

INSTITUTO OSWALDO CRUZ

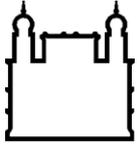
Mestrado em Biologia Celular e Molecular

FREQUÊNCIAS ALÉLICAS, PARÂMETROS ESTATÍSTICOS DE
NATUREZA FORENSE E CARACTERIZAÇÃO ÉTNICA DA
POPULAÇÃO DO RIO DE JANEIRO, ATRAVÉS DE
POLIMORFISMOS DO TIPO STR

TATIANA HESSAB DE CASTRO ARANHA

Rio de Janeiro

2012



Ministério da Saúde

FIOCRUZ
Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ
Pós-Graduação em Biologia Celular e Molecular

TATIANA HESSAB DE CASTRO ARANHA

Frequências alélicas, parâmetros estatísticos de natureza forense e caracterização étnica da população do Rio de Janeiro, através de polimorfismos do tipo STR

Dissertação apresentada ao Instituto Oswaldo Cruz como parte dos requisitos para obtenção do título de Mestre em Biologia Celular e Molecular

Orientador: Dr. Pedro Hernan Cabello

RIO DE JANEIRO

2012

Ficha catalográfica elaborada pela
Biblioteca de Ciências Biomédicas/ ICICT / FIOCRUZ - RJ

A662 Aranha, Tatiana Hessab de Castro

Frequências alélicas, parâmetros estatísticos de natureza forense e caracterização étnica da população do Rio de Janeiro, através de polimorfismos do tipo STR. / Tatiana Hessab de Castro Aranha. – Rio de Janeiro, 2012.

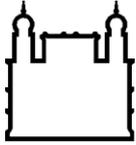
ix, 91 f. : il. ; 30 cm.

Dissertação (mestrado) – Instituto Oswaldo Cruz, Pós-Graduação em Biologia Celular e Molecular, 2012.

Bibliografia: f. 64-75

1. Genética forense. 2. Polimorfismos. 3. STR. 4. Identificação humana. 5. Rio de Janeiro. I. Título.

CDD 614.1



Ministério da Saúde

FIOCRUZ
Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ
Pós-Graduação em Biologia Celular e Molecular

AUTORA: TATIANA HESSAB DE CASTRO ARANHA

**FREQUÊNCIAS ALÉLICAS, PARÂMETROS ESTATÍSTICOS DE NATUREZA FORENSE E
CARACTERIZAÇÃO ÉTNICA DA POPULAÇÃO DO RIO DE JANEIRO, ATRAVÉS DE
POLIMORFISMOS DO TIPO STR**

ORIENTADOR: Dr. Pedro Hernan Cabello

Aprovada em: 20/08/2012

EXAMINADORES:

Prof^a. Dr^a. Maria da Graça Figueiredo Pereira Dutra - **Presidente**
Prof^a. Dr^a. Selma Lílian Sallenave Sales
Prof^a. Dr^a. Alejandra Saori Araki
Prof^a. Dr^a. Roberta Kuan Tchuen de Mello Loh
Prof^a. Dr^a. Giselda Maria Kalil de Cabello

Rio de Janeiro, 20 de agosto de 2012.

Aos meus pais, à minha irmã Thais
e ao meu marido Renato,
meus grandes amores.

Agradecimentos

Ao Dr. Pedro Cabello, obrigada pela orientação, pela oportunidade de fazer o mestrado na Fiocruz, participar do seu grupo e me desenvolver. Obrigada também pela compreensão e paciência.

Aos meus pais, agradeço o amor, o carinho e o orgulho que sentem de mim e sempre foi fonte de estímulo. Obrigada ainda pela educação a que me deram acesso. Sem isso, eu não estaria aqui. Obrigada, amo vocês!

À minha irmã e melhor amiga, Thais, obrigada pelo apoio, pelas vezes em que dormiu com as luzes acesas para eu trabalhar, pelas conversas, o carinho e amizade. Obrigada por me ter como exemplo; isso é, certamente, uma honra para mim. Vou estar sempre ao seu lado, te amo, querida!

Aos amigos dos mais diversos lugares Alexandre Sardinha, Eliza França, Fabricio e Raquel Rabello, Gabriel Malaquias, Glaucio Domingues, Luanda Oliveira, Priscila Hessab, Renata Santore, Tayane Lima e Thales Vieira, obrigada por fazerem parte da minha vida.

Aos meus colegas de trabalho do IPPGF, obrigada pela companhia. Em especial, aos Peritos do laboratório, agradeço a amizade, a colaboração, o incentivo para a realização do projeto de mestrado e por contribuírem para tornar o trabalho tão especial.

Aos estudantes das turmas 804 e 3004, do Colégio Brigadeiro Newton Braga, agradeço os ensinamentos nos campos pessoal e profissional. É um prazer ser professora de vocês.

À equipe do LGH, obrigada pela companhia e o aprendizado, nesse período de mestrado em que estive no laboratório.

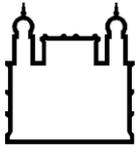
Por fim, ao meu marido Renato, agradeço o companheirismo e a parceria de quase oito anos. Pelo apoio não somente neste momento, mas na vida, nesses anos em que crescemos juntos. Pelo amor e carinho que tornam a vida muito mais colorida. Por ser você, exatamente assim como é, a coisa mais linda e apaixonante que já me aconteceu.

“Nenhuma grande vitória é possível sem
que tenha sido precedida de pequenas
vitórias sobre nós mesmos.”

Sumário

Resumo	1
Abstract.....	3
1. Introdução.....	5
1.1 A evolução dos métodos de identificação humana.....	5
1.2 Polimorfismos genéticos e a identificação humana.....	8
1.3 Marcadores STR estudados	12
1.3.1 CSF1PO.....	12
1.3.2 FGA	13
1.3.3 TH01	13
1.3.4 TPOX.....	14
1.3.5 vWA	15
1.3.6 D3S1358	15
1.3.7 D5S818	16
1.3.8 D7S820	16
1.3.9 D8S1179	17
1.3.10 D13S317	17
1.3.11 D16S539	18
1.3.12 D18S51	18
1.3.13 D21S11	19
1.3.14 D2S1338	19
1.3.15 D19S433	20
1.3.16 Penta E.....	20
1.3.17 Penta D	20
1.4 Análise estatística na identificação humana por DNA	21
1.5 Grupos étnicos que influenciaram a composição genética atual	22
2. Objetivos.....	24
2.1 Objetivos Gerais	24
2.2 Objetivos Específicos	24
3. Material e Métodos.....	25
3.1 Coleta de Amostras.....	25
3.2 Extração de DNA.....	26
3.2.1 Extração de DNA em Papel FTA®	26
3.2.2 Extração de DNA por “Salting-Out”	26
3.2.3 Extração de DNA com Resina Chelex® 100	27

3.2.4	Extração de DNA pelo Método Orgânico	28
3.3	Amplificação do material genético.....	28
3.3.1	AmpF ℓ STR $\text{\textcircled{R}}$ Identifiler $\text{\textcircled{R}}$	32
3.3.2	AmpF ℓ STR $\text{\textcircled{R}}$ Identifiler $\text{\textcircled{R}}$ Plus.....	32
3.3.3	AmpF ℓ STR $\text{\textcircled{R}}$ MiniFiler TM	33
3.3.4	PowerPlex $\text{\textcircled{R}}$ 16 System	33
3.3.5	PowerPlex $\text{\textcircled{R}}$ 16 HS System.....	34
3.4	Eletroforese Capilar.....	34
3.5	Análise dos Dados	36
3.6	Aspectos Éticos	38
4.	Resultados.....	39
5.	Discussão.....	48
5.1	Parâmetros estatísticos no conjunto total de loci.....	48
5.2	Análise da diferenciação genética	49
5.2	Frequências Alélicas.....	50
5.2.1	CSF1PO.....	51
5.2.2	FGA	52
5.2.3	TH01	52
5.2.4	TPOX.....	53
5.2.5	vWA	53
5.2.6	D3S1358	54
5.2.7	D5S818	55
5.2.8	D7S820	56
5.2.9	D8S1179	56
5.2.10	D13S317	57
5.2.11	D16S539	58
5.2.12	D18S51	59
5.2.13	D21S11	59
5.2.14	D2S1338	60
5.2.15	D19S433	60
5.2.16	Penta E.....	60
5.2.17	Penta D	61
6.	Conclusões.....	62
7.	Referências	64
8.	Anexos.....	76



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ

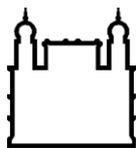
Frequências alélicas, parâmetros estatísticos de natureza forense e caracterização étnica da população do Rio de Janeiro, através de polimorfismos do tipo STR

RESUMO

DISSERTAÇÃO DE MESTRADO

A Genética Forense é um ramo da Ciência que está se expandindo a cada dia. O aprimoramento das técnicas moleculares vem trazendo uma constante renovação e relevantes discussões para a consolidação dessa área de estudo. Na identificação humana pelo perfil de DNA, além da comparação de perfis genéticos, usam-se métodos estatísticos a fim de estimar a probabilidade de um determinado perfil ser encontrado apenas a partir do doador em comparação ao de outro indivíduo qualquer da população. Essa análise é baseada em frequências obtidas de estudos feitos previamente na população. Para a realização dos casos criminais de identificação genética, o estado do Rio de Janeiro conta com o Instituto de Pesquisa e Perícias em Genética Forense (IPPGF), que é o órgão Pericial da Polícia Civil. Esse órgão recebe amostras de todo o estado, agregando um estoque representativo de amostras da população Rio de Janeiro. Dessa forma, um estudo empregando amostras do IPPGF é de grande importância na solução de casos de identificação genética criminal. Nesse contexto, o presente trabalho tem por finalidade caracterizar a população de Rio de Janeiro através das frequências alélicas de marcadores utilizados na genética forense a partir da análise de amostras de 505 indivíduos não aparentados, que doaram material biológico para confronto genético, no Instituto de Pesquisa e Perícias em Genética Forense, no período de 2008 a 2011. No presente estudo foram empregados 17 marcadores autossômicos do tipo STR

(*short tandem repeats*) e a genotipagem foi baseada nos kits de identificação humana Power Plex® 16 System / Power Plex® 16 HS System, da empresa *Promega*, e/ou AmpFISTRIdentifiler™ / AmpFISTRIdentifiler™ Plus e AmpFSTR® MiniFiler™, da empresa *Applied Biosystems*. Com base na análise dos perfis genéticos, foram calculadas as frequências alélicas dos marcadores genéticos, os parâmetros forenses, o coeficiente de endogamia (F_{IS}), realizado o teste exato do Equilíbrio Hardy-Weinberg. Nos 17 marcadores estudados, foi detectado um total de 215 alelos distintos. O *locus* D21S11, foi o mais polimórfico (25 alelos), seguido do FGA (22 alelos). O alelo 8 do TPOX apresentou a maior frequência observada. Além disso, esse *locus* teve o menor Poder de Discriminação ($PD = 0,879$), Poder de Exclusão ($PE = 0,419$) e Conteúdo de Informação Polimórfica ($PIC = 0,676$) e a maior Probabilidade de Coincidência Aleatória ($PM = 0,121$). O marcador genético Penta E apresentou o maior Poder de Discriminação ($PD = 0,983$), Poder de Exclusão ($PE = 0,761$) e Conteúdo de Informação Polimórfica ($PIC = 0,905$) e menor Probabilidade de Coincidência Aleatória ($PM = 0,017$). Foram realizados ainda análises a fim de estudar a composição étnica da população. A árvore filogenética de Neighbor-Joining, com dados de diferentes regiões brasileiras, possibilitou detectar uma semelhança entre população do Rio de Janeiro e a população do Nordeste. No entanto, não foram observados valores estatisticamente significativos na análise do índice de fixação (F_{ST}) par a par entre as populações, indicando ausência de estruturação, importante característica no caso de marcadores genéticos utilizados em testes de identificação humana. O gráfico “Triangle Plot” populacional, gerado no *software* Structure, sugere que, em média, os indivíduos não apresentam contribuição maior de determinado grupo étnico em comparação com os demais, podendo a população do Rio de Janeiro ser formada por contribuição similar dos grupos étnicos principais. De maneira, geral observamos que os alelos de maior frequência nos diversos genes estudados também se distribuem amplamente nas diversas populações mundiais.



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ

Allele frequencies, forensic parameters and ethnic composition of the Rio de Janeiro population, based on STR markers

ABSTRACT

DISSERTAÇÃO DE MESTRADO

Improvement of molecular techniques has brought a constant renewal and relevant discussion to Forensic Genetics development. In human genetic identification, after comparing genetic profiles, statistical methods are applied to estimate the probability of a given profile be found only from the donor in comparison to another individual of the population. This analysis is based on frequencies obtained from previous studies done in the population. In the state of Rio de Janeiro, the IPPGF (Instituto de Pesquisa e Perícias em Genética Forense) is responsible for performing genetic identification in criminal cases. This Institute receives samples of the whole state, joining a representative stock of Rio de Janeiro samples. In this way, a research using samples of IPPGF is very important to solve criminal case of genetic identification. Therefore, this work aims to develop an allele frequencies database from samples of 505 unrelated individuals who, from 2008 to 2011, donated biological material to IPPGF. For this purpose, we have genotyped 17 autosomal STRs (short tandem repeats) using the human identification kits PowerPlex® 16 System / PowerPlex® 16 HS System (Promega Corporation) and AmpFℓSTR™ Identifiler / AmpFℓSTR™ Identifiler Plus and AmpFℓSTR® MiniFiler™ (Applied Biosystems). Based on the genetic profiles analysis, we have calculated

the STR allele frequencies, tested for Hardy-Weinberg equilibrium, inbreeding coefficient (F_{IS}) and forensic parameters. In these 17 genetic markers, we detected a total of 215 distinct alleles. The *locus* D21S11, was the most polymorphic (25 alleles), followed by FGA (22 alleles). The TPOX allele 8 was the most frequently observed. Furthermore, this *locus* had the lowest Power of Discrimination ($PD = 0.879$), Power of Exclusion ($PE = 0.419$) e Polymorphism Information Content ($PIC = 0.676$) and the highest Probability of a Random Match ($PM = 0.121$). The marker Penta E had the highest Power of Discrimination ($PD = 0.983$), Power of Exclusion ($PE = 0.761$) and Polymorphism Information Content ($PIC = 0.905$) and the lowest Probability of a Random Match ($PM = 0.017$). Further analyzes were performed to study the ethnic composition of the population. The Neighbor-Joining phylogenetic tree (data from different Brazilian regions) showed a similarity between the Rio de Janeiro population and the Northeast Brazilian population. However, none statistical significance was observed in pairwise F_{ST} analysis among populations, indicating the absence of structuration, an important trait for of genetic markers used in human identification tests. The "Triangle Plot" graphic, designed by the Structure software, suggests that, on average, individuals have no greater contribution of a particular ethnic group and the Rio de Janeiro population may be formed by the similar contribution of the main ethnic groups. By comparing the allele frequencies obtained, we have detected that the alleles of higher frequency, in the present study, showed in general, wide distribution in various populations.

1. Introdução

1.1 A evolução dos métodos de identificação humana

A identificação humana é um processo que associa dados obtidos com um indivíduo particular da população (Clarke, 1994). No contexto criminal, considera-se que o primeiro método científico de identificação foi introduzido por Alphonse Bertillon, oficial de polícia reconhecido como criador da antropometria. Em 1879, ele elaborou um sistema de identificação baseado em caracteres antropométricos como tamanho da cabeça, comprimento dos dedos e demais traços morfológicos e cromáticos (James e Nordby, 2005). Atualmente, este sistema é classificado como não conclusivo (Espíndula, 2006, *apud* Garrido, 2009).

Já as impressões digitais e palmares, consideradas conclusivas, tiveram seus primeiros registros na China, ao menos 2000 atrás, indicando seu emprego como marcas de autenticidade. A partir do século XVII, grupos de cientistas contribuíram para o reconhecimento desse método e o desenvolvimento da papiloscopia (Espíndula, 2006, *apud* Garrido, 2009).

No hemisfério ocidental, o argentino Juan Vucetich é um pioneiro da área. Como funcionário do departamento de polícia de La Plata, o pesquisador se convenceu do valor criminal da papiloscopia, escrevendo um livro sobre o assunto. A partir de 1896, após a publicação do livro, a polícia Argentina substituiu o método de Bertillon pelas impressões digitais, sendo o primeiro caso de sua utilização registrado em 1902 (James e Nordby, 2005).

Com o tempo e a evolução das técnicas de análise, as abordagens genéticas foram implantadas na identificação humana. A descrição do sistema ABO, por Karl Landsteiner, é considerada o primeiro passo para o desenvolvimento da Genética Forense. Landsteiner, no

início do século XX, observou que os indivíduos poderiam ser classificados em diferentes grupos, baseados em seu tipo sanguíneo. Essa determinação é bastante relevante quando detectada a exclusão, no entanto, é pouco informativa nos resultados de inclusão, especialmente, no caso do grupo O, que representa 40% da população (Butler, 2005).

A evolução da Biologia Molecular gerou, nas décadas de 60 e 70, o início da análise de sequências de DNA, com a utilização de enzimas de restrição, sequenciamento de Sanger e *Southern Blotting* (Goodwin *et al.*, 2007). Seguindo esse panorama, Wyman e White (1980), publicaram artigo registrando a primeira análise de *locus* altamente polimórfico, utilizando polimorfismos de tamanho de fragmentos de restrição (RFLP).

No entanto, o grande passo para a formulação da Genética Forense, como é conhecida atualmente, ocorreu no ano de 1985, quando Alec Jeffreys desenvolveu a técnica de identificação por minissatélites (VNTR – número variável de repetições em tandem). Ele notou que determinadas regiões do DNA contêm sequências repetitivas e que o número de repetições varia de um indivíduo para o outro (Jeffreys *et al.*, 1985a).

O método de Jeffreys utilizado para a análise de VNTRs consistia na digestão das moléculas de DNA por enzimas de restrição, seguida da hibridação com sondas específicas para a região de repetição. A partir desse método, é possível identificar a quantidade de repetições que o indivíduo apresenta em cada *locus* e, ao mapear diversas regiões, pode-se estabelecer um padrão único, denominado *fingerprint* de DNA (Jeffreys *et al.*, 1985a,b).

Esta técnica foi utilizada, no ano seguinte (1986), para resolver dois casos de estupro, seguido de homicídio, num vilarejo da Inglaterra. Os crimes possuíam características similares, levando a investigação a suspeitar de uma única autoria. No entanto, um homem confessou ter matado apenas uma das vítimas. Dessa forma, o material genético desse homem foi comparado com o sêmen coletado. A análise evidenciou não ser ele o autor dos crimes e a

primeira utilização forense do exame de DNA inocentou um indivíduo que poderia ser declarado culpado. O caso ainda não estava solucionado e, a fim de resolvê-lo, foi realizada uma coleta em massa nos três vilarejos próximos. Mais de quatro mil homens realizaram exame sem que se observasse coincidência com o perfil genético das amostras do local de crime. Um ano depois, um homem num bar se vangloriou de ter doado sangue no lugar de seu amigo Colin Pitchfork. A polícia local tomou ciência, interrogou e coletou sangue de Pitchfork. Seu perfil genético apresentou coincidência com o obtido a partir dos sêmens, conferindo uma resposta ao caso (Wambaugh, 1989, *apud* Butler, 2005).

Embora tenham sido registrados exemplos de sucesso, vale ressaltar que as amostras forenses, frequentemente apresentam DNA com certo nível de degradação, o que dificulta as análises e, dependendo da técnica utilizada, as inviabiliza. Com isso, o advento da técnica de PCR (reação em cadeia da polimerase), processo que possibilita a amplificação de segmentos relativamente pequenos do DNA, foi de grande importância para o desenvolvimento da Genética Forense. O método aumentou a sensibilidade da análise de DNA, podendo, assim, ser aplicado em amostras com material degradado (Saiki *et al.*, 1988; Goodwin *et al.*, 2007).

Goodwin *et al.* (2007) salientam que, além de avanços tecnológicos, o desenvolvimento da Genética influenciou significativamente para o aperfeiçoamento do controle de qualidade da Ciência Forense. O autor sugere que o caso de homicídio ocorrido nos Estados Unidos, conhecido como caso “Castro” (Patton, 1990), promoveu a padronização e controle de qualidade da Genética Forense e de outras áreas afins. As técnicas empregadas na análise de DNA foram questionadas. Embora o Tribunal tenha considerado o exame de DNA admissível, a prova técnica específica do caso foi negada, devido à ausência de aplicação de métodos cientificamente aceitáveis.

No início da década de 1990, a descrição de sequências de DNA repetitivas do tipo STRs (repetições curtas em *tandem*) ou microssatélites, tornou possível incrementar a análise de DNA forense (Garrido, 2009). Os STRs possuem unidades repetitivas constituídas por pequenas sequências de DNA, o que também auxilia as análises de material degradado. A utilização de marcadores fluorescentes a partir de 1991 e os kits de PCR multiplex, validados em diversos laboratórios, desde 1999, ampliaram ainda mais as possibilidades de análise de material genético com intuito forense (Butler, 2005).

