensic Sci Int** 148:239-42.
- AMORIM A, ALVES C, GUSMÃO L, PEREIRA L (2006) Extended Northern Portuguese database on 21 autosomal STRs in genetic identification. **Prog Forensic Genet 11 Int Cong Ser** 1288:364–6.
- BEIGUELMAN, B. **Dinâmica dos genes nas famílias e nas populações**. 2. ed. Ribeirão Preto: Sociedade Brasileira de Genética, 1996.
- BELEZA S, ALVES C, REIS F, AMORIM A, CARRACEDO A, GUSMÃO L (2004) 17 STR data (AmpF/STR Identifiler and Powerplex 16 System) from Cabinda (Angola). **Forensic Sci Int** 141:193-6.
- BENINCASA V (2003) **Velhas Fazendas: arquitetura e cotidiano nos campos de Araraquara 1830-1930**. EdUFSCAR, São Carlos.
- BONACCORSO, N. **Análise Forense de DNA**. 2004. 26f. Monografia (Concurso de Ingresso para Professor da ACADEPOL) – Academia de Polícia de São Paulo, São Paulo.

- BOTSTEIN, D.; WHITE, R.L.; SKOLMICK, H. et al (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphism. **American Journal of Human Genetics** 32:314-33.
- BRISIGHELLI F, CAPELLI C, BOSCHI I, GARAGNANI P, LAREU MV, PASCALI VL, CARRACEDO A (2009) Allele frequencies of fifteen STRs in a representative sample of the Italian population. **Forensic Sci Int: Genet** 3:e29–e30.
- BUTLER JM. **Forensic DNA typing**. 2nd ed. London: Elsevier Academic Press, 2005.
- CALABREZ MCT (1999) **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.
- CAMACHO MV, BENITO C, FIGUEIRAS AM (2007) Allelic frequencies of the 15 STR loci included in the AmpFISTR Identifier PCR Amplification Kit in an autochthonous sample from Spain. **Forensic Sci Int** 173:241-5.
- CORREA AMM (2008) **Araraquara 1720-1930: um capítulo da história do café em São Paulo**. Cultura Acadêmica, São Paulo.
- DEL CASTILLO DM, PERONE C, DE QUEIROZ AR, MOURÃO PH, DE SOUZA VASCONCELLOS L, DO NASCIMENTO MA, JANUARIO JN (2009) Populational genetic data for 15 STR markers in the Brazilian population of Minas Gerais. **Leg Med (Tokyo)** 11:45-7.
- EXCOFFIER L, LAVAL G, SCHNEIDER S (2005) Arlequin ver. 3.0: an integrated software package for population genetics data analysis. **Evol Bioinform Online** 1:47–50.

- FELSENSTEIN J (1989) PHYLIP: Phylogeny Inference Package (version 3.2). **Cladistics** 5:164–166.
- FERREIRA LB, MENDES-JUNIOR CT, WIEZEL CE, LUIZON MR, SIMÕES AL (2006) Genomic ancestry of a sample population from the state of São Paulo, Brazil. **Am J Hum Biol** 18:702-5.
- FRANÇA GV (2001) **Medicina Legal**. 6. ed., Rio de Janeiro, Guanabara-Kogan, cap. 3, p.32-63.
- FRIDMAN C, DOS SANTOS PC, KOHLER P, GARCIA CF, LOPEZ LF, MASSAD E, GATTÁS GJ (2008) Brazilian population profile of 15 STR markers. **Forensic Sci Int Genet** 2:e1-4.
- GODINHO NMO, GONTIJO CC, DINIZ MECG, FALCÃO-ALENCAR G, DALTON GC, AMORIM CEG, BARCELOS RSS, KLAUTAU-GUIMARÃES MN, OLIVEIRA SF (2008) Regional patterns of genetic admixture in South America. **Forensic Sci Int: Genet Suppl Ser** 1:329–330.
- HUTZ MH, CALLEGARI-JACQUES SM, ALMEIDA SEM, ARMBORST T, SALZANO FM (2002) Low levels of STRP variability are not universal in American Indians. **Hum Biol** 74: 791–806.
- JEFFREYS AJ, WILSON V, THEIN SL (1985) Hipervariable “minisatellite” regions in human DNA. **Nature**, v.314, p.67-73.
- JORDE LB, WATKINS WS, BAMSHAD MJ (2001) Population genomics: a bridge from evolutionary history to genetic medicine. **Human Mol. Genetics**, v.10, n.20, p.2199-2207.
- KIDO A, DOBASHI Y, FUJITANI N, HARA M, SUSUKIDA R, KIMURA H, OYA M (2007) Population data on the AmpF/STR Identifier loci in Africans and Europeans from South Africa. **Forensic Sci Int** 168:232-5.