1.2 Polimorfismos genéticos e a identificação humana

O genoma humano eucromático compreende, aproximadamente, três bilhões de nucleotídeos (*International Human Genome Sequencing Consortium*, 2004). Uma parcela menor do que 5% é constituída por região codificante, ao passo que sequências repetitivas representam ao menos 50% do genoma (*International Human Genome Sequencing Consortium*, 2001). Essas sequências repetitivas podem ser agrupadas em cinco classes distintas: repetições derivadas de transposon; pseudogenes; SSR (do inglês, *Simple Sequence Repeat*), regiões repetitivas em tandem; duplicações de segmentos do genoma de 10 a 300 Kb; blocos de sequências repetitivas como centrômeros, telômeros e braços curtos de cromossomos acrocêntricos (*International Human Genome Sequencing Consortium*, 2001).

As regiões não-codificantes do DNA estão sujeitas a menor pressão seletiva e, portanto, à maior taxa de mutação, sendo mais polimórficas e bastante úteis para a identificação humana (Regateiro, 2007). Os polimorfismos são gerados pelas mutações, que alteram as sequências de DNA. Estas podem ser identificadas como mutação gênica ou de ponto, quando há alteração em poucos pares de base adjacentes; ou mutação cromossômica, quando segmentos de um cromossomo ou cromossomos inteiros sofrem modificação (Griffiths *et al.*, 2004).

As mutações gênicas mais comuns são a substituição de nucleotídeo e a inserção/deleção de bases. As substituições ocorrem quando um par de bases substitui outro. Quando ela se dá por alteração de base de uma mesma categoria (purina por purina ou pirimidina por pirimidina), denomina-se transição. Quando ocorre com alteração de categoria química, dá-se o nome transverso (Griffiths *et al.*, 2004).

Os SNPs (polimorfismos de base única) são polimorfismos gerados pela substituição de base, resultado da mutação de um único nucleotídeo, apresentando, no genoma humano, uma ocorrência, de cerca de 12 a 16 milhões. Estima-se que eles sejam responsáveis pela maior parte da variabilidade genética individual (Regateiro, 2007). Regiões altamente polimórficas, como a região hipervariável do DNA mitocondrial, apresentam maior ocorrência destes polimorfismos. Possuem aplicação forense e para estudos populacionais (Helyar, 2011; Yamaguchi-Kabata *et al.*, 2008; Freedman *et al.*, 2004).

Os Indels são polimorfismos de inserção-deleção. Recentemente, essas mutações também estão sendo incluídas em kits de identificação humana, mas assim como os SNPs, atualmente, não constituem os principais marcadores de identificação, sendo aplicados em estudos populacionais (Santos *et al.*, 2010; Behar *et al.*, 2007).

Sequências satélites são regiões repetitivas em *tandem*, geradas por repetição sequencial da mesma unidade; elas podem ser classificadas como minissatélites/VNTRs (número variável de repetições em tandem) ou microssatélites/STRs (repetições curtas em tandem). Nos minissatélites, as unidades repetitivas possuem, usualmente, entre 9 e 100 pares de bases. Nos microssatélites, as unidades repetitivas são constituídas por sequências com 2 a 6 pares de bases (Tautz, 1993). Essas sequências constituem cerca de 3% do genoma humano (*International Human Genome Sequencing Consortium*, 2001).

Os polimorfismos do tipo STR são gerados por “deslizamentos” da DNA polimerase, no momento da replicação, associado a mecanismo de reparo deficiente (Ellegren, 2004). Esses deslizamentos originam inserções e deleções no número original de repetições, constituindo a base das diferenças alélicas.

Dentre os diversos tipos de STRs, os de repetições de tetranucleotídeos são mais comumente utilizados, por serem mais frequentes do que os pentanucleotídeos e apresentarem menor taxa de erro no momento da amplificação do que di e trinucleotídeos. Dessa forma, reduzem a ocorrência de artefatos, como os *stutters*, produtos de PCR com número de repetição menor do que o alelo verdadeiro (Butler, 2005).

Os microssatélites apresentam importante aplicação forense. Comparados aos VNTRs, possuem unidades de repetição de menor tamanho, o que auxilia a amplificação de amostras degradadas. Além disso, conforme Butler (2005), os STRs constituem a melhor solução, com grande poder de discriminação e rápida velocidade de análise.

Na identificação humana por DNA, a escolha de marcadores do tipo STR envolve o preenchimento de requisitos que os tornem apropriados para esse fim. Dentre os critérios adotados para a seleção, podemos citar ausência de associação com patologias detectadas até o momento, baixa taxa de mutação (menor do 0,3%), alto nível de independência (associação aleatória de alelos), alto poder de discriminação (maior do que 0,9) (Hares, 2012), robustez (alta frequência de resultados bem sucedidos), reprodutibilidade, baixa frequência de *stutters* e produtos de PCR variando entre 90 e 500 pares de base (Gill *et al.* 1994).

Com base em características como essas, em 1997, o FBI (*Federal Bureau of Investigation*) estabeleceu 13 marcadores moleculares do tipo STR que deveriam ser utilizados em sua base de dados (CODIS – *Combined DNA Index System*) (Butler, 2006). Atualmente, os marcadores selecionados são utilizados em diversos laboratórios de Genética

Forense, usualmente associados a dois marcadores adicionais, além do sistema de identificação de gênero amelogenina, utilizando este gene que possui cópias homólogas nos cromossomos X (106 pares de base) e Y (112 pares de base), permitindo a identificação masculina ou feminina. Mais recentemente, os STRs do CODIS estão sendo revisados e há previsão de inclusão de novos marcadores, o que aumentaria o poder de identificação da análise (Hares, 2012). Existe estudo, inclusive, que recomenda a utilização de ao menos 15 marcadores genéticos (Moroni *et al.*, 2011).

Os 13 *loci* do CODIS são CSF1PO, FGA, TH01, TPOX, vWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51 e D21S11 (Budowle *et al.*, 1998), conforme figura 1.1. Outros quatro loci também são amplamente empregados. São eles: D2S1338 e D19S433, presentes no *kit* de identificação humana AmpF/STR® Identifiler™, e Penta D e Penta E, presentes no *kit* de identificação humana PowerPlex® 16 System.

A nomenclatura utilizada para nomear esses marcadores aplica os seguintes critérios: quando estão presentes num gene, o nome do gene é utilizado na designação; se estiverem localizados fora da região de um gene, são nomeados de acordo com o cromossomo no qual estão localizados. Nesse caso, a primeira letra é o “D”, de DNA, seguido pelo número do cromossomo, posteriormente, da letra “S” (se for uma sequência de cópia única) e de um número final, que indica a ordem de caracterização daquele marcador dentro do cromossomo (Butler, 2005).

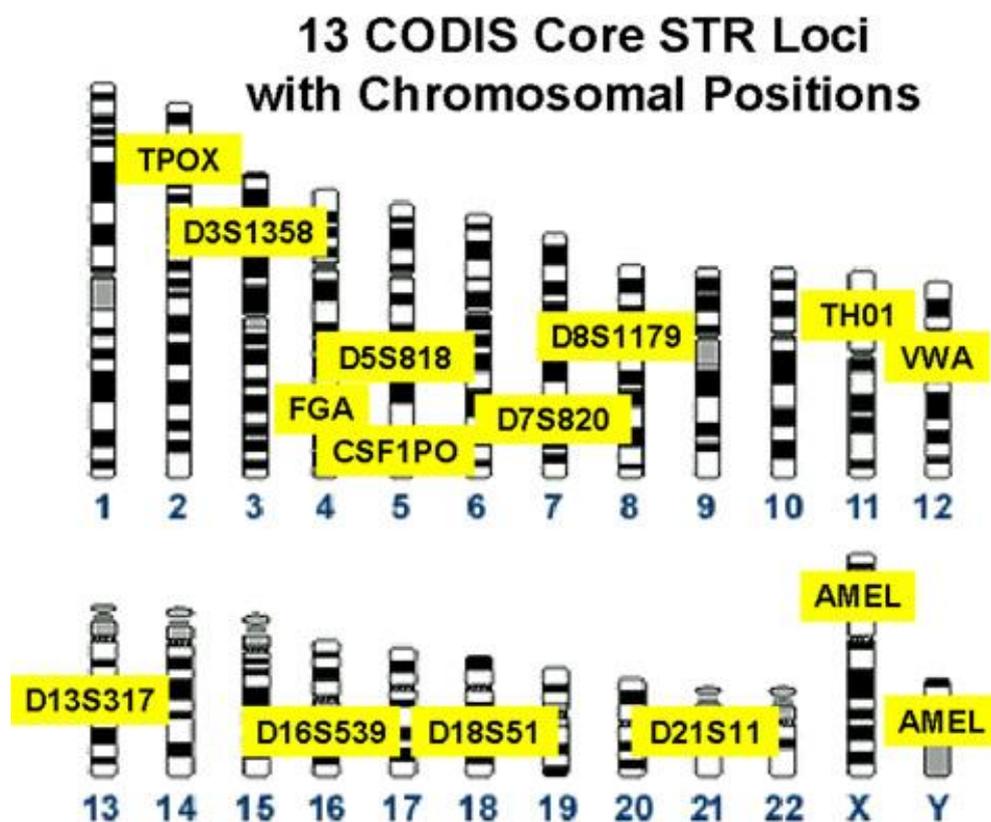


Figura 1.1: Marcadores genéticos do CODIS (*Combined DNA Index System*) (STRBase).

1.3 Marcadores STR estudados

1.3.1 CSF1PO

O marcador genético CSF1PO é um STR de repetição simples presente no sexto íntron do *c-fms* proto-oncogene para o receptor CSF-1, localizado no braço longo do cromossomo 5. Seus alelos possuem a unidade de repetição AGAT (Figura 1.2) (Hammond *et al.*, 1994), possuindo, usualmente, de 6 a 15 repetições, com a presença de alguns microvariantes (Butler, 2005). Margolis-Nunno *et al.* (2001) detectaram ainda a presença do alelo 16.

```

11821 ggagcacaca ctccagggca gtgttccaac ctgagtctgc caaggactag caggttgcta
11881 accaccctgt gtctcagttt tcctacctgt aaaatgaaga tattaacagt aactgccttc
11941 atagatagaa gatagataga ttatagatagat agatagatag atagatagat agatagatag
12001 atagatagat aggaagtact tagaacaggg tctgacacag gaaatgctgt ccaagtgtgc
12061 accaggagat agtatctgag aaggctcagt ctggcaccat gtgggttggg tgggaacctg
12121 gaggtctggag aatgggctga agatggccag tgggtgtgtgg aagagtctga gatgcagggg

```

Figura 1.2: Sequência evidenciando a região de repetição do *locus* CSF1PO (GenBank: U63963.1).

1.3.2 FGA

O *locus* FGA (também denominado FIBRA ou HUMFIBRA) é formado por repetição composta do tetranucleotídeo CTTT (Figura 1.3), flanqueado por repetições degeneradas, sendo encontrado no terceiro íntron do alfa fibrinogênio humano (braço longo do cromossomo 4). Contém de 12.2 a 51.2 repetições, totalizando 80 alelos distintos, o que o torna um dos *loci* mais polimórficos utilizados na identificação humana (Butler, 2005). Vale observar que a inserção ou deleção de dinucleotídeos gera os microvariantes x.2 (Barber *et al.*, 1996), bastante frequentes para esse marcador (Butler, 2005).

```

2761 attaggacat cttactggc attcatggaa ggctgcaggg cataacatta tccaaaagtc
2821 aaatgcccc taggttttga actcacagat taaactgtaa ccaaataaa attaggcata
2881 ttacaagct agtttctttc tttctttttt ctctttcttt ctttctttct ttctttcttt
2941 ctttctttct ttctttcttt ctttctcttt ctttctttc ttctttcttt ttttctttc
3001 aattacagac aaatcactca gcagctactt caataacat attttctgatt tcagaccgtg
3061 ataataccta caaccgagtg tcagaggatc tgagaagcag aattgaagtc ctgaagcgca

```

Figura 1.3: Sequência evidenciando a região de repetição do *locus* FGA (GenBank: M64982.1).

1.3.3 TH01

O marcador do tipo STR TH01 localiza-se no íntron 1 do gene da tirosina hidroxilase, no braço curto do cromossomo 11. A unidade de repetição do *locus* é formada pelo motivo

TCAT (Figura 1.4) (Polymeropoulos *et al.*, 1991). Existem referências ao mesmo marcador como TC11 e HUMTH01. O microvariante 9.3 é bastante comum na população caucasiana (Butler, 2005).

```
1021 gcctgcctgc ttggggagggc agccccaagg cccttcccag gctctagcag cagctcatgg
1081 tgggggggtcc tgggcaaata gggggcaaaa ttcaaagggg atctggggctc tggggtgatt
1141 cccattggcc tggttcctccc ttatttccct cattcattca ttcattcatt cattcattca
1201 ttcatcacc atggagtctg tggttccctgt gacctgcact cggaagccct gtgtacaggg
1261 gactgtgtgg gccaggctgg ataatcggga gcttttcagc ccacaggagg ggtcttcggg
1321 gcctccttgg gcaactcagaa ccttggggctc cctggcacat ttaaaatggg tttttattta
```

Figura 1.4: Sequência evidenciando a região de repetição do *locus* TH01 (GenBank: D00269.2).

1.3.4 TPOX

O marcador genético TPOX (também reportado na literatura como hTPO) é encontrado no íntron 10 do gene da tireoide peroxidase humana, próximo ao final do braço curto do cromossomo 2. Sua unidade de repetição é formada pela sequência AATG (Figura 1.5) e este é o menos polimórfico dos STRs do CODIS (Butler, 2005). O alelo mais comum desse marcador pode apresentar frequências maiores do que 60% em algumas populações, mas a baixa taxa de mutação torna-o bastante útil em determinados testes de paternidade (Butler, 2006). Conforme Hares (2012) propõe, essa é uma importante característica para a seleção de um marcador STR com intuito forense.

```
1741 ccatgttccc actggcctgt ggggtccccc atagatcgta agcccaggag gaagggctgt
1801 gtttcagggc tgtgatcact agcaccaga accgtcgact ggcacagaac aggcacttag
1861 ggaaccctca ctgaatgaat gaatgaatga atgaatgaat gaatgaatga atgaatgttt
1921 gggcaaataa acgctgacaa ggacagaagg gcctagcggg aagggaaacag gagtaagacc
1981 agcgcacagc ccgacttgtg ttcagaagac ctgggatttg acctgaggag ttcaattttg
```

Figura 1.5: Sequência evidenciando a região de repetição do *locus* TPOX (GenBank: M68651.1).

1.3.5 vWA

O vWA é uma repetição de tetranucleotídeo composto, constituído pelas unidades de repetição TCTA, TCTG e TCCA (Figura 1.6) (Walsh *et al.*, 1996), intercaladas, no íntron 40 do Fator de von Willebrand (braço curto do cromossomo 12). No mesmo íntron, são encontradas outras duas regiões de repetição. No entanto, estas não são polimórficas como a região utilizada na identificação humana (Kimpton *et al.*, 1992). São observados ainda dados na literatura referindo-se a este marcador como vWF ou VWA (Butler, 2005).

```
1561 cctataaggt agagttccca ccttccagaa gaagaaacag gtctagagga tccaagttga
1621 cttggctgag atgtgaaagc cctagtggat gataagaata atcagtatgt gacttggatt
1681 gatctatctg tctgtctgtc tgtctatcta tctatctatc tatctatcta tctatctatc
1741 tatctatcta tctatccatc tatccatcca tcctatgtat ttatcatctg tcctatctct
1801 atctaacctt tgtatctatt tatcatctat cctgtctcta tctatccttt gtatctatca
1861 tctatocctat ctctatctaa gctatatatc tatttatcat ctatcctcta tcctatctct
```

Figura 1.6: Sequência evidenciando a região de repetição do *locus* vWA (GenBank: M25858.1).

1.3.6 D3S1358

O *locus* D3S1358 está localizado no braço curto do cromossomo 3 (Butler, 2005). Ele é formado por tetranucleotídeo composto, gerado pelas repetições AGAT e AGAC (Figura 1.7).

```
77641 atcccaggta cttgggaggc tgaggcagga ggactgcttg agcccaggag tttgaggctg
77701 tagtgagcta tgattccccc actgcagtcc aatctgggtg acagagcaag accctgtctc
77761 atagatagat agatagatag atagatagat agatagatag atagatagat agatagacag
77821 atagatacat gcaagcctct gttgatttca tgagtataag agatgcccc aaaggcacag
77881 ggaatacaca ccacagaaa atagatccct gggcagaagt gggcaagtga atatggccag
77941 catgccatt ctggagcagt gccctggcag ctgcagtcct cacctgggaa tagcttttcc
```

Figura 1.7: Sequência evidenciando a região de repetição do *locus* D3S1358 (GenBank: AC099539.2).

1.3.7 D5S818

O marcador D5S818 localiza-se no braço longo do cromossomo 5. Sua unidade de repetição é formada pela sequência AGAT (Figura 1.8), apresentando alelos de 7 a 16 repetições. Ele é um dos menores marcadores presentes nos kits de identificação humana e costuma ser amplificado mesmo no caso de amostras degradadas (Butler, 2005).

```
1 tctaattaa gtggtgtccc agataatctg tactaataaa agtatatattt aatagcaagt
61 atgtgacaag ggtgattttc ctctttggta tccttatgta atattttgaa gatagataga
121 tagatagata gatagataga tagatagata gataggtaga tagaggata aataaggata
181 cagatatagn taaaatggt gtaaactgtg gctatgattg gaatcacttg gctaaaaagc
241 gctnaagcnt tcctctgnga gaggcaatta ctttttttct taggnactnc ctcancagtc
```

Figura 1.8: Sequência evidenciando a região de repetição do *locus* D5S818 (GenBank: G08446.1).

1.3.8 D7S820

O D7S820 é formado por uma repetição de tetranucleotídeo simples, constituído pela sequência GATA (Figura 1.9). É encontrado no braço longo do cromossomo 7. Os alelos microvariantes desse marcador ocorrem devido à variação do número de nucleotídeos da cauda poli T, presente 13 nucleotídeos após a unidade de repetição (Butler, 2005).

```
1 aatttttgta ttttttttag agacggggtt tcaccatggt ggtcaggctg actatggagt
61 tattttaagg ttaatatata taaagggtat gatagaacac ttgtcatagt ttagaacgaa
121 ctaacgatag atagatagat agatagatag atagatagat agatagatag atagacagat
181 tgatagtttt tttttatctc actaaatagt ctatagtaaa catttaatta ccaatatttg
241 gtgcaattct gtcaatgagg ataaatgtgg aatcgttata attcttaaga atatatattc
```

Figura 1.9: Sequência evidenciando a região de repetição do *locus* D7S820 (GenBank: G08616.1).

1.3.9 D8S1179

O marcador genético D8S1179 é classificado como um STR composto, encontrado no braço longo do cromossomo 8. Ele é inicialmente formado pela repetição do tetranucleotídeo TCTA (Figura 1.10), porém nos casos de alelos com mais de 13 repetições o *locus* é constituído pelas repetições TCTG, desde a segunda ou terceira posição a partir da extremidade 5' da região TCTA (Butler, 2005).

```
1  tggcaactta  tatgtatddd  tgtatdddcat  gtgtacattc  gtatctatct  atctatctat
61  ctatctatct  atctatctat  ctatctatct  attccccaca  gtgaaaataa  tctacaggat
121 aggtaaataa  attaaggcat  attcagcaa  tgggatacgn  tacagtgatg  aaaatgaact
181 aattatagct  acgtgaaact  atactcatgn  acacaatttg  gtaaaagaaa  ctgggaacaa
241 gaatacatac  ggtddddgnc  agctgtgcta  ttttacattc  ccaacaacaa  tgcacagggt
```

Figura 1.10: Sequência evidenciando a região de repetição do *locus* D8S1179 (GenBank: G08710.1).

1.3.10 D13S317

D13S317 é composto pela repetição do tetranucleotídeo simples TATC (Figura 1.11), presente no braço longo do cromossomo 13. Os alelos mais comuns contêm de 7 a 15 repetições, embora tenham sido observados também os alelos 5, 6 e 16 (Butler, 2005).

```
1  tgggatgggt  tgctggacat  ggtatcacag  aagtctggga  tgtggaggag  agttcatttc
61  tttagtggc  atccgtgact  ctctggactc  tgacccatct  aacgcctatc  tgtatdddaca
121 aatacattat  ctatctatct  atctatctat  ctatctatct  atctatctat  ctatctatca
181 atcatctatc  tatctdddctg  tctgtctddd  tgggctgcct  atggctcaac  ccaagttgaa
241 ggaggagatt  tgaccaacaa  ttcaagctct  ctgaatatgt  tttgaa
```

Figura 1.11: Sequência evidenciando a região de repetição do *locus* D13S317 (GenBank: G09017.1).

1.3.11 D16S539

O *locus* D16S539, localizado no braço longo do cromossomo 16, é formado pela unidade de repetição GATA (Figura 1.12). Apresenta nove alelos comuns, dentre eles o alelo 5 e os de 8 a 15 unidades de repetição (Butler, 2005).

```
121 agtgccagat gctcgttgtg cacaaatcta aatgcagaaa agcactgaaa gaagaatcca
181 gaaaaccaca gttcccattt ttatatggga gcaaacaaaag gcagatccca agctcttcct
241 cttccctaga tcaatacaga cagacagaca ggtggataga tagatagata gatagataga
301 tagatagata gatagatc attgaaagac aaaacagaga tggatgatag atacatgctt
361 acagatgcac acacaaacgt aatgggtatn aaaaatngga tncactcttg tanggttggt
```

Figura 1.12: Sequência evidenciando a região de repetição do *locus* D16S539 (GenBank: G07925.1).

1.3.12 D18S51

D18S51 está localizado no braço longo do cromossomo 18 e é constituído pela repetição da sequência AGAA (Figura 1.13) (Barber & Parkin, 1996; Straub *et al.*, 1993). Os microvariantes x.2 são originados pela inserção de dois pares de base, na região flanqueadora 3' (Barber & Parkin, 1996). Há registro de mais de 50 alelos desse marcador, sendo um dos mais polimórficos do CODIS (Butler, 2005).

```
61 aaaaatacaa aaatnagttg ggcattggtg cacgtgcctg tagtctcagc tacttgcagg
121 gctgaggcag gaggagttct tgagcccaga aggttaaggc tgcagtgagc catgttcatg
181 ccaactgcaact tcaactctgag tgacaaattg agaccttgtc tcagaaaagaa agaaagaaaag
241 aaagaaagaa agaaagaaaag aangaagaa agaaagtaag aaaaagagag ggaaagaaaag
301 agaaanagna aanaaatagt agcaactggt attgtaagac atctccacac accagagaag
361 ttaatttttaa ttttaacatg ttaagaacag agagaagcca acatgtccac cttagggtga
```

Figura 1.13: Sequência evidenciando a região de repetição do *locus* D18S51 (GenBank: L18333.1).

1.3.13 D21S11

O marcador genético D21S11 é formado por um tetranucleotídeo complexo, observado no braço longo do cromossomo 21. Um número variável de repetições do tipo TCTA e TCTG flanqueia uma região de 43 pares de base, composta pela sequência {[TCTA]₃ TA [TCTA]₃ TCA [TCTA]₂ TCCA TA} (Figura 1.14) (Butler, 2005). Os microvariantes x.2 são produzidos devido à inserção do dinuclotídeo TA no final 3' da região de repetição (Brinkmann *et al.* 1996).

```
61 caagagaaag ataggggagc ttgtagatgg tctggttatgg gacttttctc agtctccata
121 aatatgtgag tcaattcccc aagtgaattg ccttctatct atctatctat ctgtctgtct
181 gtctgtctgt ctgtctatct atctatatct atctatctat catctatcta tctatctatc
241 tatctatcta tctatctatc tatcgtctat ctatccagtc tatctacntc ctatnnagtc
301 tgtctctgga gaacattgac taatacaaca tctttaatat atcacagttt aatttcaagt
361 tatatcatat cacttcatac attatataaa accttacagt gtttctccct tctcagtggt
```

Figura 1.14: Sequência evidenciando a região de repetição do *locus* D21S11 (GenBank: M84567.1).

1.3.14 D2S1338

O marcador genético D2S1338 está localizado no braço longo do cromossomo 2. É constituído pela repetição dos tetranucleotídeos compostos TGCC e TTCC (Figura 1.15) (Butler, 2007).

```
1 gtggaggaa gccagtgat ttggaacag aaatggcttg gccttgctg cctgcctgcc
61 tgcctgcctt ccttccttcc ttccttctt ccttccttcc ttccttctt ccttcctgca
121 atcctttaac ttactgaata actcattatt atgggccncc tgcaggtacc atgctaggta
181 ctagggatgt aggcataaac actgacaagg gcctctggga ctggcattct ggtaggaaaa
```

Figura 1.15: Sequência evidenciando a região de repetição do *locus* D2S1338 (GenBank: G08202.1).

1.3.15 D19S433

O *locus* D19S433 está localizado no braço longo do cromossomo 19. É formado pela repetição sequencial do tetranucleotídeo AAGG (Figura 1.16) (Butler, 2007).

```
1 aaagctataa ttgtaccact gcactccagc ctgggcaaca gaataagatt ctgttgaaagg
61 aaagaaggta ggaaggaagg aaggaaggaa ggaaggaagg aaggaaggaa ggaaggagag
121 aggtagaaag agagaagatt tttattcggg taatgggtgc accaaaatat cagaaatcac
181 tgctaaagaa cttattcatg taaccaacac cacctgttcc ttaaaaacct attgaaataa
```

Figura 1.16: Sequência evidenciando a região de repetição do *locus* D19S433 (GenBank: G08036.1).

1.3.16 Penta E

O *locus* Penta E está localizado no braço longo do cromossomo 15. De forma diferenciada dos demais, é formado pela repetição sequencial do pentanucleotídeo AAAGA (*Promega Corporation*).

1.3.17 Penta D

O marcador genético Penta D está localizado no braço longo do cromossomo 21. Assim como Penta E, é formado pela repetição sequencial do pentanucleotídeo AAAGA (*Promega Corporation*).

1.4 Análise estatística na identificação humana por DNA

A análise de DNA é baseada na comparação de perfis genéticos obtidos a partir de uma amostra de referência (fonte biológica conhecida) e de uma amostra questionada (fonte biológica desconhecida). Essa comparação tem o intuito de detectar, por exemplo, se as amostras são provenientes de um mesmo doador ou se são provenientes de indivíduos que possuem vínculo genético.

Entretanto, segundo o NRC (*National Research Council*, 1996), não seria cientificamente justificável tratar uma coincidência de perfis genéticos como uma prova de identidade, sem conhecer a frequência com que essa identidade ocorre na população. Dessa forma, quando se realiza a análise, a comparação dos perfis é seguida do cálculo da razão de verossimilhança, que é a razão de probabilidades do mesmo evento sob duas hipóteses diferentes.

Considerando-se que os marcadores genéticos em estudo encontram-se em Equilíbrio de Hardy-Weinberg, calcula-se a razão de verossimilhança da seguinte forma: no caso de homozigotos, utiliza-se a razão entre 1 (hipótese de que o doador da amostra de referência é a fonte da evidência) e p^2 ; quando se trata de heterozigotos, utiliza-se a razão entre 1 e $2pq$, em que p e q são frequências de alelos na população (Butler, 2005).

Nos casos de vínculo genético, a exemplo da paternidade, a razão de verossimilhança é o índice de paternidade. Para obtê-la, utiliza-se a razão entre: 1) a probabilidade dos alelos do perfil genético do possível filho, coincidentes com o suposto pai (ou suposta mãe), serem oriundas desse pai e 2) de serem oriundas de um indivíduo selecionado aleatoriamente na população (Butler, 2005). Esse cálculo também tem por base as frequências alélicas dos marcadores genéticos. Dessa forma, fica evidente a importância da determinação dessas frequências para a solução dos casos de identificação humana, através da análise de DNA.

1.5 Grupos étnicos que influenciaram a composição genética atual

A população brasileira é uma das mais heterogêneas do ponto de vista étnico. Sua estrutura atual sofreu influência de diversos grupos étnicos, principalmente grupos europeus, ameríndios e africanos. No século XVI, quando da chegada dos portugueses ao continente americano, existiam cerca de 2,4 milhões de indígenas habitando o país (IBGE, 2000).

Num primeiro momento, os grupos indígenas perceberam a chegada do europeu numa visão mítica. Frente à invasão europeia, os indígenas defenderam seu modo de ser, mas foram superados pelas armas e as enfermidades trazidas pelos portugueses. Aos que chegavam, o Novo Mundo era fonte de ganhos. Do ponto de vista da Igreja, a expansão teve a finalidade de evangelização e formação de uma Igreja universal, com atuação europeia na conversão desses grupos em povos cristãos.

O início da formação do povo brasileiro ocorreu, principalmente, através da vinda de integrantes europeus do sexo masculino que se utilizaram do *cunhadismo*, costume indígena de incorporar à sua comunidade aqueles que recebessem uma moça índia como esposa. Após aceitar a moça, o europeu colocava os parentes a seu serviço, para conforto pessoal e produção de mercadorias, e passava a poder ter relações sexuais com todos os parentes de sua geração. Não só portugueses, mas franceses e espanhóis participaram da implantação europeia na costa brasileira, através desse modelo. Tal costume fez surgir uma numerosa camada de mestiços que ocupou o país (Ribeiro, 2006).

Em 1570, a dominação portuguesa estava assentada, com uma população de 30 a 40 mil habitantes, dos quais mais de 75% era composto por mamelucos. A maioria dos indígenas “desapareceu” e as regiões açucareiras, inicialmente formadas por mamelucos, se transformaram pela vinda cada vez mais maciça de escravos africanos. Os primeiros contingentes de negros africanos chegaram ao Brasil desde cerca de 1540 e, com o

desenvolvimento da economia açucareira, passaram a ser trazidos em maior número. Vieram, principalmente, da costa ocidental da África, capturados nas centenas de povos tribais que falavam dialetos e línguas não inteligíveis uns aos outros. Admite-se que chegaram homens e mulheres africanos, havendo variação dos dados em relação à sua proporção (Ribeiro, 2006).

A partir do século XIX, acrescenta-se à miscigenação entre os primeiros grupos étnicos a contribuição de imigrantes alemães, italianos e espanhóis. Já no século XX, o Brasil recebeu ainda imigrantes japoneses, sírios e libaneses e, mais recentemente, imigrantes latino-americanos (Fausto, 2000).

Atualmente, a classificação brasileira em grupo étnico é baseada no questionário do Instituto Brasileiro de Geografia e Estatística (IBGE), que utiliza o critério de autoclassificação da cor da pele. Sabe-se que essa e outras características físicas externas são traços superficiais, que apenas refletem adaptações ao clima e a outras variáveis ambientais.

Segundo a Pesquisa Nacional por Amostra de Domicílios (PNAD), realizada em 2009, a população brasileira atual é composta por 190 milhões de habitantes. Desse contingente, 48,2% é representado por indivíduos de cor branca, 6,9% preta, 44,2% parda e 0,7% amarela ou indígena (IBGE, 2010).

No Estado do Rio de Janeiro, estimativas da mistura étnica baseados nos sistemas eritrocitários ABO e Rh, numa amostra de mais de 32.000 indivíduos mostrou que a proporção de genes de origem caucasiana neste estado era de 52%, seguido de 40% de origem africana e os 8% restantes de origem ameríndia (Lopez-Camelo *et al.*, 1996). Segundo dados do IBGE, essa população é composta por 15 milhões de habitantes, dos quais 55,8% é formada por indivíduos de cor branca, 11,1%, de cor preta, 32,6% de cor parda e 0,4% amarela ou indígena (IBGE, 2010).

2. Objetivos

2.1 Objetivos Gerais

Realizar o estudo dos marcadores genéticos CSF1PO, FGA, TH01, TPOX, vWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, D2S1338, D19S433, Penta E e Penta D, a partir da análise de uma amostragem de 505 indivíduos não aparentados, que doaram amostras para confronto genético, no Instituto de Pesquisa e Perícias em Genética Forense, no período de 2008 a 2011.

2.2 Objetivos Específicos

- a) Construir um banco de dados de frequências alélicas;
- b) Testar a hipótese de Equilíbrio de Hardy-Weinberg nos diferentes sistemas;
- c) Calcular o Coeficiente de Endogamia (F_{IS});
- d) Calcular os diversos parâmetros de interesse forense;
- e) Analisar a composição étnica da população em estudo.

3. Material e Métodos

3.1 Coleta de Amostras

Foram utilizadas 505 amostras provenientes de coletas de indivíduos não aparentados que doaram material biológico para confronto genético, no Instituto de Pesquisa e Perícias em Genética Forense (IPPGF), no período de 2008 a 2011, e amostras questionadas encaminhadas ao IPPGF, no mesmo período. Vale ressaltar que este Instituto é o órgão Pericial da Polícia Civil e recebe amostras de todo o estado, agregando um estoque representativo da população. A figura 3.1 indica o instituto de origem das amostras questionadas encaminhadas.

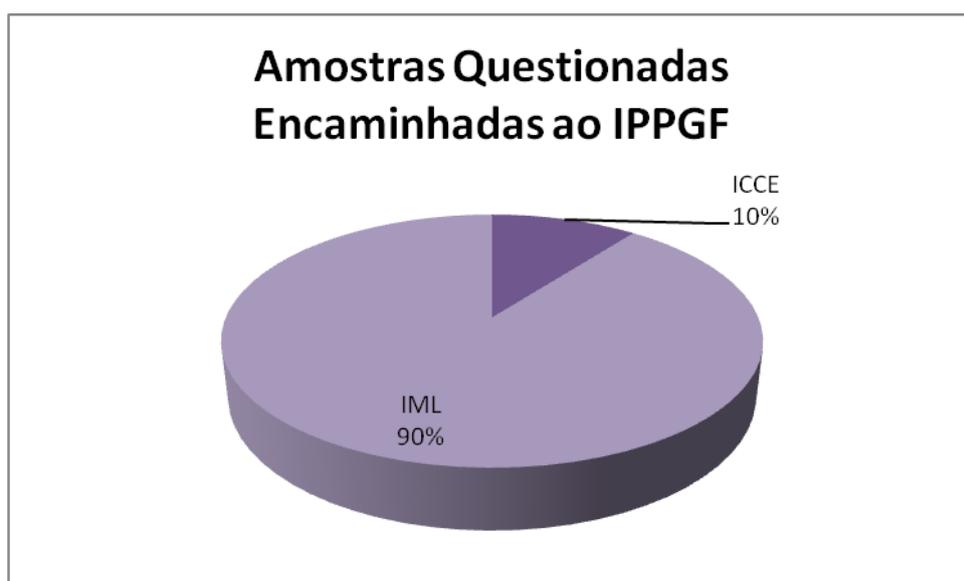


Figura 3.1: Instituto de origem do material biológico encaminhado ao IPPGF, no período de 2008 a 2010. ICCE – Instituto de Criminalística Carlos Éboli. IML – Instituto Médico Legal.

As amostras dos doadores foram, primariamente, coletadas por punção venosa, com armazenamento de sangue em tubo do tipo Vacutainer®. Este método foi substituído pela lancetação dactilar, armazenando sangue seco em papel especialmente desenvolvido para este

fim (FTA) e, posteriormente, pela coleta de esfregaço de mucosa oral, através da utilização de suabe. As amostras coletadas e o material extraído foram preservados a -20° C.

3.2 Extração de DNA

3.2.1 Extração de DNA em Papel FTA®

A extração de DNA em papel FTA se iniciou com a retirada de amostras de sangue seco contido no papel, com auxílio de um “puncher” (picotador), que promove o corte de pequenos discos, com cerca de 1,2 mm de diâmetro cada. Cada disco foi colocado em um microtubo, seguido da adição de “FTA Purification Reagent”, da empresa Whatman®, que permaneceu agindo por 20 minutos.

Posteriormente, o reagente foi removido e este procedimento foi repetido. Em seguida, os fragmentos foram lavados duas vezes com 200 µL de água estéril (ddH₂O) e os discos foram secos em um termociclador por 20 minutos. A amplificação do material genético foi realizada diretamente nesses microtubos, de acordo com o protocolo dos fabricantes dos kits de amplificação.

3.2.2 Extração de DNA por “Salting-Out”

O DNA genômico total das amostras coletadas por punção venosa foi extraído utilizando o protocolo de extração por “salting-out” (Miller *et al.*, 1988). Primeiramente, foi aliquotado 300 µL da amostra de sangue e adicionado de Tampão SLR (Tris-HCl 10mM – NaCl 10mM – MgCl₂ 5mM). As amostras foram centrifugadas por 3 minutos a 3800 rpm e o sobrenadante, desprezado. Essa etapa foi repetida até a lise completa das hemáceas. Posteriormente, acrescentou-se 500 µl de Tampão SLB (Tris-HCl 20mM – NaCl 50mM –

EDTA 10mM – SDS 0,2%), 7 µl de Proteinase K 20 mg/µL (BioAmerica) e as amostras foram incubadas a 56° C durante 2 horas.

Após a incubação, foi adicionado 200 µL de cloreto de sódio (NaCl) 5M seguido da centrifugação durante 10 minutos a 10.000 rpm. O sobrenadante foi transferido para outro tubo, adicionando-se duas vezes o volume de isopropanol absoluto gelado, por 15 minutos à -80° C.

Posteriormente, as amostras foram centrifugadas por 20 minutos a 10.000 rpm. O pellet foi lavado com 1 mL de etanol 70% gelado, seguido da centrifugação por 15 minutos a 10000 rpm. O álcool foi descartado e o DNA foi seco em uma centrífuga a vácuo (*Speed Vac*) durante 20 minutos. O DNA foi ressuspensão em 30 µL de TE⁻⁴ (Tris-EDTA) e quantificado por espectrofotometria no equipamento GeneQuant *pro*, da empresa Amersham Biosciences.

3.2.3 Extração de DNA com Resina Chelex® 100

O DNA genômico total das amostras de mucosa oral, contidas em suabe foi extraído utilizando o protocolo de extração com resina Chelex® 100 (Barber, 2004). Primeiramente, foi adicionado 1 mL de PBS (tampão fosfato-salino) às amostras, que foram incubadas à temperatura ambiente, por 30 minutos.

Após a incubação, as amostras foram centrifugadas a 12.000 rpm por 1 minuto. Posteriormente, parte do sobrenadante foi descartada, mantendo quantidade suficiente para cobrir o precipitado.

Em seguida, foi adicionado 200 µL de Chelex® a cada amostra, que foi incubada a 56°C durante 30 minutos. As amostras foram, então, centrifugadas por 3 minutos a 12.000

rpm e o DNA, quantificado por espectrofotometria no equipamento GeneQuant *pro*, da empresa Amersham Biosciences.

3.2.4 Extração de DNA pelo Método Orgânico

O DNA genômico total das demais amostras (amostras questionadas) foi extraído utilizando o protocolo de extração pelo método orgânico, modificado de Butler (2005). Foi adicionado 300 µL de Tampão de Extração de Manchas (Tris-HCl 10mM – NaCl 100mM – EDTA 10mM – SDS 1%), 24 µL de Proteinase K 20 mg/µL (BioAmerica) e 48 µL de DTT (ditiotreitól) e as amostras foram incubadas a 56° C durante 2 horas.

Após a incubação, foi adicionado fenol: clorofórmio: álcool isoamílico (25:24:1 v/v) seguido da centrifugação durante 5 minutos a 12.000 rpm. O sobrenadante foi retirado e o DNA precipitado pela adição de 10% de cloreto de sódio (NaCl) 5M e igual volume de isopropanol absoluto gelado, por 15 minutos à -80° C.

Em seguida, as amostras foram centrifugadas por 20 minutos a 10.000 rpm. O pellet foi lavado com 1 mL de etanol 70% gelado, seguido da centrifugação por 15 minutos a 10000 rpm. O álcool foi descartado e o DNA foi seco em uma centrífuga a vácuo (*Speed Vac*) durante 20 minutos. O DNA foi ressuspenso em 30 µL de TE⁻⁴ (Tris-EDTA) e quantificado por espectrofotometria no equipamento GeneQuant *pro*, da empresa Amersham Biosciences.

3.3 Amplificação do material genético

Nesse trabalho, foram realizadas amplificações, com o emprego dos sistemas de identificação humana AmpF ℓ STR® Identifiler®, AmpF ℓ STR® Identifiler® Plus e AmpF ℓ STR® MiniFiler™, da empresa Applied Biosystems, e PowerPlex® 16 System e

PowerPlex® 16 HS System, da empresa Promega. Parte das amostras foi amplificada com a utilização dos kits da Applied Biosystems e outra parcela com a utilização dos kits da Promega. Algumas amostras foram amplificadas com ambos os sistemas.

Na análise de STR, com a utilização desses kits, diversos fragmentos são amplificados numa mesma reação, através da técnica de PCR multiplex. Os iniciadores utilizados são marcados por fluoróforos distintos. Sendo assim, as sequências amplificadas contêm o iniciador capaz de emitir fluorescência de determinado comprimento de onda, que é detectado através de um analisador genético.

Os sistemas de amplificação AmpF ℓ STR® Identifiler® e AmpF ℓ STR® Identifiler® Plus proporcionam a amplificação simultânea dos treze marcadores autossômicos pertencentes à lista do CODIS, mais o marcador de gênero amelogenina, além de dois marcadores autossômicos adicionais. Eles resultam na amplificação dos mesmos *loci* e diferem no protocolo de utilização, na inclusão da Taq DNA polimerase no tampão do AmpF ℓ STR® Identifiler® Plus e na otimização para superar inibição, melhorar a linha de base dos eletroferogramas e o balanceamento de picos dos heterozigotos (*Life Technologies Corporation*, 2012; http://marketing.appliedbiosystems.com/mk/get/IDPLUS_COMPARISON_ID_KIT). Os marcadores amplificados a partir desses kits podem ser visualizados na tabela 3.1.