- KOHLRAUSCH FB, CALLEGARI-JACQUES SM, TSUNETO LT, PETZL-ERLER ML, HILL K, HURTADO AM, SALZANO FM, HUTZ MH (2005) Geography influences microsatellite polymorphism diversity in Amerindians. **Am J Phys Anthropol** 126: 463–470.
- LOPES V, SERRA A, GAMERO J, SAMPAIO L, Balsa F, OLIVEIRA C, BATISTA L, CORTE-REAL F, VIEIRA DN, VIDE MC, ANJOS MJ, CARVALHO M (2009) Allelic frequency distribution of 17 STRs from Identifiler and PowerPlex-16 in Central Portugal area and the Azores archipelago. **Forensic Sci Int: Genet** 4:e1–e7.
- MELO MM, CARVALHO M, LOPES V, ANJOS MJ, SERRA A, VIEIRA DN, SEQUEIROS J, CORTE-REAL F (2010) Genetic study of 15 STRs loci of Identifiler system in Angola population. **Forensic Sci Int** 4(5):e153-7.
- NETO RSM, Curso de Genética Forense, 56º Congresso Brasileiro de Genética, Guarujá, 2010.
- PANETO GG. Análise de polimorfismos do DNA mitocondrial em indivíduos residentes na Grande São Paulo para aplicação na Identificação Humana. 2010. Tese (Doutorado em Biociências e Biotecnologia Aplicadas à Farmácia) - Faculdade de Ciências Farmacêuticas, UNESP, Araraquara.
- PEREIRA L, ALSHAMALI F, ANDREASSEN R, BALLARD R, CHANTRATITA W, CHO NS, COUDRAY C, DUGOUJON JM, ESPINOZA M, GONZÁLEZ-ANDRADE F, HADI S, IMMEL UD, MARIAN C, GONZALEZ-MARTIN A, MERTENS G, PARSON W, PERONE C, PRIETO L, TAKESHITA H, VILLALOBOS HR, ZENG Z, ZHIVOTOVSKY L, CAMACHO R, FONSECA NA (2011) PopAffiliator: online calculator for individual affiliation to a major population group based on 17 autosomal short tandem repeat genotype profile. **Int J Legal Med** 125(5):629-36.

- POIARES LA, OSORIO PS, SPANHOL FA, COLTRE SC, RODENBUSCH R, GUSMÃO L, LARGURA A, SANDRINI F, DA SILVA CMD (2010) Allele frequencies of 15 STRs in a representative sample of the Brazilian population. **Forensic Sci Int: Genet** 4:e61–e63.
- RODRIGUES EL, MACHADO FB, ARRUDA MM, DE MOURA-NETO RS, MEDINA-ACOSTA E (2009) Genetic data on 15 STR autosomal loci for a sample population of the Northern Region of the State of Rio de Janeiro, Brazil. **Forensic Sci Int Genet** 4:e25-6.
- SHETE S, TIWARI H, ELSTON RC (2000). On estimating the heterozygosity and polymorphism information content value. **Theor Popul Biol** 57:265-71.
- SNUSTAD DP, SIMMONS MJ. **Fundamentos de Genética**. 4. ed. Rio de Janeiro: Guanabara Koogan, 2008.
- SUBRAMANIAN S, MISHRA RK, SINGH L (2003) Genome-wide analysis of microsatellite repeats in humans: their abundance and density in specific genomic regions. **Genome Biology**, v.4, R13.
- TEREBA A (1999) Tools for analysis of population statistics. **Profiles in DNA** 2:14–16.
- WEEDN VW, SWARNEN SL. 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.
- (<http://dna-view.com/exclusn.htm>).