O sistema de identificação AmpF ℓ STR® MiniFiler™ é um sistema de mini STR, utilizado na amplificação de amostras complexas, em que se obteve resultados pouco satisfatórios. Esse kit é utilizado para a amplificação simultânea de oito marcadores STR, dos quais sete pertencem ao CODIS, mais o marcador do gênero amelogenina, conforme tabela 3.1 (*Life Technologies Corporation*, 2007).

Os kits de identificação humana PowerPlex® 16 System e PowerPlex® 16 HS System promovem a amplificação simultânea de treze marcadores autossômicos pertencentes à lista do CODIS, além de dois marcadores autossômicos adicionais, diferentes daqueles presentes nos kits AmpFℓSTR® Identifiler® e AmpFℓSTR® Identifiler®, e do marcador de gênero amelogenina (Tabela 3.1). Assim como no caso da Applied Biosystems, o sistema PowerPlex® 16 HS System é uma versão otimizada do PowerPlex® 16 System, com inclusão de uma enzima hot-start no tampão do kit (*Promega Corporation*).

CODIS	AmpFℓSTR® Identifiler®/ AmpFℓSTR® Identifiler® Plus	AmpFℓSTR® MiniFiler™	PowerPlex® 16 System/ PowerPlex® 16 HS System
CSF1PO	CSF1PO	CSF1PO	CSF1PO
FGA	FGA	FGA	FGA
D7S820	D7S820	D7S820	D7S820
D13S317	D13S317	D13S317	D13S317
D16S539	D16S539	D16S539	D16S539
D18S51	D18S51	D18S51	D18S51
D21S11	D21S11	D21S11	D21S11
TH01	TH01	D2S1338	TH01
TPOX	TPOX		TPOX
vWA	vWA		vWA
D3S1358	D3S1358		D3S1358
D5S818	D5S818		D5S818
D8S1179	D8S1179		D8S1179
	D2S1338		Penta E
	D19S433		Penta D

Tabela 3.1: Marcadores genéticos presentes no CODIS e nos kits de amplificação utilizados.

3.3.1 AmpF ℓ STR® Identifiler®

A amplificação foi realizada através de reações de PCR contendo 5,25 μ L do Master Mix, 2,75 μ L do Primer Set, que continha os *primers* para a amplificação simultânea dos marcadores autossômicos, 0,25 μ L de AmpliTaq Gold® DNA Polymerase e 1 ng de DNA molde (5 μ L), em um volume final de reação de 12,5 μ L.

Foi utilizada a seguinte programação de termociclagem: uma etapa inicial de 11 minutos a 95°C; 28 ciclos contendo 1 minuto de desnaturação a 94°C, 1 minuto de hibridação a 59°C e 1 minuto de extensão a 72°C; além de 1 ciclo de 60 minutos a 60°C.

Houve a inclusão de controles positivos, com o DNA controle 9947A, que acompanhava o kit de amplificação, e negativos, que consistiam de reações com adição de água substituindo o DNA, para detectar possíveis contaminações.

3.3.2 AmpF ℓ STR® Identifiler® Plus

As reações de PCR foram realizadas a partir da adição de 5,0 μ L do Master Mix, 2,5 μ L do Primer Set, que continha os *primers* para a amplificação simultânea dos marcadores autossômicos, e 1 ng de DNA molde (5 μ L), em um volume final de reação de 12,5 μ L.

Para proceder com a reação, utilizamos a seguinte programação de termociclagem: uma etapa inicial de 11 minutos a 95°C; 28 ciclos contendo 20 segundos de desnaturação a 94°C e 3 minutos de hibridação e extensão a 59°C; além de 1 ciclo de 10 minutos a 60°C.

Houve a inclusão de controles positivos, com o DNA controle 9947A, que acompanhava o kit de amplificação, e negativos, que consistiam de reações com adição de água substituindo o DNA.

3.3.3 AmpF ℓ STR® MiniFiler™

A amplificação foi realizada através de reações de PCR contendo 5,0 μ L do Master Mix, 2,5 μ L do Primer Set, que continha os iniciadores para a amplificação simultânea dos marcadores autossômicos, e 0,5 ng de DNA molde (5 μ L), em um volume final de reação de 12,5 μ L.

Utilizamos a seguinte programação de termociclagem: uma etapa inicial de por 11 minutos a 95°C; 30 ciclos contendo 20 segundos de desnaturação a 94°C, 2 minutos de hibridação a 59°C e 1 minuto de extensão a 72°C; além de 1 ciclo de 45 minutos a 60°C.

Foram incluídos controles positivos, com o DNA controle 007, que acompanhava o kit de amplificação, e negativos, que consistiam de reações com adição de água substituindo o DNA.

3.3.4 PowerPlex® 16 System

As reações de PCR continham 7,1 μ L água estéril (ddH₂O), 1,25 μ L do Master Mix, 1,25 μ L do Primer Pair Mix, que continha os *primers* para a amplificação simultânea dos marcadores autossômicos, 0,4 μ L de AmpliTaq Gold® DNA Polymerase e 1 ng de DNA molde (2,5 μ L), em um volume final de reação de 12,5 μ L.

Para proceder com a reação, utilizamos a seguinte programação de termociclagem: uma etapa inicial de 1 minuto a 96°C; 10 ciclos contendo 30 segundos de desnaturação a 94°C, 30 segundos de hibridação a 60°C e 45 segundos de extensão a 70°C; 22 ciclos contendo 30 segundos de desnaturação a 90°C, 30 segundos de hibridação a 60°C e 45 segundos de extensão a 70°C ; além de 1 ciclo de 30 minutos a 60°C.

Houve a inclusão de controles positivos, com o DNA controle 9947A, que acompanhava o kit de amplificação, e negativos, que consistiam de reações com adição de água substituindo o DNA.

3.3.5 PowerPlex® 16 HS System

A amplificação foi realizada através de reações de PCR contendo 2,5 µL do Master Mix, 1,25 µL do Primer Pair Mix, que continha os iniciadores para a amplificação simultânea dos marcadores autossômicos, e 0,5 ng de DNA molde (8,75 µL), em um volume final de reação de 12,5 µL.

Foi utilizada a seguinte programação de termociclagem: uma etapa inicial de 2 minutos a 96°C; 10 ciclos contendo 30 segundos de desnaturação a 94°C, 30 segundos de hibridação a 60°C e 45 segundos de extensão a 70°C; 22 ciclos contendo 30 segundos de desnaturação a 90°C, 30 segundos de hibridação a 60°C e 45 segundos de extensão a 70°C; além de 1 ciclo de 30 minutos a 60°C.

Foram incluídos controles positivos, com o DNA controle 9947A, que acompanhava o kit de amplificação, e negativos, que consistiam de reações com adição de água substituindo o DNA.

3.4 Eletroforese Capilar

Após a PCR, os produtos amplificados foram analisados no ABI PRISM™ 3100 Avant Genetic Analyser, da empresa Applied Biosystems. Nele, os fragmentos de DNA amplificados passaram por um processo de eletroinjeção, entrando no capilar e sendo direcionados para a janela de detecção.

Com isso, o produto da PCR passou por um feixe de laser e seus fluoróforos foram excitados. A luz emitida pelos fluoróforos foi detectada e a informação, processada através de um computador (<http://www.biocompare.com/Articles/ProductReview/395/Applied-Biosystems-3130-and-3130xl-Genetic-Analyzers.html>). Este procedimento gerou um gráfico do sinal analítico em função do tempo, denominado eletroferograma, conforme Figura 3.2.

A eletroforese capilar ocorreu com auxílio do *software* ABI Prism 3100 Data Collection. A análise dos eletroferogramas foi realizada no *software* GeneMapper, da empresa Applied Biosystems.

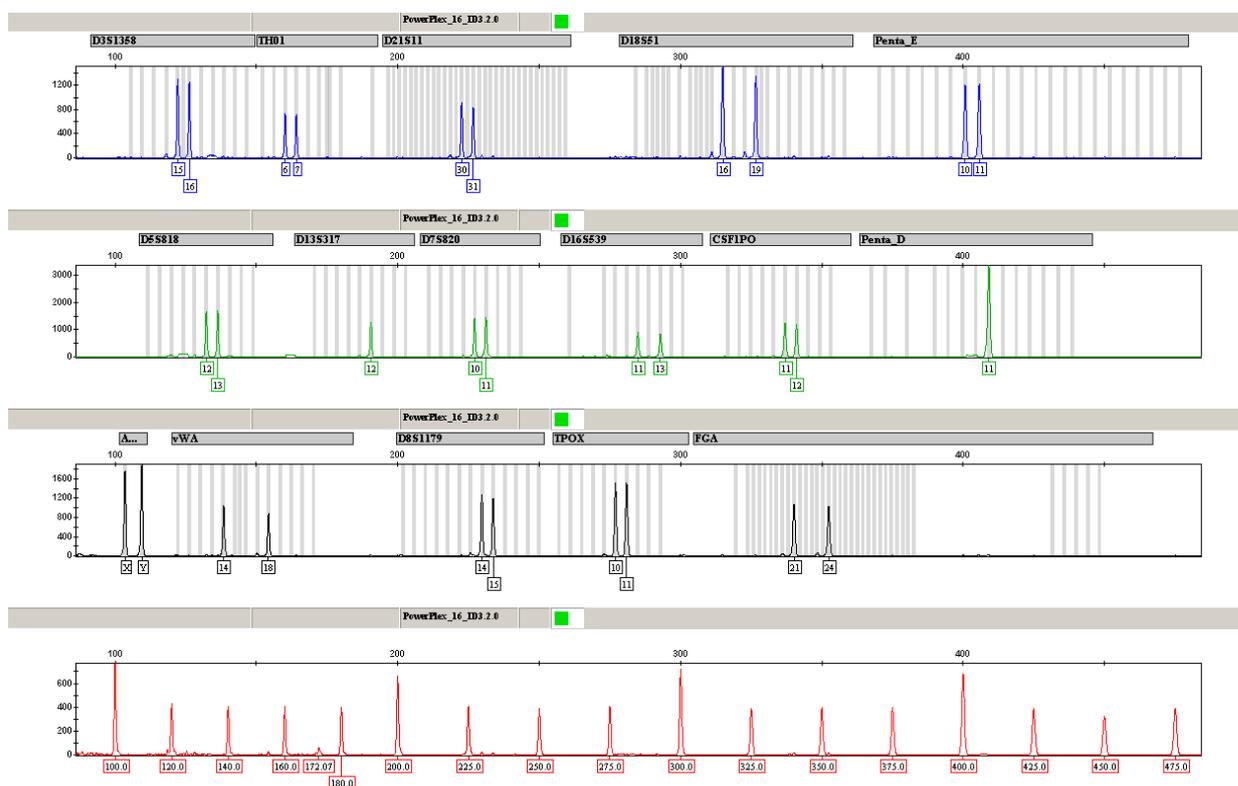


Figura 3.2: Eletroferograma, ilustrando um perfil genético de STR autossômico, com seus diferentes marcadores genéticos (IPPGF). *Loci* apresentando um único pico representam homozigose e dois picos indicam heterozigose.

3.5 Análise dos Dados

Os perfis genéticos dos indivíduos estudados foram organizados através da utilização do programa Microsoft Excel 2010. Em seguida, o *software* Convert v. 1.31 (Glaubitz, 2004) foi empregado a fim de converter o arquivo para o formato de *input* dos softwares Genepop v. 4.0 (Raymond & Rousset, 1995; Rousset, 2008) e Arlequin v. 3.11 (Excoffier *et al.*, 2005).

O programa Genepop v. 4.0 (Raymond & Rousset, 1995; Rousset, 2008) foi aplicado para calcular o valor de coeficiente de endogamia de Wright (F_{IS}) e as frequências alélicas obtidas para os marcadores estudados. A estatística F_{IS} de Wright expressa a correlação média de união dos gametas em relação às frequências gênicas da própria linhagem. Valores de F_{IS} variam entre -1 e 1, de forma que valores positivos indicam uma tendência à segregação, com endogamia que ultrapassa o nível esperado de acasalamentos ao acaso. Valores de F_{IS} próximos de zero são esperados em panmixia e valores negativos indicam que o acasalamento entre indivíduos da mesma linhagem foi evitado, levando a um excesso de heterozigotos (Wright, 1965).

O teste exato do Equilíbrio de Hardy-Weinberg foi realizado no *software* Arlequin v. 3.11 (Excoffier *et al.*, 2005).

Com o intuito de conhecer as relações evolutivas entre a população do Rio de Janeiro e as populações das diversas regiões brasileiras, foi desenhada uma árvore filogenética, no programa TreeView v. 1.4 (Page, 1996), através da comparação dos resultados observados com dados dos marcadores genéticos FGA, TH01, vWA, D3S1358, D8S1179, D16S539, D18S51, D21S11, D2S1338, D19S433 de diferentes populações. Essa árvore foi baseada no modelo de agrupamento de vizinhos (Neighbor-Joining) de F_{ST} par-a-par.

A matriz de do índice de fixação (F_{ST}) (Wright, 1965) par a par, foi obtida no programa GDA v. 1.1 (Lewis & Zaykin, 2002), utilizando 10.000 permutações. O F_{ST} par a par mede o grau de diferenciação genética das subpopulações (Nei, 1977), através do cálculo da probabilidade de encontro de dois gametas idênticos por descendência. Dessa forma, quanto maior o valor observado no índice de fixação, menor será o fluxo gênico entre as populações.

A fim de estudar a diferenciação genética da população e analisar sua mistura étnica, foi realizada ainda uma análise Bayesiana, implementada no software Structure v. 2.3.3 (Pritchard *et al.*, 2000). Empregou-se um valor de *burnin* de 100.000 permutações e *Markov Chain Montecarlo* (MCMC) com 1.000.000 de permutações. As análises foram realizadas através do modelo de mistura (*admixture model*), em que cada amostra pode apresentar ancestrais de mais de uma população. Foram realizadas dez réplicas para cada valor de K (K= 1 a K= 10).

Os parâmetros estatísticos forenses intitulados Probabilidade de Coincidência Aleatória (PM), Poder de Discriminação (PD), Poder de Exclusão (PE) e Conteúdo de Informação Polimórfica (PIC) foram calculados com a utilização do programa PowerStats v. 1.2 (*Promega Corporation*). As definições desses parâmetros são descritas a seguir.

A Probabilidade de Coincidência Aleatória (PM) é obtida a partir da soma dos quadrados das frequências genotípicas (Adam, 2010). Dessa forma, é a probabilidade de duas pessoas não relacionadas compartilharem um mesmo perfil genético (Jobling & Gill, 2004). Pode ser expresso ainda na forma de $1/PM$.

Poder de Discriminação (PD) é a probabilidade de dois indivíduos selecionados aleatoriamente não possuírem o mesmo perfil genético. É definido por 1 menos a soma dos quadrados das frequências genotípicas (Adam, 2010).

Poder de Exclusão (PE) é a probabilidade de um *locus* em questão excluir uma paternidade erroneamente atribuída e pode ser obtido pela fórmula $PE = h^2 (1 - (1-h) h^2)$, em que h = heterozigosidade (Butler, 2005).

O Conteúdo de Informação Polimórfica (PIC) reflete a probabilidade com que o genótipo da prole permite a dedução do genótipo parental para um determinado *locus* (Guo & Elston, 1999). Esse parâmetro pode ser determinado através da fórmula

$1 - \left(\sum_{i=1}^n p_i^2 \right) - \sum_{i=1}^{n-1} \sum_{j=i+1}^n 2p_i^2 p_j^2$, onde n = número de alelos e p = frequência alélica (Botstein, 1980; Buckleton *et al.*, 2005).

3.6 Aspectos Éticos

O projeto de pesquisa do presente trabalho foi aprovado em seus aspectos éticos pelo Comitê de Ética em Pesquisa/Fiocruz, de acordo com o protocolo número 604/11.

4. Resultados

A análise de 17 marcadores autossômicos do tipo STR, de indivíduos não aparentados, revelou um total de 215 alelos distintos. O número de indivíduos estudados para cada marcador genético pode ser observado na Tabela 4.1.

<i>Locus</i>	Nº de indivíduos
CSF1PO	498
FGA	494
TH01	504
TPOX	498
vWA	502
D3S1358	503
D5S818	502
D7S820	499
D8S1179	498
D13S317	503
D16S539	502
D18S51	497
D21S11	500
D2S1338	211
D19S433	209
Penta E	283
Penta D	283

Tabela 4.1: Tamanho amostral estudado para cada marcador genético.

A Tabela 4.2 ilustra o número de alelos obtidos, faixa de variação e alelos mais frequentes em cada marcador genético.

<i>Locus</i>	Faixa de Variação dos Alelos	Nº de Alelos	Alelos mais Frequentes	
			Alelo	Frequência
CSF1PO	7 - 15	9	12	0,312
			11	0,276
FGA	13 - 32	22	22	0,170
			21	0,163
TH01	4 - 10	8	7	0,261
			9	0,192
TPOX	5 - 13	9	8	0,432
			11	0,261
vWA	12 - 22	11	17	0,250
			16	0,234
D3S1358	12 - 20	9	15	0,307
			16	0,284
D5S818	7 - 15	9	12	0,347
			11	0,300
D7S820	7 - 14	8	10	0,300
			11	0,209
D8S1179	8 - 18	11	14	0,294
			13	0,264
D13S317	7 - 15	9	12	0,335
			11	0,298
D16S539	5 - 14	9	11	0,284
			12	0,235
D18S51	9 - 27	18	15	0,170
			16	0,156
D21S11	19 - 38	25	30	0,242
			29	0,216
D2S1338	15 - 28	13	17	0,182
			19	0,152
D19S433	10 - 17	13	13	0,258
			14	0,249
Penta E	5 - 23	18	12	0,159
			11	0,125
Penta D	2.2 - 17	14	9	0,182
			12	0,166

Tabela 4.2: Número de alelos obtidos, faixa de variação e alelos mais frequentes.

A Tabela 4.3 apresenta as frequências observadas dos alelos, para os marcadores genéticos estudados.

Alelos	CSF1PO	FGA	TH01	TPOX	vWA	D3S1358
4			0,001			
5			0,002	0,001		
6			0,169	0,032		
7	0,021		0,261	0,011		
8	0,029		0,182	0,432		
9	0,022		0,192	0,142		
9.3			0,184			
10	0,262		0,010	0,076		
11	0,276			0,261	0,012	
12	0,312			0,040		0,004
13	0,069	0,001		0,005	0,005	0,002
14	0,007				0,076	0,089
14.2						
15	0,001				0,180	0,307
16					0,234	0,284
17		0,001			0,250	0,198
18		0,006			0,149	0,104
18.2		0,010				
19		0,071			0,064	0,008
19.2		0,001				
20		0,095			0,023	0,003
21		0,163			0,006	
21.2		0,003				
22		0,170			0,001	
22.2		0,006				
23		0,158				
23.2		0,002				
24		0,156				
25		0,085				
26		0,050				
27		0,013				
28		0,004				
29		0,001				
30.2		0,001				
31.2		0,002				
32		0,001				

Tabela 4.3: Frequências de alelos dos 17 marcadores do tipo STR.

Alelos	D5S818	D7S820	D8S1179	D13S317	D16S539	D21S11
5					0,001	
7	0,018	0,011		0,001		
8	0,025	0,187	0,008	0,075	0,021	
9	0,041	0,124	0,012	0,069	0,151	
10	0,042	0,300	0,056	0,052	0,101	
11	0,300	0,209	0,072	0,298	0,284	
12	0,347	0,144	0,116	0,335	0,235	
12.1					0,001	
13	0,210	0,023	0,264	0,119	0,181	
14	0,015	0,001	0,294	0,05	0,025	
15	0,003		0,138	0,002		
16			0,027			
17			0,010			
18			0,002			
19						0,001
20						0,001
23.2						0,001
24.2						0,001
25						0,001
25.2						0,001
27						0,026
28						0,142
29						0,216
29.2						0,004
30						0,242
30.2						0,030
31						0,068
31.2						0,103
32						0,008
32.2						0,071
33						0,004
33.2						0,041
34						0,003
34.2						0,004
35						0,018
35.2						0,001
36						0,011
37						0,001
38						0,001

Tabela 4.3 (continuação): Frequências de alelos dos 17 marcadores do tipo STR.

Alelos	D18S51	D2S1338	D19S433	Penta E	Penta D
2.2					0,044
3.2					0,004
5				0,067	0,030
6					0,002
7				0,083	0,030
8				0,087	0,055
9	0,001			0,019	0,182
10	0,008		0,002	0,078	0,120
10.2	0,001				
11	0,004		0,036	0,125	0,150
12	0,114		0,115	0,159	0,166
12.2			0,012		
13	0,106		0,258	0,097	0,155
13.2			0,060		
14	0,128		0,249	0,069	0,048
14.2			0,031		
15	0,170	0,002	0,134	0,046	0,012
15.2			0,053		
16	0,156	0,043	0,017	0,049	
16.2			0,031		
17	0,130	0,182	0,002	0,037	0,002
18	0,089	0,062		0,032	
19	0,051	0,152		0,021	
20	0,025	0,135		0,014	
21	0,008	0,038		0,005	
22	0,006	0,100		0,005	
23	0,001	0,104		0,005	
24	0,002	0,102			
25		0,064			
26		0,014			
27	0,001				
28		0,002			

Tabela 4.3 (continuação): Frequências de alelos dos 17 marcadores do tipo STR.

Na tabela 4.4, são apresentados os principais parâmetros estatísticos forenses obtidos para os marcadores genéticos estudados.

	CSF1PO	FGA	TH01	TPOX	vWA	D3S1358	D5S818	D7S820	D8S1179
Poder de Discriminação (PD)	0,895	0,966	0,886	0,879	0,939	0,909	0,896	0,928	0,935
Probabilidade de Coincidência Aleatória (PM)	0,105	0,034	0,114	0,121	0,061	0,091	0,104	0,072	0,065
1 / PM	9.536	29.426	8.753	8.288	16.299	11.027	9.625	13.937	15.296
Poder de Exclusão (PE)	0,489	0,706	0,543	0,419	0,652	0,543	0,458	0,525	0,581
Conteúdo de Informação Polimórfica (PIC)	0,774	0,850	0,702	0,676	0,792	0,729	0,699	0,764	0,774
	D13S317	D16S539	D18S51	D21S11	D2S1338	D19S433	Penta E	Penta D	
Poder de Discriminação (PD)	0,915	0,922	0,971	0,896	0,972	0,898	0,983	0,963	
Probabilidade de Coincidência Aleatória (PM)	0,085	0,078	0,029	0,104	0,028	0,102	0,017	0,037	
1 / PM	11.718	12.821	34.488	9.573	35.874	9.772	57.794	27.103	
Poder de Exclusão (PE)	0,505	0,573	0,748	0,457	0,716	0,434	0,761	0,602	
Conteúdo de Informação Polimórfica (PIC)	0,737	0,766	0,865	0,717	0,871	0,716	0,905	0,842	

Tabela 4.4: Principais parâmetros estatísticos forenses, no conjunto total de *loci*.

Na tabela 4.5, podem ser observados os dados da heterozigosidade observada, o cálculo da heterozigosidade esperada pelo equilíbrio de Hardy-Weinberg, o valor do coeficiente de endogamia (F_{IS}) e o resultado do teste do equilíbrio genético.

Locus	Heterozigosidade Obs.	Heterozigosidade Exp.	F_{IS}	P-valor
CSF1PO	0,73896	0,75172	0,0170	0,72746
FGA	0,85830	0,87204	0,0158	0,15842
TH01	0,80754	0,80049	-0,0088	0,03191*
TPOX	0,69478	0,71753	0,0317	0,42023
vWA	0,82869	0,81818	-0,0128	0,06828
D3S1358	0,76938	0,76747	-0,0025	0,41222
D5S818	0,71912	0,74196	0,0308	0,42382
D7S820	0,75952	0,79516	0,0449	0,33503
D8S1179	0,79116	0,80262	0,0143	0,89844
D13S317	0,69478	0,71753	0,0292	0,42023
D16S539	0,78685	0,79800	0,0140	0,00611**
D18S51	0,87726	0,87912	0,0021	0,37653
D21S11	0,81600	0,85136	0,0416	0,27268
D2S1338	0,86256	0,88504	0,0255	0,36549
D19S433	0,80383	0,83219	0,0342	0,28953
Penta E	0,88339	0,91341	0,0329	0,03092*
Penta D	0,81979	0,87048	0,0583	0,36658

Tabela 4.5: Teste exato para hipótese de Equilíbrio de Hardy-Weinberg e valores do coeficiente de endogamia (F_{IS}). Valores significativos indicados com asterisco (** $p < 0,01$; * $p < 0,05$).

Na Figura 4.1, é possível visualizar a árvore filogenética, baseada no agrupamento de vizinhos (Neighbor-Joining) de F_{ST} par-a-par, incluindo a população do Rio de Janeiro (IPPGF) e populações das diferentes regiões brasileiras.

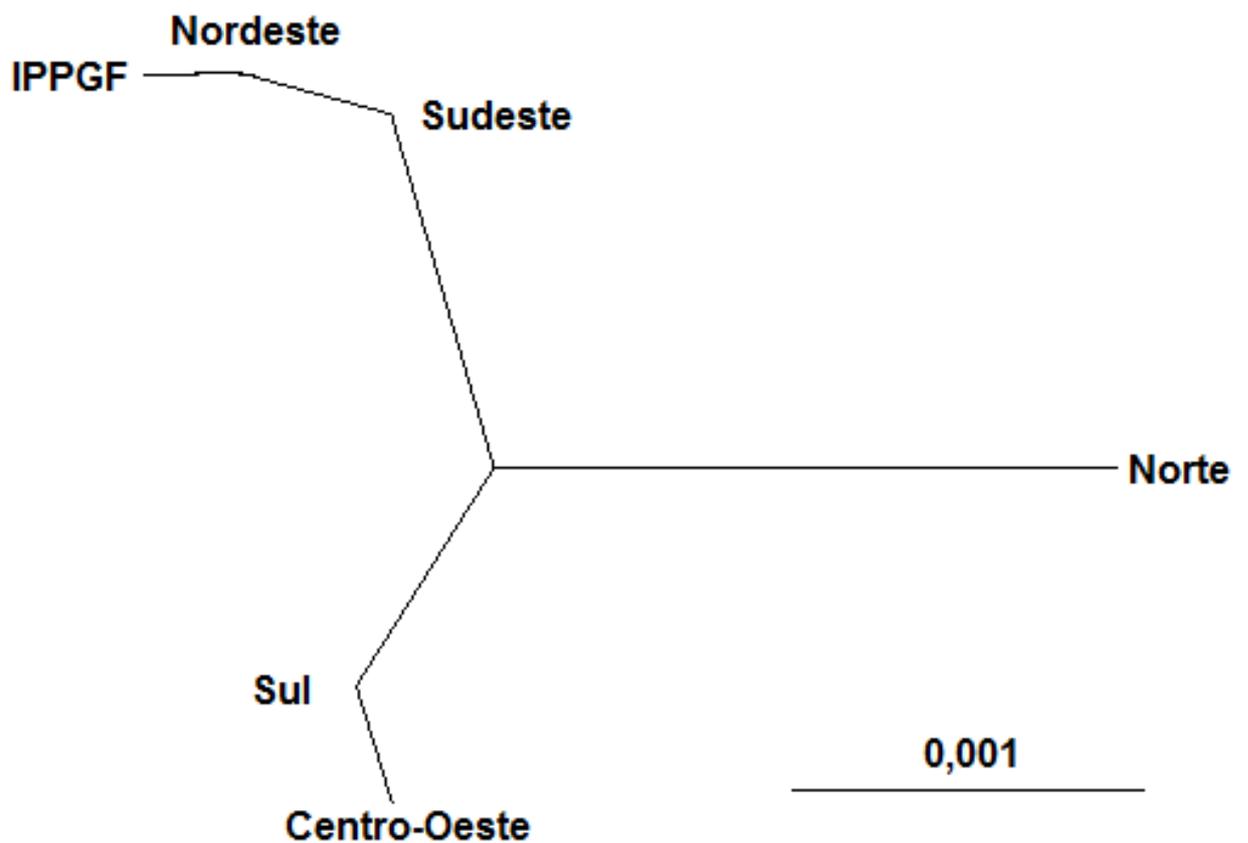


Figura 4.1: Árvore filogenética, baseada no agrupamento de vizinhos (Neighbor-Joining) de F_{ST} par-a-par, incluindo a população do Rio de Janeiro (IPPGF) e populações das diferentes regiões brasileiras.

Não foram observados valores estatisticamente significativos nas comparações entre a população do Rio de Janeiro e as populações das demais regiões brasileiras, na análise de estruturação a partir de comparações de F_{ST} par a par (Tabela 4.6).

	IPPGF	Norte	Nordeste	Centro-Oeste	Sudeste	Sul
IPPGF	-----					
Norte	0.003474	-----				
Nordeste	0.000233	0.003128	-----			
Centro-Oeste	0.002802	0.003046	0.002682	-----		
Sudeste	0.000492	0.002693	0.000067	0.001652	-----	
Sul	0.002063	0.001961	0.002201	-0.000142	0.001001	-----

Tabela 4.6: Valores de F_{ST} par a par entre as populações brasileiras.

A estrutura “Triangle Plot” populacional, baseada na análise dos dezessete *loci*, realizada no software Structure v. 2.3.3 (Pritchard *et al.*, 2000), sugere uma centralização dos pontos, que representam os indivíduos da população (Figura 4.2).

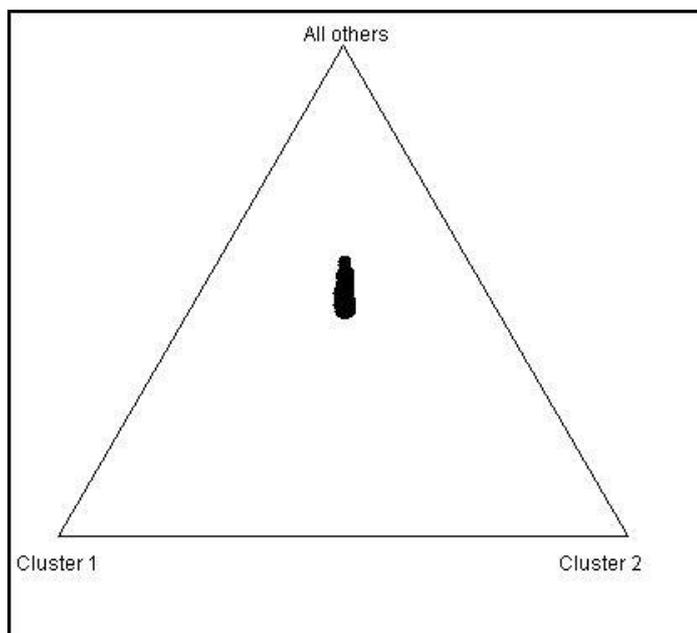


Figura 4.2: “Triangle Plot” populacional, baseada na análise dos dezessete *loci* STR estudados.

5. Discussão

5.1 Parâmetros estatísticos no conjunto total de *loci*

As análises realizadas evidenciaram que os marcadores genéticos TH01 e D7S820 apresentaram o menor número de alelos (Tabela 4.3). Este resultado é comparável ao observado em outros estudos conduzidos na população brasileira (Chula *et al.*, 2009; Gattapaglia *et al.*, 2001). Por outro lado, o marcador genético D21S11, com um total de 25 alelos, foi o mais polimórfico, resultado similar ao obtido em outro trabalho realizado na população do Rio de Janeiro (Góes *et al.*, 2004) e em populações brasileiras (Chula *et al.*, 2009; Francez *et al.*, 2011; Manta *et al.*, 2012).

Após o D21S11, o marcador genético FGA apresentou o maior número de alelos. Segundo Bulter (2005) esse é um dos *loci* mais polimórficos utilizados na identificação humana por DNA. Os microvariantes x.2, gerados pela inserção ou deleção de dinucleotídeos (Barber *et al.*, 1996), contribuíram para essa variação característica do marcador.

Conforme pode ser verificado na tabela 4.3, o alelo 8, do marcador genético TPOX, foi o mais frequentemente observado. O marcador apresentou o menor Poder de Discriminação (PD = 0,879) dentre os analisados, dado concordante com Hares (2012), além dos menores Poder de Exclusão (PE = 0,419) e Conteúdo de Informação Polimórfica (PIC = 0,676) e a maior Probabilidade de Coincidência Aleatória (PM = 0,121) (Tabela 4.4). Apesar do resultado, a sua baixa taxa de mutação, importante característica para a seleção de um marcador STR com intuito forense (Hares, 2012), pode torná-lo útil em testes de paternidade (Butler, 2006).

O *locus* Penta E apresentou o maior Poder de Discriminação (PD = 0,983), Poder de Exclusão (PE = 0,761) e Conteúdo de Informação Polimórfica (PIC = 0,905) e a menor

Probabilidade de Coincidência Aleatória (PM = 0,017) (Tabela 4.4), demonstrando ser esse um importante marcador genético para a identificação humana. Vale observar, entretanto que se trata de um pentanucleotídeo e, devido ao maior tamanho do fragmento gerado na amplificação (de 379 a 484 pares de base – STRBase), algumas amostras complexas apresentam maior dificuldade de amplificação para este marcador.

Seguido do *locus* Penta E, o marcador genético D2S1338 apresentou o maior Poder de Discriminação (PD = 0,971), Conteúdo de Informação Polimórfica (PIC = 0,905) e a menor Probabilidade de Coincidência Aleatória (PM = 0,029), enquanto o *locus* D18S51 apresentou o maior Poder de Exclusão (PE = 0,748) (Tabela 4.4).

O teste exato para a hipótese de Equilíbrio de Hardy-Weinberg demonstrou que três marcadores genéticos apresentavam desvio significativo em relação ao equilíbrio, são eles: TH01, D16S539 e Penta E (Tabela 4.5). Com o intuito de reduzir o erro estatístico do tipo I, foi realizada a correção de Bonferroni, (α/k), sendo $\alpha = 0,05$ e $k=17$ (número de loci). Dessa forma, um valor de p menor do que 0,0029 foi requerido para o teste de Equilíbrio ser considerado significativo. Com isso, ficou evidenciado que os desvios observados nos *loci* TH01, D16S539 e Penta E podem ser considerados puramente casuais.

Como não foram observados valores estatisticamente significativos, através da análise do coeficiente de endogamia de Wright (F_{IS}), não foi possível observar uma tendência à endogamia ou que o acasalamento entre indivíduos da mesma linhagem tenha sido evitado. Vale citar que o F_{IS} médio obtido na população foi de 0,0196.

5.2 Análise da diferenciação genética

A árvore filogenética, desenvolvida pelo método de Neighbor-Joining, possibilitou detectar uma semelhança da população do Rio de Janeiro com a população do Nordeste

brasileiro (Figura 4.1). Esta população possui maior contribuição de afrodescendentes do que as demais populações brasileiras, razão pela qual a população estudada pode ter se aproximado da população nordestina.

No entanto, não foram observados valores estatisticamente significativos na análise de F_{ST} par a par entre a população do Rio de Janeiro e as populações das demais regiões brasileiras (Tabela 4.6). Este resultado mostra ausência de estruturação, importante característica no caso de marcadores genéticos utilizados em testes de identificação humana. Não é desejável existirem diferenças significativas nas frequências, porque poderia gerar aumento na probabilidade de erros estatísticos, como a superestimativa do índice de paternidade, em decorrência da estruturação (Francez *et al.*, 2011).

Em relação ao gráfico “Triangle Plot” populacional, gerado no *software* Structure, com base nos 17 marcadores genéticos autossômicos, pode-se perceber uma centralização dos pontos, que representam os indivíduos da população (Figura 4.2). Tal configuração sugere que, em média, os indivíduos não apresentam contribuição maior de determinado grupo étnico em comparação com os demais, podendo a população ser formada por contribuição similar dos grupos étnicos principais.

A fim de complementar as análises, foram realizadas ainda comparações das frequências alélicas obtidas no presente trabalho, com as frequências dos marcadores estudados em diversas populações.

5.2 Frequências Alélicas

As frequências alélicas obtidas no presente estudo adicionaram novos dados para *loci* amplamente empregados nos laboratórios de Genética Forense. O *locus* Penta D não possuía

descrição de frequências de nenhum alelo para a população do Rio de Janeiro. Além disso, os seguintes alelos não possuíam frequência descrita para a mesma população: 13, 19.2, 30.2, 31.2 e 32 (FGA); 5 (TH01); 5 (TPOX); 22 (vWA); 18 (D8S1179); 5 e 12.1 (D16S539); 24 (D18S51); 19, 20, 23.2, 35.2 e 37 (D21S11). Vale observar que não foi detectado nenhum alelo exclusivo da população.

5.2.1 CSF1PO

A comparação dos dados obtidos com as frequências de diversas localidades permitiu detectar que, o alelo 12, o de maior frequência detectado (0,312), no *locus* CSF1PO, para a população do Rio de Janeiro, também foi o de maior frequência em Indígenas da Amazônia Brasileira (Mendes-Junior, 2001), no Egito (Coudray *et al.*, 2007a), Itália (Brisighelli *et al.*, 2009), Japão (Okamoto *et al.*, 2003), Líbia (Immel *et al.*, 2006), Marrocos – Asni (Coudray *et al.*, 2007b), Moçambique (Alves *et al.*, 2004), Norte e Nordeste de Portugal (Pinheiro *et al.*, 2005; Alves *et al.*, 2011) e Síria (Abdin *et al.*, 2003). Além disso, foi o segundo mais frequente no Sul da Espanha – Andaluzia (Gaibar *et al.*, 2010), Centro de Portugal (Lopes *et al.*, 2009) e Tanzânia (Forward *et al.*, 2008), sendo portanto, de grande frequência em diversas populações.

Por outro lado, o alelo 11, segundo mais frequente para o mesmo marcador (0,276), teve destaque nas populações da Espanha (Andaluzia) e Sul da Espanha (Coudray *et al.*, 2007c; Gaibar *et al.*, 2010), Itália (Brisighelli *et al.*, 2009), Japão (Okamoto *et al.*, 2003), Marrocos – Bouhria (Coudray *et al.*, 2007 b), Namíbia (Muro *et al.*, 2008) e Síria (Abdin *et al.*, 2003).

5.2.2 FGA

No caso do marcador genético FGA, o alelo mais frequente, o 22 (0,170) foi também o de maior frequência nas populações da Alemanha (Reichenpfader *et al.*, 2003), China (Gusmão *et al.*, 2010), Sul da Espanha (Gaibar *et al.*, 2010), Itália (Brisighelli *et al.*, 2009), Japão (Okamoto *et al.*, 2003), Líbia (Immel *et al.*, 2006), Marrocos – Asni (Coudray *et al.*, 2007b), Namíbia (Muro *et al.*, 2008) e o segundo mais frequente no Centro e Nordeste de Portugal (Lopes *et al.*, 2009; Pinheiro *et al.*, 2005).

Já o alelo 21, segundo mais observado (0,163), destacou-se nas populações da Alemanha (Reichenpfader *et al.*, 2003), Egito (Coudray *et al.*, 2007a), Espanha – Andaluzia (Coudray *et al.*, 2007c), Itália (Brisighelli *et al.*, 2009), Marrocos – Bouhria (Coudray *et al.*, 2007b), Centro, Norte e Nordeste de Portugal (Lopes *et al.*, 2009; Amorim *et al.*, 2001; Pinheiro *et al.*, 2005) e Síria (Abdin *et al.*, 2003).

5.2.3 TH01

O *locus* TH01 apresentou o alelo 7 como o de maior frequência (0,261). Este alelo também foi o de maior ocorrência nas populações de Indígenas da Amazônia Brasileira (Mendes-Junior, 2001; Manta *et al.*, 2012), no Marrocos – Asni (Coudray *et al.*, 2007b), Moçambique (Alves *et al.*, 2004), Namíbia (Muro *et al.*, 2008) e Tanzânia (Forward *et al.*, 2008) e o segundo mais frequente no Egito (Coudray *et al.*, 2007a) e Japão (Okamoto *et al.*, 2003).

O alelo 9 deste marcador apresentou a segunda maior ocorrência na população estudada (frequência de 0,192). Este alelo apresentou frequência alta na Espanha – Andaluzia (Coudray *et al.*, 2007c), Itália (Brisighelli *et al.*, 2009), Japão (Okamoto *et al.*, 2003), Líbia

(Immel *et al.*, 2006), Marrocos – Asni e Bouhria (Coudray *et al.*, 2007b), Centro de Portugal (Lopes *et al.*, 2009) e Síria (Abdin *et al.*, 2003).

Vale ressaltar que, para o marcador TH01, o microvariante 9.3 é amplamente distribuído em populações europeias como é o caso da Alemanha (Reichenpfader *et al.*, 2003), Espanha (Andaluzia) e Sul da Espanha (Coudray *et al.*, 2007c; Gaibar *et al.*, 2010), Centro, Norte e Nordeste de Portugal (Manco *et al.*, 2008; Lopes *et al.*, 2009; Pinheiro *et al.*, 2005; Alves *et al.*, 2011). Esse alelo apresentou a frequência de 0,184 no presente estudo.

5.2.4 TPOX

Para o marcador genético TPOX, os alelos 8 e 11 foram os mais frequentemente observados na população do Rio de Janeiro (0,432 e 0,261 respectivamente) e em diversas populações (Abdin *et al.*, 2003; Okamoto *et al.*, 2003; Pinheiro *et al.*, 2005; Immel *et al.*, 2006; Coudray *et al.*, 2007a,b,c; Manco *et al.*, 2008; Muro *et al.*, 2008; Brisighelli *et al.*, 2009; Lopes *et al.*, 2009; Gaibar *et al.*, 2010; Alves *et al.*, 2011), indicando uma homogeneidade na distribuição dos alelos desse *locus* em variadas regiões.

5.2.5 vWA

No *locus* vWA, o alelo 17 apresentou a maior frequência (0,250), assim como em estudos nas populações da Alemanha (Reichenpfader *et al.*, 2003), Egito (Coudray *et al.*, 2007a), Sul da Espanha (Gaibar *et al.*, 2010), Itália (Brisighelli *et al.*, 2009), Japão (Okamoto *et al.*, 2003), Líbia (Immel *et al.*, 2006), Marrocos – Asni e Bouhria (Coudray *et al.*, 2007b), Centro Portugal (Lopes *et al.*, 2009; Pinheiro *et al.*, 2005) e Síria (Abdin *et al.*, 2003) e a segunda maior em Indígenas da Amazônia Brasileira (Terenas) (Manta *et al.*, 2012), na China

(Gusmão *et al.*, 2010), Espanha – Andaluzia (Coudray *et al.*, 2007c), Norte e Nordeste de Portugal (Amorim *et al.*, 2001).

Seguindo o 17, o alelo 16 apresentou frequência de 0,234, também observado com alta frequência em Indígenas da Amazônia Brasileira (Terenas) (Manta *et al.*, 2012), no Egito (Coudray *et al.*, 2007a), Espanha (Andaluzia) e Sul da Espanha (Coudray *et al.*, 2007c; Gaibar *et al.*, 2010), Itália (Brisighelli *et al.*, 2009), Líbia (Immel *et al.*, 2006), Moçambique (Alves *et al.*, 2004), Namíbia (Muro *et al.*, 2008), Centro, Norte e Nordeste de Portugal (Lopes *et al.*, 2009; Amorim *et al.*, 2001; Alves *et al.*, 2011; Pinheiro *et al.*, 2005), Síria (Abdin *et al.*, 2003) e Tanzânia (Forward *et al.*, 2008).

Vale observar que, para este marcador genético, o alelo 15 foi um dos mais frequentes em populações africanas, como no Marrocos – Asni (Coudray *et al.*, 2007b), Moçambique (Alves *et al.*, 2004), Namíbia (Muro *et al.*, 2008) e Tanzânia (Forward *et al.*, 2008). Na população do Rio de Janeiro, este alelo apresentou frequência de 0,180.

5.2.6 D3S1358

O alelo 15, o de maior frequência detectada (0,307), no *locus* D3S1358, para a população do Rio de Janeiro, apresentou um padrão similar em Indígenas da Amazônia Brasileira (Terenas) (Manta *et al.*, 2012), na China (Gusmão *et al.*, 2010), Sul da Espanha (Gaibar *et al.*, 2010), Japão (Okamoto *et al.*, 2003), Líbia (Immeal *et al.*, 2006), Centro, Norte e Nordeste de Portugal (Lopes *et al.*, 2009; Amorim *et al.*, 2001; Pinheiro *et al.*, 2005; Gusmão *et al.*, 2000) e Síria (Abdin *et al.*, 2003), não sendo o mais frequente nas populações africanas analisadas. No entanto, apresentou a segunda maior frequência na Alemanha (Reichenpfader *et al.*, 2003), Itália (Brisighelli *et al.*, 2009), Moçambique (Alves *et al.*, 2004), Namíbia (Muro *et al.*, 2008) e Tanzânia (Forward *et al.*, 2008).

O alelo 16 do marcador D3S1358, segundo em frequência (0,284), foi observado com alta frequência em Indígenas da Amazônia Brasileira (Terenas) (Manta *et al.*, 2012), na Alemanha (Reichenpfader *et al.*, 2003), China (Gusmão *et al.*, 2010), Egito (Coudray *et al.*, 2007a), Itália (Brisighelli *et al.*, 2009), Japão (Okamoto *et al.*, 2003), Marrocos – Asni e Bouhria (Coudray *et al.*, 2007b), Moçambique (Alves *et al.*, 2004), Namíbia (Muro *et al.*, 2008), Centro, Norte e Nordeste de Portugal (Lopes *et al.*, 2009; Amorim *et al.*, 2001; Gusmão *et al.*, 2000), Síria (Abdin *et al.*, 2003) e Tanzânia (Forward *et al.*, 2008).

5.2.7 D5S818

No *locus* D5S818, o 12, alelo mais observado na população estudada (frequência de 0,347), também apresentou a maior ocorrência nas populações do Sul da Espanha (Coudray *et al.*, 2007c), Itália (Brisighelli *et al.*, 2009), Líbia (Immel *et al.*, 2006), Marrocos – Asni e Bouhria (Coudray *et al.*, 2007b), Moçambique (Alves *et al.*, 2004), Centro, Norte e Nordeste de Portugal (Lopes *et al.*, 2009; Amorim *et al.*, 2001; Pinheiro *et al.*, 2005; Alves *et al.*, 2011), e segunda maior nas populações do Egito (Coudray *et al.*, 2007a), Espanha – Andaluzia (Coudray *et al.*, 2007c), Japão (Okamoto *et al.*, 2003), Namíbia (Muro *et al.*, 2008) e Síria (Abdin *et al.*, 2003).

O alelo 11, segundo mais observado no presente trabalho (frequência de 0,300), para o mesmo marcador genético, apresentou uma frequência significativa em populações distribuídas em diferentes regiões (Gusmão *et al.*, 2000; Amorim *et al.*, 2001; Abdin *et al.*, 2003; Okamoto *et al.*, 2003; Pinheiro *et al.*, 2005; Immel *et al.*, 2006; Coudray *et al.*, 2007a,b,c; Brisighelli *et al.*, 2009; Lopes *et al.*, 2009; Gusmão *et al.*, 2010; Gaibar *et al.*, 2010; Alves *et al.*, 2011).

5.2.8 D7S820

Para o marcador genético D7S820, o alelo 10 apresentou a maior frequência (0,300), assim como no Egito (Coudray *et al.*, 2007a), Espanha – Andaluzia (Coudray *et al.*, 2007c), Itália (Brisighelli *et al.*, 2009), Líbia (Immel *et al.*, 2006), Marrocos – Bouhria (Coudray *et al.*, 2007b), Moçambique (Alves *et al.*, 2004), Namíbia (Muro *et al.*, 2008), Centro Norte e Nordeste de Portugal (Manco *et al.*, 2008; Lopes *et al.*, 2009; Amorim *et al.*, 2001; Gusmão *et al.*, 2000; Pinheiro *et al.*, 2005), Síria (Abdin *et al.*, 2003) e Tanzânia (Forward *et al.*, 2008). Além disso, foi detectada a segunda maior ocorrência no Japão (Okamoto *et al.*, 2003) e Marrocos – Asni (Coudray *et al.*, 2007b).

Seguido do 10, o alelo 11 apresentou frequência de 0,209, também observado com frequência alta em diversas populações (Amorim *et al.*, 2001; Abdin *et al.*, 2003; Okamoto *et al.*, 2003; Pinheiro *et al.*, 2005; Immel *et al.*, 2006; Alves *et al.*, 2004; Coudray *et al.*, 2007a,b,c; Forward *et al.*, 2008; Manco *et al.*, 2008; Brisighelli *et al.*, 2009; Lopes *et al.*, 2009; Gaibar *et al.*, 2010; Gusmão *et al.*, 2010; 2010; Alves *et al.*, 2011; Manta *et al.*, 2012).

5.2.9 D8S1179

Para este marcador genético, o alelo 14 (0,294) foi o de maior ocorrência na população em estudos e em Moçambique (Alves *et al.*, 2004), Namíbia (Muro *et al.*, 2008) e Tanzânia (Forward *et al.*, 2008), não sendo identificado como o mais frequente nas populações europeias analisadas. No entanto, foi o segundo mais frequente em Indígenas da Amazônia Brasileira (Terenas) (Manta *et al.*, 2012), na Alemanha (Reichenpader *et al.*, 2003), China (Gusmão *et al.*, 2010), Espanha (Andaluzia) e Sul da Espanha (Coudray *et al.*, 2007c; Gaibar *et al.*, 2010), Itália (Brisighelli *et al.*, 2009), Japão (Okamoto *et al.*, 2003), Marrocos –

Bouhria (Coudray *et al.*, 2007b) e Portugal (Lopes *et al.*, 2009; Amorim *et al.*, 2001; Gusmão *et al.*, 2000; Pinheiro *et al.*, 2005).

O alelo 13, segundo em número de observações no presente trabalho (frequência de 0,264), foi detectado como um alelo de alta frequência em diversas populações (Gusmão *et al.*, 2000; Amorim *et al.*, 2001; Abdin *et al.*, 2003; Okamoto *et al.*, 2003; Reichenpfader *et al.*, 2003; Alves *et al.*, 2004; Pinheiro *et al.*, 2005; Immel *et al.*, 2006; Coudray *et al.*, 2007a,b,c; Forward *et al.*, 2008; Muro *et al.*, 2008; Brisighelli *et al.*, 2009; Lopes *et al.*, 2009; Gaibar *et al.*, 2010; Gusmão *et al.*, 2010; Manta *et al.*, 2012).

5.2.10 D13S317

O alelo 12, pertencente ao *locus* D13S317 apresentou a maior frequência, no presente trabalho (0,335), e nas populações da Líbia (Immel *et al.*, 2006), Marrocos – Asni e Bouhria (Coudray *et al.*, 2007b), Moçambique (Alves *et al.*, 2004), Namíbia (Muro *et al.*, 2008) e Tanzânia (Forward *et al.*, 2008) Egito (Coudray *et al.*, 2007a) e foi o segundo mais frequente na Espanha – Andaluzia (Coudray *et al.*, 2007c), Itália (Brisighelli *et al.*, 2009), Centro, Norte e Nordeste de Portugal (Manco *et al.*, 2008; Lopes *et al.*, 2009; Amorim *et al.*, 2001; Gusmão *et al.*, 2000; Pinheiro *et al.*, 2005) e Síria (Abdin *et al.*, 2003).

Para o marcador D13S317, o segundo alelo mais frequente na população foi o 11 (0,298), que apresenta ampla distribuição em diversas populações (Gusmão *et al.*, 2000; Amorim *et al.*, 2001; Abdin *et al.*, 2003; Okamoto *et al.*, 2003; Alves *et al.*, 2004; Pinheiro *et al.*, 2005; Immel *et al.*, 2006; Coudray *et al.*, 2007a,b,c; Forward *et al.*, 2008; Manco *et al.*, 2008; Muro *et al.*, 2008; Brisighelli *et al.*, 2009; Lopes *et al.*, 2009; Gusmão *et al.*, 2010; Manta *et al.*, 2012).

Vale observar que, em populações asiáticas, o alelo 8 foi o mais recorrente (Okamoto *et al.*, 2003; Gusmão *et al.*, 2010). Para este mesmo alelo foi obtida uma frequência de 0,075, na população do Rio de Janeiro. Já o microvariante 10.2, o de maior frequência em Indígenas da Amazônia Brasileira (Terenas) (Manta *et al.*, 2012), não foi detectado no presente estudo.

5.2.11 D16S539

Em relação ao marcador genético D16S539, o alelo 11 apresentou a maior frequência na população estudada (0,284), o mesmo ocorreu em estudos com Indígenas da Amazônia Brasileira (Terenas) (Manta *et al.*, 2012), na Alemanha (Reichenpfader *et al.*, 2003), Egito (Coudray *et al.*, 2007a), Espanha – Andaluzia Coudray *et al.* (2007c), Itália (Brisighelli *et al.*, 2009), Líbia (Immel *et al.*, 2006), Marrocos – Bouhria (Coudray *et al.*, 2007b), Moçambique (Alves *et al.*, 2004), Centro de Portugal (Manco *et al.*, 2008) e Tanzânia (Forward *et al.*, 2008). Na Áustria (Reichenpfader *et al.*, 2003), Sul da Espanha (Gaibar *et al.*, 2010), Japão (Okamoto *et al.*, 2003), Namíbia (Muro *et al.*, 2008), Centro e Nordeste Portugal (Lopes *et al.*, 2009; Pinheiro *et al.*, 2005) e Síria (Abdin *et al.*, 2003), o alelo 11 foi o segundo maior em frequência observada.

Já o alelo 12, no mesmo marcador, foi o segundo mais frequente na população do Rio de Janeiro (0,235), apresentando frequência elevada também em Indígenas da Amazônia Brasileira (Terenas) (Manta *et al.*, 2012), na Alemanha e Áustria (Reichenpfader *et al.*, 2003), Egito (Coudray *et al.*, 2007a), Espanha (Andaluzia) e Sul da Espanha (Coudray *et al.*, 2007c; Gaibar *et al.*, 2010), Itália (Brisighelli *et al.*, 2009), Líbia (Immel *et al.*, 2006), Marrocos – Asni e Bouhria (Coudray *et al.*, 2007b), Centro e Nordeste de Portugal (Lopes *et al.*, 2009; Pinheiro *et al.*, 2005) e Síria (Abdin *et al.*, 2003), sendo, portanto, de ampla distribuição.

5.2.12 D18S51

Para o marcador D18S51, o alelo 15 foi o de maior frequência (0,170). Dentre as populações comparadas, este alelo apresentou maior frequência apenas na Alemanha (Reichenpfader *et al.*, 2003), China (Gusmão *et al.*, 2010) e Nordeste de Portugal (Gusmão *et al.*, 2000), e a segunda maior frequência apenas no Centro de Portugal (Lopes *et al.*, 2009), não sendo um alelo de alta frequência nas variadas populações.

Após o alelo 15, o 16 (0,156) apresentou a maior frequência alélica, também amplamente observado em estudos na Líbia (Immel *et al.*, 2006), Marrocos – Asni (Coudray *et al.*, 2007b), Moçambique (Alves *et al.*, 2004), Namíbia (Muro *et al.*, 2008), Norte e Nordeste de Portugal (Amorim *et al.*, 2001; Pinheiro *et al.*, 2005).

5.2.13 D21S11

No caso do *locus* D21S11, o alelo 30 apresentou a maior frequência (0,242). Nas populações de Indígenas da Amazônia Brasileira (Terenas) (Manta *et al.*, 2012), do Sul da Espanha (Gaibar *et al.*, 2010), Japão (Okamoto *et al.*, 2003), Líbia (Immel *et al.*, 2006), Centro, Norte e Nordeste de Portugal (Lopes *et al.*, 2009; Amorim *et al.*, 2001; Gusmão *et al.*, 2000; Pinheiro *et al.*, 2005), esse foi o alelo de maior ocorrência. Na Alemanha (Reichenpfader *et al.*, 2003), China (Gusmão *et al.*, 2010), Egito (Coudray *et al.*, 2007a), Espanha – Andaluzia (Coudray *et al.*, 2007c), Itália (Brisighelli *et al.*, 2009), Marrocos – Asni e Bouhria (Coudray *et al.*, 2007b) e Namíbia (Muro *et al.*, 2008), foi o segundo maior em observações.

O alelo 29, foi o segundo mais frequente para o mesmo marcador (0,216), similar ao ocorrido em diversas populações (Gusmão *et al.*, 2000; Amorim *et al.*, 2001; Abdin *et al.*, 2003; Okamoto *et al.*, 2003; Reichenpfader *et al.*, 2003; Alves *et al.*, 2004; Immel *et al.*,

2006; Coudray *et al.*, 2007a,b,c; Forward *et al.*, 2008; Brisighelli *et al.*, 2009; Lopes *et al.*, 2009; Gaibar *et al.*, 2010; Gusmão *et al.*, 2010).

5.2.14 D2S1338

Para o marcador genético D2S1338, o alelo 17 foi o mais frequente (0,182), assim como em estudos na Alemanha e Áustria (Reichenpfader *et al.*, 2003), Espanha (Andaluzia) e Sul da Espanha (Coudray *et al.*, 2007c; Gaibar *et al.*, 2010), Marrocos – Asni e Bouhria (Coudray *et al.*, 2007b), Centro de Portugal (Lopes *et al.*, 2009).

O alelo 19, segundo em frequência observada (0,152), foi frequente na Namíbia (Muro *et al.*, 2008) e Tanzânia (Forward *et al.*, 2008), não sendo um alelo muito frequente nas diversas regiões.

5.2.15 D19S433

No do *locus* D19S433, os alelos 13 e 14 apresentaram as maiores frequências (0,258 e 0,249, respectivamente). Estes também foram os alelos mais frequentes na Alemanha e Áustria (Reichenpfader *et al.*, 2003), Egito (Coudray *et al.*, 2007a), Espanha (Andaluzia) e Sul da Espanha (Coudray *et al.*, 2007c; Gaibar *et al.*, 2010), Marrocos – Asni e Bouhria (Coudray *et al.*, 2007b), Moçambique (Alves *et al.*, 2004), Namíbia (Muro *et al.*, 2008), Centro Portugal (Lopes *et al.*, 2009) e Tanzânia (Forward *et al.*, 2008).

5.2.16 Penta E

Em relação ao marcador genético Penta E, o alelo 12 foi o mais frequente (0,159), fato que ocorreu em trabalhos na Itália (Brisighelli *et al.*, 2009), Centro e Nordeste de Portugal

(Lopes *et al.*, 2009; Pinheiro *et al.*, 2005). Em Indígenas da Amazônia Brasileira (Terenas) (Manta *et al.*, 2012) e na Líbia (Immel *et al.*, 2006), foi o segundo mais frequente. Nas populações do Japão (Okamoto *et al.*, 2003) e da Líbia (Immel *et al.*, 2006), o alelo 11, segundo em frequência na população estudada (0,125), apresentou frequência elevada.

5.2.17 Penta D

O alelo 9, pertencente ao *locus* Penta D, apresentou a maior ocorrência (frequência de 0,182), coincidente com a população do Japão (Okamoto *et al.*, 2003). Na Itália (Brisighelli *et al.*, 2009), Líbia (Immel *et al.*, 2006), Centro e Nordeste de Portugal (Lopes *et al.*, 2009; Pinheiro *et al.*, 2005), este foi o segundo alelo mais frequente. O alelo 12, segundo em ocorrência na população do Rio de Janeiro (frequência de 0,166), apresentou a segunda maior frequência em Indígenas da Amazônia Brasileira (Terenas) (Manta *et al.*, 2012) e na Síria (Abdin *et al.*, 2003).

6. Conclusões

Nesta dissertação, foram analisados 17 marcadores do tipo STR, de interesse forense, na população do Rio de Janeiro. As principais conclusões referentes ao estudo são:

1. Os marcadores genéticos TH01 e D7S820 apresentaram o menor número de alelos;
2. O *locus* D21S11 foi o mais polimórfico, seguido do marcador genético FGA;
3. O alelo 8, do marcador genético TPOX, foi o mais frequentemente observado e o *locus* TPOX apresentou o menor Poder de Discriminação (PD), Poder de Exclusão (PE) e Conteúdo de Informação Polimórfica (PIC) e a maior Probabilidade de Coincidência Aleatória (PM), dentre os analisados;
4. No *locus* Penta E, foi observado o maior Poder de Discriminação (PD), Poder de Exclusão (PE) e Conteúdo de Informação Polimórfica (PIC) e a menor Probabilidade de Coincidência Aleatória (PM);
5. O marcador genético D2S1338 apresentou o segundo maior Poder de Discriminação (PD), Conteúdo de Informação Polimórfica (PIC) e a segunda menor Probabilidade de Coincidência Aleatória (PM), enquanto o *locus* D18S51 apresentou o segundo maior Poder de Exclusão (PE);
6. Foram observadas frequências ainda não descritas, na população do Rio de Janeiro, para o *locus* Penta D e para os seguintes alelos: 13, 19.2, 30.2, 31.2 e 32 (FGA); 5 (TH01); 5 (TPOX); 22 (vWA); 18 (D8S1179); 5 e 12.1 (D16S539); 24 (D18S51); 19, 20, 23.2, 35.2 e 37 (D21S11);

7. O teste exato para a hipótese de Equilíbrio de Hardy-Weinberg demonstrou que os marcadores genéticos estudados não apresentaram desvio significativo em relação ao equilíbrio;
8. O valor médio do coeficiente de endogamia de Wright (F_{IS}) foi de 0,0196;
9. A árvore filogenética, desenvolvida pelo método de Neighbor-Joining, possibilitou detectar uma semelhança entre população do Rio de Janeiro e a população do Nordeste brasileiro;
10. Não foram observados valores estatisticamente significativos na análise de F_{ST} par a par entre a população do Rio de Janeiro e as populações das demais regiões brasileiras, indicando ausência de estruturação, importante característica no caso de marcadores genéticos utilizados em testes de identificação humana;
11. O gráfico “Triangle Plot” populacional, gerado no *software* Structure, sugere que, em média, os indivíduos não apresentam contribuição maior determinado grupo étnico em comparação com os demais, podendo a população ser formada por similar contribuição dos grupos étnicos principais;
12. A comparação das frequências alélicas, dos marcadores genéticos estudados, permitiu observar que, de forma geral, os alelos frequentemente observados na população do Rio de Janeiro apresentaram ampla distribuição nas diversas populações mundiais.

7. Referências

- Abdin, L.; Shimadaa, I.; Brinkmanna, B.; Hohoff, C. (2003) Analysis of 15 short tandem repeats reveals significant differences between the Arabian populations from Morocco and Syria. *Legal Medicine* 5: S150-S155.
- Adam, C. (2010) *Essential Mathematics and Statistics for Forensic Science*. John Wiley & Sons, Grã Bretanha, 1ª Ed, 354 p.
- Alves, C.; Gusmão, L.; Damascenob, 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 Science International* 139: 201-205.
- Alves, C.; Gusmão, L.; Pereira, L.; Amorim, A. (2011) STR data (CD4, CSF1PO, F13A01, FES/FPS, MBPB, TH01, TPOX) from North Portugal. *Forensic Science International* 123: 76-77.
- Amorim, A.; Gusmão, L.; Alves, C. (2001) STR data (AmpFISTR Pro®ler Plus) from north Portugal. *Forensic Science International* 115: 119-121.
- Barber, M.D.; McKeown, B.J.; Parkin, B.H. (1996) Structural variation in the alleles of a short tandem repeat system at the human alpha fibrinogen locus. *International Journal of Legal Medicine* 108 (4): 180–185.
- Barber, M.D.; Parkin, B.H. (1996) Sequence analysis and allelic designation of the two short tandem repeat loci D18S51 and D8S1179. *International Journal of Legal Medicine* 109 (2): 62-65.
- Barber, P. (2004) Extraction Protocol: Chelex.
<http://www.eeb.ucla.edu/Faculty/Barber/Web%20Protocols/Protocol2.pdf>.

- Behar, D. B.; Rosset, S.; Blue-Smith, J.; Balanovsky, O.; Tzur, S.; Comas, D.; Mitchell, R. J.; Quintana-Murci, L.; Tyler-Smith, C.; Wells, R. S.; The Genographic Consortium (2007) The Genographic Project Public Participation Mitochondrial DNA Database. *PLoS Genetics* 3 (6): 1083-1095.
- Botstein, D.; White, R.L.; Skolnick, M.; Davis, R.W. (1980) Construction of a Genetic Linkage Map in Man Using Restriction Fragment Length Polymorphisms. *American Journal of Human Genetics* 32: 314–331.
- Brinkmann, B.; Meyer, E.; Junge, A. (1996) Complex mutational events at the HumD21S11 locus. *Human Genetics* 98: 60-64.
- Brisighelli, F.; Capelli, C.; Boschi, I.; Garagnani, P.; Lareu, M. V.; Pascali, V. L.; Carracedo, A. (2009) Allele frequencies of fifteen STRs in a representative sample of the Italian population. *Forensic Science International: Genetics* 3 (2): e29-e30.
- Buckleton, J.; Triggs, C. M.; Walsh, S.J. (2005) *Forensic DNA Evidence Interpretation*. CRC Press, 534 p.
- Budowle B.; Moretti T.R.; Niezgoda S.J.; Brown B.L. (1998). CODIS and PCR-based short tandem repeat loci: law enforcement tools. *Proceedings of the Second European Symposium on Human Identification*. Madison, WI: Promega Corporation. 1998, 73-88.
- Butler, J. M. (2005) *Forensic DNA typing: biology & technology behind STR markers*. Elsevier Academic Press, 2^a Ed., 660 p.
- Butler, J. M. (2006). *Genetics and Genomics of Core Short Tandem Repeat Loci Used in Human Identity Testing*. *J Forensic Sci* 51 (2): 253-265.

- Butler, J. M. (2007) Short tandem repeat typing technologies used in human identity testing
BioTechniques 43 (4): Sii–Sv.
- Chula, F. G. L.; Rodenbusch, R.; Schumacher, S.; Grandi, T.; Michelon, C. T.; Gastaldo, A.
Z.; Costi, C.; Carvalho, B.; Silva, C. M. D. (2009) 15 STR loci frequencies with mutation
rates in the population from Rio Grande do Sul, Southern Brazil. Forensic Science
International: Genetics 3: e35-e38.
- Clarke, R. (1994) Human Identification in Information Systems: Management Challenges and
Public Policy Issues. Information Technology e People 7: 4, 6-37.
- Coudray, C.; Guitard, E. ; El-Chennawi, F. ; Larrouy, G. ; Dugoujon, JM. (2007a) Allele
frequencies of 15 short tandem repeats (STRs) in three Egyptian populations of different
ethnic groups Forensic Science International 169: 260-265.
- Coudray, C.; Guitard, E. ; Keyser-Tracqui, C. ; Melhaoui, M. ; Cherkaoui, M.; Larrouy, G. ;
Dugoujon, JM. (2007b) Population genetic data of 15 tetrameric short tandem repeats
(STRs) in Berbers from Morocco Forensic Science International 167: 81-86.
- Coudray, C.; Calderon, R.; Guitard, E.; Ambrosio, B.; González-Martín, A.; Dugoujon, JM.
(2007c) Allele frequencies of 15 tetrameric short tandem repeats (STRs) in Andalusians
from Huelva (Spain). Forensic Science International 168: e21-e24.
- Ellegren, H. (2004) Microsatellites: simple sequences with complex evolution. Nature
Reviews Genetics 5: 435–445.
- Espíndula, A. (2006) Perícia Criminal e Cível. Uma visão completa para peritos e usuários da
perícia. Millennium, 2ª Ed, 480 p.

- Excoffier, L.; Laval, G.; Schneider, S. (2005). Arlequin ver. 3.0: an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* 1: 47-50.
- Fausto, B. (2000) *Fazer a América: a imigração em massa para a América Latina*. São Paulo, Ed. USP, 2ª ed.
- Francez, P. A. C.; Rodrigues, E. M. R. R.; Frazão, G. F.; Borges, N. D. R.; Santos, S. E. B. (2011) Allelic frequencies and statistical data obtained from 12 codis STR loci in an admixed population of the Brazilian Amazon. *Genetics and Molecular Biology* 34 (1): 35-39.
- Forward, B. W.; Eastman, M. W.; Nyambo, T. B.; Ballard, R. E. (2008) AMPFISTR®R Identifiler™ STR Allele Frequencies in Tanzania, Africa. *J Forensic Sci* 53 (1): 245-247.
- Freedman, M. L.; Reich, D.; Penney, K. L.; McDonald, G. J.; Mignault, A. A.; Patterson, N.; Gabriel, S. B.; Topol, E. J.; Smoller, J. W.; Pato, C. N.; Pato, M. T.; Petryshen, T. L.; Kolonel, L. N.; Lander, E. S.; Sklar, P.; Henderson, B.; Hirschhorn, J. N.; Altshule, D. (2004) Assessing the impact of population stratification on genetic association studies. *Nature Genetics* 36: 388-393.
- Gaibar, M.; Esteban, E.; Moral, P.; Gómez-Gallego, F.; Santiago, C.; Bandrés, F.; Luna, F.; Fernández-Santander, F. (2010) STR genetic diversity in a Mediterranean population from the south of the Iberian Peninsula. *Annals of Human Biology* 37 (2): 254-267.
- Garrido, R. G. (2009) *Evolução dos Processos de Identificação Humana: das características antropométricas ao DNA*. *Genética na Escola* 5 (2): 38-40.

- Gattapaglia, D.; Schmidt, A. B.; Costa e Silva, C.; Stringher, C.; Fernandes, A. P.; Ferreira, M. E. (2001) Brazilian population database for the 13 STR loci of the AmpFI STR Profiler Plus and Cofiler multiplex kits. *Forensic Science International* 118: 91-94.
- Gill, P.; Kimpton, C.P.; d'Aloja, E.; Andersen, J.F.; Bar, W.; Brinkmann, B.; Holgersson, S.; Johnsson, V.; Kloosterman, A.D.; Lareu, M.V.; Nellesmann, L.; Pfitzinger, H.; Phillips, C.P.; Schmitter, H.; Schneider, P.M.; Stenersen, M. (1994) Report of the European DNA profiling group (EDNAP) – towards standardisation of short tandem repeat (STR) loci. *Forensic Science International* 65 (1): 51–59.
- Glaubitz J.C. (2004) CONVERT: A user-friendly program to reformat diploid genotypic data for commonly used population genetic software packages. *Molecular Ecology Notes* 4: 309-310.
- Góes, A. C. S.; Silva, D. A.; Gila, E. H. F.; Silva, M. T. D.; Pereirab, R. W.; Carvalho, E. F. (2004) Allele frequencies data and statistic parameters for 16 STR loci - D19S433, D2S1338, CSF1PO, D16S539, D7S820, D21S11, D18S51, D13S317, D5S818, FGA, Penta E, TH01, vWA, D8S1179, TPOX, D3S1358 - in the Rio de Janeiro population, Brazil. *Forensic Science International* 140: 131–132.
- Griffiths, A. J. F.; Wessler, S. R.; Lewontin, R. C.; Carroll, S. B. (2004) *An Introduction to Genetic Analysis*. W. H. Freeman, Eighth Edition, 800 p.
- Gusmão, L.; Prata, M. J.; Sánchez-Diz, P.; Lareu, M. V.; Carracedo, A.; Alves, C.; Martins, N.; Amorim, A. (2010) STR data for the AmpFI STR profiler plus loci from Macau (China). *Forensic Science International* 123: 74-75.

- Gusmão, L.; Sánchez-Diz, P.; Alves, C.; Lareu, M. V.; Carracedo, A.; Amorim, A. (2000) Genetic diversity of nine STRs in two northwest Iberian populations: Galicia and northern Portugal. *Int J Legal Med* 114: 109-113.
- Hammond, H.A.; Jin, L.; Zhong, Y.; Caskey, C.T.; Chakraborty, R. (1994) Evaluation of 13 short tandem repeat loci for use in personal identification applications. *American Journal of Human Genetics* 55: 175-189.
- Hares D.R. (2012). Expanding the CODIS core loci in the United States Forensic. *Sci Int Genet.* 6 (1): 52-4.
- Helyar, S.J. (2011) Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges. *Molecular Ecology Resources* 11: 123-136.
- http://marketing.appliedbiosystems.com/mk/get/IDPLUS_COMPARISON_ID_KIT, acesso em 28/05/2012.
- <http://www.biocompare.com/Articles/ProductReview/395/Applied-Biosystems-3130-and-3130xl-Genetic-Analyzers.html>, acesso em 29/05/2012.
- IBGE – Instituto Brasileiro de Geografia e Estatística (2000). *Brasil: 500 anos de povoamento*. IBGE. Rio de Janeiro.
- IBGE – Instituto Brasileiro de Geografia e Estatística (2010) *Síntese de indicadores sociais: uma análise das condições de vida da população brasileira 2010*. IBGE. Rio de Janeiro.
- Immel, U. D.; Erhuma, M.; Mustafa, T.; Kleiber, M.; Klintschar, M. (2006) Population genetic analysis in a Libyan population using the PowerPlex 16 system. *International Congress Series* 1288: 421- 423.

- International Human Genome Sequencing Consortium (2001). Initial sequencing and analysis of the human genome. *Nature* 409: 860–921.
- International Human Genome Sequencing Consortium (2004) Finishing the euchromatic sequence of the human genome. *Nature* 431: 931-945.
- James, S. H.; Nordby, J. J. (2005) *Forensic science: An introduction to scientific and investigative techniques*. CRC Press, 2^a Ed, 778 p.
- Jeffreys, A. J.; Wilson, V.; Thein, S. L. (1985a) Individual-specific 'fingerprints' of human DNA. *Nature* 316: 76-79.
- Jeffreys, A. J.; Wilson, V.; Thein, S. L. (1985b) Hypervariable 'minisatellite' regions in human DNA. *Nature* 314: 67-73.
- Jobling, M. A.; Gill, P. (2004) Encoded evidence: DNA in forensic analysis. *Nature Reviews Genetics* 5: 739-751.
- Kimpton, C.; Walton, A.; Gill, P. (1992) A further tetranucleotide repeat. *Human Molecular Genetics* 1 (4), 287.
- Lewis, P. O.; Zaykin, D. (2002) GDA - Genetic Data Analysis: version 1.1 for Windows 95/NT. Disponível em: <http://www.eeb.uconn.edu/people/plewis/software.php>.
- Life Technologies Corporation (2007) AmpFISTR® MiniFiler™ PCR Amplification Kit User Guide. Printed in the U.S.A.
- Life Technologies Corporation (2012) AmpFISTR® Identifiler® Plus PCR Amplification Kit User Guide. Printed in the U.S.A.

- Lopes, V.; Serra, A.; Gamero, J.; Sampaio, L.; Balsa, F.; Oliveira, C.; Batista, L.; Corte-Real, F.; Vieira, D. N.; Vide, M. C.; Anjos, M. J.; Carvalho, M. (2009) Allelic frequency distribution of 17 STRs from Identifiler and PowerPlex-16 in Central Portugal area and the Azores archipelago. *Forensic Science International: Genetics* 4: e1-e7.
- Lopez-Camelo, J. S.; Cabello, P. H.; Dutra, M. G. (1996) A simple model for the estimation of congenital malformation frequency in racially mixed populations. *Brazilian Journal of Genetics*, 19 (4): 659-663.
- Manco, L.; Arede, P.; Silva, I.; Alvarez, M.; Abade, A. (2008) Estudo populacional de 6 STRs (TH01, TPOX, CSF1PO, D7S820, D13S317 e D16S539) na região Centro de Portugal. *Antropologia Portuguesa* 24/25: 141-148.
- Manta, F. S. N.; Silva, D. A.; Sánchez-Diz, P.; Aguiar, J. I. A.; Cabello, P. H.; Carvalho, E. F. (2012) Terena Amerindian group autosomal STR data: comparison studies with other Brazilian populations. *Mol Biol Rep* 39 :4455-4459.
- Margolis-Nunno, H.; Brenner, L.; Cascardi, J.; Kobilinsky, L. (2001) *Journal of Forensic Sciences* 46: 1480–1483.
- Mendes-Junior, C. T. (2001) Frequências Alélicas de Polimorfismos de DNA do Tipo STR em Indígenas da Amazônia Brasileira. Dissertação de Mestrado. Universidade de São Paulo, Ribeirão Preto.
- Miller, S.A.; Dykes, D. D.; Polesky, H.F. (1988) A Simple Salting Out Procedure for Extracting DNA from Human Nucleated Cells. *Nucleic Acids Research* 6: 1215.

- Moroni, R.; Gasbarra, D.; Arjas, E.; Lukka, M.; Ulmanen, I. (2011) Effects of Reference Population and Number of STR Markers on positive evidence in Paternity Testing. *J Forensic Res* 2: 119.
- Muro, T.; Fujihara, J.; Imamura, S.; Nakamura, H.; Yasuda, T.; Takeshita, H. (2008) Allele frequencies for 15 STR loci in Ovambo population using AmpFISTR® Identifiler Kit. *Legal Medicine* 10: 157-159.
- National Research Council. Committee on DNA Forensic Science: An Update (1996). The Evaluation of Forensic DNA Evidence. National Academies Press, Washington, D.C., 254 p.
- Nei, M., (1977). F-statistics and analysis of gene diversity in subdivided populations. *Annals of Human Genetics* 41, 225-233.
- Okamoto, O.; Yamamoto, Y.; Inagaki, S.; Yoshitome, K.; Ishikawa, T.; Imabayashi, K.; Miyaishi, S.; Ishizu, H. (2003) Analysis of Short Tandem Repeat (STR) Polymorphisms by the Power Plex 16 System and Capillary Electrophoresis: Application to Forensic Practice. *Acta. Med Okayama* 52 (2): 59-71.
- Page, R. D. (1996) TreeView: An application to display phylogenetic trees on personal computers. *Comput Appl Biosci* 12: 357-358.
- Patton, S. M. (1990) DNA Fingerprinting: the Castro Case. *Harvard Journal of Law & Technology* 3: 223-240.
- Pritchard, J. K.; Stephens, M.; Donnelly, P. (2000) Inference of Population Structure Using Multilocus Genotype Data. *Genetics* 155: 945–959.

- Pinheiro, M. F.; Cainé, L.; Pontes, L.; Abrantes, D.; Lima, G.; Pereira, M. J.; Rezende, P. (2005) Allele frequencies of sixteen STRs in the population of Northern Portugal. *Forensic Science International* 148: 221-223.
- Polymeropoulos, M.H.; Xiao, H.; Rath, D.S.; Merrill, C.R. (1991) Tetranucleotide repeat polymorphism at the human tyrosine hydroxylase gene (TH). *Nucleic Acids Research* 19: 3753.
- Promega Corporation. PowerPlex® 16 HS System Technical Manual. Printed in USA.
- Raymond M.; Rousset F, 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *J. Heredity* 86: 248-249.
- Regateiro, F. J. (2007) *Manual de Genética Médica*. Imprensa da Universidade de Coimbra: 1ª Ed., 2ª reimpressão, 496 p.
- Reichenpfer, B.; Immelb, U.; Klitscharb, M. (2003) Population data on the AmpFI STR SGM plus PCR amplification kit in Germans and Austrians. *Forensic Science International* 132: 84-86.
- Ribeiro, D. (2006) *O povo brasileiro: A formação e o sentido do Brasil*. Companhia das Letras, 10ª reimpressão, 435 p.
- Rousset, F., 2008. Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. *Mol. Ecol. Resources* 8: 103-106.
- Saiki, R. K.; Gelfand, D. H ; Stoffel, S.; Scharf, S. J.; Higuchi, R.; Horn, G. T.; Mullis, K. B.; Erlich, H. A. (1988) Primer-Directed Enzymatic Amplification of DNA with a Thermostable DNA Polymerase. *Science* 239: 487-491.

- Santos, N. P. C.; Ribeiro-Rodrigues, E. M.; Ribeiro-dos-Santos, A. K. C.; Pereira, R.; Gusmão, L.; Amorim, A.; Guerreiro, J. F.; Zago, M. A.; Matte, C.; Hutz, M. H.; Santos, S. E.B. (2010) Assessing individual interethnic admixture and population substructure using a 48–insertion-deletion (INSEL) ancestry-informative marker (AIM) panel. *Hum. Mutat.* 31: 184–190.
- Straub, R. E.; Speer, M.C.; Luo, Y.; Rojas, K.; Overhauser, J.; Ott, J.; Gilliam, T. C. (1993) A microsatellite genetic linkage map for human chromosome 18. *Genomics* 15: 48-56.
- STRBase: <http://www.cstl.nist.gov/strbase>
- Tautz, D. (1993) Notes on definition and nomenclature of tandemly repetitive DNA sequences. In Pena, D.J., Chakraborty, R., Eppelen, J.T. and Jeffreys, A.J. (eds), *DNA Fingerprinting: State of the Science*, Birkhauser Verlag: Basel: 21–28.
- Walsh, P. S.; Fildes, N. J.; Reynolds, R. (1996) Sequence analysis and characterization of stutter products at the tetranucleotide repeat locus vWA. *Nucleic Acids Research* 24 (14): 2807-2812.
- Wambaugh, J. (1989) *The Blooding*. New York, Bantam Books; <http://www.forensic.gov.uk>.
- Wright, S. (1965) The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution* 19: 395-420.
- Wyman, A. R.; White, R. (1980) A highly polymorphic locus in human DNA. *Genetics* 77: 6754-6758.
- Guo, X.; Elston, R. C. (1999) Linkage Information Content of Polymorphic Genetic Markers. *Hum Hered* 49: 112-118.

Yamaguchi-Kabata, Y.; Nakazono, K.; Takahashi, A.; Saito, S.; Hosono, N.; Kubo, M.; Nakamura, Y.; Kamatani, N. (2008) Japanese population structure, based on SNP genotypes from 7003 individuals compared to other ethnic groups: effects on population-based association studies. *The American Journal of Human Genetics* 83 (4): 445-456.

8. ANEXOS

CSF1PO			
População	Alelos mais Frequentes	Frequência	Referência
Rio de Janeiro	12	0,312	População Estudada
	11	0,276	
Egito - Andaima	12	0,323	Coudray et al., 2007a
	10	0,318	
Espanha - Andaluzia	10	0,338	Coudray et al., 2007c
	11	0,286	
Sul da Espanha	12	0,286	Gaibar <i>et al.</i> , 2010
	11	0,361	
Indígenas da Amazônia Brasileira	10	0,278	Mendes-Junior, 2001
	12	0,393	
Itália	10	0,284	Brisighelli <i>et al.</i> , 2009
	12	0,325	
Japão	11	0,317	Okamoto <i>et al.</i> , 2003
	12	0,427	
Líbia	11	0,203	Immel <i>et al.</i> , 2006
	12	0,330	
Marrocos - Asni	10	0,316	Coudray <i>et al.</i> , 2007b
	12	0,338	
Marrocos - Bouhria	10	0,300	Coudray <i>et al.</i> , 2007b
	11	0,370	
Moçambique	12	0,288	Alves <i>et al.</i> , 2004
	10	0,349	
Namíbia	10	0,275	Muro <i>et al.</i> , 2008
	11	0,269	
Centro Portugal	11	0,264	Lopes <i>et al.</i> , 2009
	12	0,334	
Nordeste Portugal	12	0,325	Pinheiro <i>et al.</i> , 2005
	10	0,325	
Norte Portugal	10	0,305	Alves <i>et al.</i> , 2011
	12	0,315	
Síria	10	0,309	Abdin <i>et al.</i> , 2003
	12	0,339	
Tanzânia	11	0,311	Forward <i>et al.</i> , 2008
	10	0,288	
	12	0,279	

Tabela 8.1: Alelos mais frequentes, no *loci* CSF1PO, em diferentes populações.

FGA			
População	Alelos mais Frequentes	Frequência	Referência
Rio de Janeiro	22	0,170	População Estudada
	21	0,163	
Alemanha	22	0,212	Reichenpfader <i>et al.</i> , 2003
	21	0,175	
China	22	0,224	Gusmão <i>et al.</i> , 2010
	23	0,184	
Egito - Andaima	23	0,182	Coudray <i>et al.</i> , 2007a
	21	0,157	
Espanha - Andaluzia	21	0,272	Coudray <i>et al.</i> , 2007c
	24	0,154	
Sul da Espanha	22	0,183	Gaibar <i>et al.</i> , 2010
	24	0,172	
Indígenas da Amazônia Brasileira (Terenas)	24	0,350	Manta <i>et al.</i> , 2012
	25	0,205	
Itália	22	0,189	Brisighelli <i>et al.</i> , 2009
	21	0,167	
Japão	22	0,240	Okamoto <i>et al.</i> , 2003
	24	0,178	
Líbia	22	0,191	Immel <i>et al.</i> , 2006
	24	0,167	
Marrocos - Asni	22	0,195	Coudray <i>et al.</i> , 2007b
	23	0,176	
Marrocos - Bouhria	21	0,226	Coudray <i>et al.</i> , 2007b
	24	0,163	
Moçambique	24	0,191	Alves <i>et al.</i> , 2004
	23	0,188	
Centro Portugal	21	0,193	Lopes <i>et al.</i> , 2009
	22	0,178	
Norte de Portugal	21	0,190	Amorim <i>et al.</i> , 2001
	23	0,173	
Nordeste Portugal	23	0,188	Gusmão <i>et al.</i> , 2000
	22	0,168	
Nordeste Portugal	22	0,180	Pinheiro <i>et al.</i> , 2005
	21	0,170	
Namíbia	22	0,185	Muro <i>et al.</i> , 2008
	24	0,177	
Síria	24	0,279	Abdin <i>et al.</i> , 2003
	23	0,172	
	21	0,172	

Tabela 8.2: Alelos mais frequentes, no *loci* FGA, em diferentes populações.

TH01			
População	Alelos mais Frequentes	Frequência	Referência
Rio de Janeiro	7	0,261	População Estudada
	9	0,192	
Alemanha	9.3	0,331	Reichenpfader <i>et al.</i> , 2003
	6	0,226	
Egito - Andaima	9	0,343	Coudray <i>et al.</i> , 2007a
	7	0,268	
Espanha - Andaluzia	9	0,250	Coudray <i>et al.</i> , 2007c
	9.3	0,224	
Sul da Espanha	6	0,256	Gaibar <i>et al.</i> , 2010
	9	0,211	
Indígenas da Amazônia Brasileira (Terenas)	9.3	0,211	Manta <i>et al.</i> , 2012
	7	0,487	
Indígenas da Amazônia Brasileira	6	0,043	Mendes-Junior, 2001
	7	0,446	
Itália	6	0,269	Brisighelli <i>et al.</i> , 2009
	9	0,222	
Japão	9	0,415	Okamoto <i>et al.</i> , 2003
	7	0,285	
Líbia	9	0,267	Immel <i>et al.</i> , 2006
	6	0,243	
Marrocos - Asni	7	0,295	Coudray <i>et al.</i> , 2007b
	9	0,248	
Marrocos - Bouhria	9	0,293	Coudray <i>et al.</i> , 2007b
	6	0,207	
Moçambique	7	0,342	Alves <i>et al.</i> , 2004
	8	0,299	
Namíbia	7	0,400	Muro <i>et al.</i> , 2008
	8	0,364	
Centro Portugal	9.3	0,293	Manco <i>et al.</i> , 2008
	9	0,231	
Centro Portugal	9.3	0,328	Lopes <i>et al.</i> , 2009
	6	0,210	
Nordeste Portugal	9.3	0,285	Pinheiro <i>et al.</i> , 2005
	6	0,202	
Norte de Portugal	9.3	0,267	Alves <i>et al.</i> , 2011
	6	0,212	
Síria	6	0,267	Abdin <i>et al.</i> , 2003
	9	0,254	
Tanzânia	7	0,366	Forward <i>et al.</i> , 2008
	8	0,233	

Tabela 8.3: Alelos mais frequentes, no *loci* TH01, em diferentes populações.

TPOX			
População	Alelos mais Frequentes	Frequência	Referência
Rio de Janeiro	8	0,432	População Estudada
	11	0,261	
Egito - Andaima	8	0,424	Coudray et al., 2007a
	11	0,258	
Espanha - Andaluzia	8	0,482	Coudray et al., 2007c
	11	0,294	
Sul da Espanha	8	0,483	Gaibar <i>et al.</i> , 2010
	11	0,289	
Indígenas da Amazônia Brasileira	11	0,470	Mendes-Junior, 2001
	12	0,294	
Itália	8	0,537	Brisighelli <i>et al.</i> , 2009
	11	0,254	
Japão	8	0,483	Okamoto <i>et al.</i> , 2003
	11	0,308	
Líbia	8	0,500	Immel <i>et al.</i> , 2006
	11	0,250	
Marrocos - Asni	8	0,438	Coudray <i>et al.</i> , 2007b
	11	0,267	
Marrocos - Bouhria	8	0,447	Coudray <i>et al.</i> , 2007b
	11	0,365	
Namíbia	8	0,336	Muro <i>et al.</i> , 2008
	11	0,315	
Centro Portugal	8	0,552	Manco <i>et al.</i> , 2008
	11	0,294	
Centro Portugal	8	0,498	Lopes <i>et al.</i> , 2009
	11	0,303	
Nordeste Portugal	8	0,528	Pinheiro <i>et al.</i> , 2005
	11	0,273	
Norte de Portugal	8	0,246	Alves <i>et al.</i> , 2011
	11	0,335	
Síria	8	0,450	Abdin <i>et al.</i> , 2003
	11	0,248	
Tanzânia	9	0,284	Forward <i>et al.</i> , 2008
	11	0,257	

Tabela 8.4: Alelos mais frequentes, no *loci* TPOX, em diferentes populações.

vWA				
População	Alelos mais Frequentes	Frequência	Referência	
Rio de Janeiro	17	0,250	População Estudada	
	16	0,234		
Alemanha	17	0,278	Reichenpfader <i>et al.</i> , 2003	
	18	0,225		
China	14	0,299	Gusmão <i>et al.</i> , 2010	
	17	0,230		
Egito - Andaima	17	0,328	Coudray <i>et al.</i> , 2007a	
	16	0,242		
Espanha - Andaluzia	16	0,272	Coudray <i>et al.</i> , 2007c	
	17	0,254		
Sul da Espanha	17	0,289	Gaibar <i>et al.</i> , 2010	
	16	0,283		
Indígenas da Amazônia Brasileira (Terenas)	16	0,509	Manta <i>et al.</i> , 2012	
	17	0,252		
Itália	17	0,286	Brisighelli <i>et al.</i> , 2009	
	16	0,237		
Japão	17	0,277	Okamoto <i>et al.</i> , 2003	
	18	0,209		
Líbia	17	0,291	Immel <i>et al.</i> , 2006	
	16	0,214		
Marrocos - Asni	15	0,257	Coudray <i>et al.</i> , 2007b	
	17	0,257		
Marrocos - Bouhria	17	0,269	Coudray <i>et al.</i> , 2007b	
	18	0,212		
Moçambique	16	0,306	Alves <i>et al.</i> , 2004	
	15	0,236		
Namíbia	16	0,218	Muro <i>et al.</i> , 2008	
	15	0,200		
Centro Portugal	17	0,253	Lopes <i>et al.</i> , 2009	
	16	0,246		
Nordeste Portugal	17	0,277	Pinheiro <i>et al.</i> , 2005	
	16	0,248		
Norte de Portugal	16	0,253	Amorim <i>et al.</i> , 2001	
	17	0,234		
Síria	17	0,331	Abdin <i>et al.</i> , 2003	
	16	0,269		
Tanzânia	16	0,262	Forward <i>et al.</i> , 2008	
	15	0,242		

Tabela 8.5: Alelos mais frequentes, no *loci* vWA, em diferentes populações.

D3S1358				
População	Alelos mais Frequentes	Frequência	Referência	
Rio de Janeiro	15	0,307	População Estudada	
	16	0,284		
Alemanha	16	0,244	Reichenpfader <i>et al.</i> , 2003	
	15	0,235		
China	15	0,351	Gusmão <i>et al.</i> , 2010	
	16	0,322		
Egito - Andaima	17	0,293	Coudray <i>et al.</i> , 2007a	
	16	0,247		
Espanha - Andaluzia	17	0,241	Coudray <i>et al.</i> , 2007c	
	18	0,211		
Sul da Espanha	15	0,233	Gaibar <i>et al.</i> , 2010	
	18	0,233		
Indígenas da Amazônia Brasileira (Terenas)	15	0,608	Manta <i>et al.</i> , 2012	
	16	0,220		
Itália	16	0,257	Brisighelli <i>et al.</i> , 2009	
	15	0,240		
Japão	15	0,384	Okamoto <i>et al.</i> , 2003	
	16	0,305		
Líbia	15	0,320	Immel <i>et al.</i> , 2006	
	17	0,352		
Marrocos - Asni	16	0,271	Coudray <i>et al.</i> , 2007b	
	17	0,281		
Marrocos - Bouhria	16	0,361	Coudray <i>et al.</i> , 2007b	
	17	0,279		
Moçambique	16	0,380	Alves <i>et al.</i> , 2004	
	15	0,303		
Namíbia	16	0,372	Muro <i>et al.</i> , 2008	
	15	0,274		
Centro Portugal	15	0,258	Lopes <i>et al.</i> , 2009	
	16	0,202		
Nordeste Portugal	15	0,325	Gusmão <i>et al.</i> , 2000	
	16	0,230		
Nordeste Portugal	15	0,245	Pinheiro <i>et al.</i> , 2005	
	17	0,230		
Norte de Portugal	15	0,289	Amorim <i>et al.</i> , 2001	
	16	0,241		
Síria	15	0,277	Abdin <i>et al.</i> , 2003	
	16	0,277		
Tanzânia	16	0,427	Forward <i>et al.</i> , 2008	
	15	0,233		

Tabela 8.6: Alelos mais frequentes, no *loci* D3S1358, em diferentes populações.

D5S818			
População	Alelos mais Frequentes	Frequência	Referência
Rio de Janeiro	12	0,347	População Estudada
	11	0,300	
China	11	0,374	Gusmão <i>et al.</i> , 2010
	10	0,213	
Egito - Andaima	11	0,318	Coudray <i>et al.</i> , 2007a
	12	0,268	
Espanha - Andaluzia	11	0,377	Coudray <i>et al.</i> , 2007c
	12	0,355	
Sul da Espanha	12	0,394	Gaibar <i>et al.</i> , 2010
	11	0,356	
Itália	12	0,382	Brisighelli <i>et al.</i> , 2009
	11	0,272	
Japão	11	0,319	Okamoto <i>et al.</i> , 2003
	12	0,254	
Líbia	12	0,325	Immel <i>et al.</i> , 2006
	11	0,257	
Marrocos - Asni	12	0,371	Coudray <i>et al.</i> , 2007b
	11	0,257	
Marrocos - Bouhria	12	0,361	Coudray <i>et al.</i> , 2007b
	13	0,288	
Moçambique	12	0,370	Alves <i>et al.</i> , 2004
	11	0,236	
Namíbia	13	0,379	Muro <i>et al.</i> , 2008
	12	0,313	
Centro Portugal	12	0,379	Lopes <i>et al.</i> , 2009
	11	0,354	
Nordeste Portugal	12	0,381	Gusmão <i>et al.</i> , 2000
	11	0,352	
Nordeste Portugal	12	0,403	Pinheiro <i>et al.</i> , 2005
	11	0,335	
Norte de Portugal	12	0,374	Amorim <i>et al.</i> , 2001
	11	0,346	
Síria	12	0,372	Abdin <i>et al.</i> , 2003
	11	0,281	
Tanzânia	12	0,380	Forward <i>et al.</i> , 2008
	13	0,207	

Tabela 8.7: Alelos mais frequentes, no *loci* D5S818, em diferentes populações.

D7S820			
População	Alelos mais Frequentes	Frequência	Referência
Rio de Janeiro	10	0,300	População Estudada
	11	0,209	
China	11	0,368	Gusmão <i>et al.</i> , 2010
	12	0,201	
Egito - Andaima	10	0,409	Coudray <i>et al.</i> , 2007a
	11	0,197	
Espanha - Andaluzia	10	0,303	Coudray <i>et al.</i> , 2007c
	11	0,232	
Indígenas da Amazônia Brasileira (Terenas)	11	0,332	Manta <i>et al.</i> , 2012
	12	0,263	
Itália	10	0,300	Brisighelli <i>et al.</i> , 2009
	11	0,234	
Japão	11	0,316	Okamoto <i>et al.</i> , 2003
	10	0,229	
Sul da Espanha	11	0,272	Gaibar <i>et al.</i> , 2010
	12	0,189	
Líbia	10	0,359	Immel <i>et al.</i> , 2006
	11	0,233	
Marrocos - Asni	11	0,348	Coudray <i>et al.</i> , 2007b
	10	0,314	
Marrocos - Bouhria	10	0,313	Coudray <i>et al.</i> , 2007b
	11	0,279	
Moçambique	10	0,391	Alves <i>et al.</i> , 2004
	11	0,204	
Namíbia	10	0,331	Muro <i>et al.</i> , 2008
	8	0,272	
Centro Portugal	10	0,271	Manco <i>et al.</i> , 2008
	11	0,207	
Centro Portugal	10	0,280	Lopes <i>et al.</i> , 2009
	11	0,232	
Nordeste Portugal	10	0,266	Gusmão <i>et al.</i> , 2000
	11	0,206	
Nordeste Portugal	10	0,278	Pinheiro <i>et al.</i> , 2005
	11	0,200	
Norte de Portugal	10	0,258	Amorim <i>et al.</i> , 2001
	11	0,207	
Síria	10	0,319	Abdin <i>et al.</i> , 2003
	11	0,273	
Tanzânia	10	0,391	Forward <i>et al.</i> , 2008
	8	0,196	
	11	0,196	

Tabela 8.8: Alelos mais frequentes, no *loci* D7S820, em diferentes populações.

D8S1179			
População	Alelos mais Frequentes	Frequência	Referência
Rio de Janeiro	14	0,294	População Estudada
	13	0,264	
Alemanha	13	0,347	Reichenpfader <i>et al.</i> , 2003
	14	0,179	
China	13	0,195	Gusmão <i>et al.</i> , 2010
	14	0,184	
Egito - Andaima	13	0,247	Coudray <i>et al.</i> , 2007a
	15	0,227	
Espanha - Andaluzia	13	0,268	Coudray <i>et al.</i> , 2007c
	14	0,224	
Sul da Espanha	13	0,278	Gaibar <i>et al.</i> , 2010
	14	0,239	
Indígenas da Amazônia Brasileira (Terenas)	13	0,315	Manta <i>et al.</i> , 2012
	14	0,262	
Itália	13	0,312	Brisighelli <i>et al.</i> , 2009
	14	0,226	
Japão	13	0,251	Okamoto <i>et al.</i> , 2003
	14	0,195	
Líbia	13	0,301	Immel <i>et al.</i> , 2006
	15	0,218	
Marrocos - Asni	15	0,243	Coudray <i>et al.</i> , 2007b
	13	0,238	
Marrocos - Bouhria	13	0,260	Coudray <i>et al.</i> , 2007b
	14	0,216	
Moçambique	14	0,320	Alves <i>et al.</i> , 2004
	13	0,211	
Namíbia	14	0,477	Muro <i>et al.</i> , 2008
	13	0,177	
Centro Portugal	13	0,313	Lopes <i>et al.</i> , 2009
	14	0,219	
Nordeste Portugal	13	0,325	Gusmão <i>et al.</i> , 2000
	14	0,200	
Nordeste Portugal	13	0,290	Pinheiro <i>et al.</i> , 2005
	14	0,240	
Norte de Portugal	13	0,301	Amorim <i>et al.</i> , 2001
	14	0,207	
Síria	13	0,296	Abdin <i>et al.</i> , 2003
	15	0,179	
Tanzânia	14	0,298	Forward <i>et al.</i> , 2008
	13	0,207	
	15	0,207	

Tabela 8.9: Alelos mais frequentes, no *loci* D8S1179, em diferentes populações.

D13S317			
População	Alelos mais Frequentes	Frequência	Referência
Rio de Janeiro	12	0,335	População Estudada
	11	0,298	
China	8	0,299	Gusmão <i>et al.</i> , 2010
	11	0,201	
Egito - Andaima	12	0,369	Coudray <i>et al.</i> , 2007a
	11	0,273	
Sul da Espanha	15	0,233	Gaibar <i>et al.</i> , 2010
	18	0,233	
Espanha - Andaluzia	11	0,320	Coudray <i>et al.</i> , 2007c
	12	0,219	
Indígenas da Amazônia Brasileira (Terenas)	10.2	0,261	Manta <i>et al.</i> , 2012
	11	0,261	
Itália	11	0,300	Brisighelli <i>et al.</i> , 2009
	12	0,299	
Japão	8	0,243	Okamoto <i>et al.</i> , 2003
	11	0,243	
Líbia	12	0,369	Immel <i>et al.</i> , 2006
	11	0,216	
Marrocos - Asni	12	0,371	Coudray <i>et al.</i> , 2007b
	11	0,362	
Marrocos - Bouhria	12	0,433	Coudray <i>et al.</i> , 2007b
	11	0,293	
Moçambique	12	0,384	Alves <i>et al.</i> , 2004
	11	0,299	
Namíbia	12	0,503	Muro <i>et al.</i> , 2008
	11	0,308	
Centro de Portugal	11	0,361	Manco <i>et al.</i> , 2008
	12	0,243	
Centro Portugal	11	0,332	Lopes <i>et al.</i> , 2009
	12	0,253	
Nordeste Portugal	11	0,320	Gusmão <i>et al.</i> , 2000
	12	0,285	
Nordeste Portugal	11	0,302	Pinheiro <i>et al.</i> , 2005
	12	0,295	
Norte de Portugal	11	0,335	Amorim <i>et al.</i> , 2001
	12	0,254	
Síria	11	0,332	Abdin <i>et al.</i> , 2003
	12	0,252	
Tanzânia	12	0,404	Forward <i>et al.</i> , 2008
	11	0,307	

Tabela 8.10: Alelos mais frequentes, no *loci* D13S317, em diferentes populações.

D16S539			
População	Alelos mais Frequentes	Frequência	Referência
Rio de Janeiro	11	0,284	População Estudada
	12	0,235	
Alemanha	11	0,299	Reichenpfader <i>et al.</i> , 2003
	12	0,291	
Áustria	12	0,335	Reichenpfader <i>et al.</i> , 2003
	11	0,310	
Egito - Andaima	11	0,278	Coudray <i>et al.</i> , 2007a
	12	0,263	
Espanha - Andaluzia	11	0,329	Coudray <i>et al.</i> , 2007c
	12	0,303	
Sul da Espanha	12	0,278	Gaibar <i>et al.</i> , 2010
	11	0,267	
Indígenas da Amazônia Brasileira (Terenas)	11	0,410	Manta <i>et al.</i> , 2012
	12	0,184	
Itália	11	0,314	Brisighelli <i>et al.</i> , 2009
	12	0,270	
Japão	9	0,280	Okamoto <i>et al.</i> , 2003
	11	0,243	
Líbia	11	0,422	Immel <i>et al.</i> , 2006
	12	0,180	
Marrocos - Asni	12	0,310	Coudray <i>et al.</i> , 2007b
	11	0,243	
Marrocos - Bouhria	11	0,346	Coudray <i>et al.</i> , 2007b
	12	0,178	
Moçambique	11	0,296	Alves <i>et al.</i> , 2004
	9	0,243	
Namíbia	9	0,279	Muro <i>et al.</i> , 2008
	11	0,277	
Centro de Portugal	11	0,338	Manco <i>et al.</i> , 2008
	13	0,216	
Centro Portugal	12	0,287	Lopes <i>et al.</i> , 2009
	11	0,286	
Nordeste Portugal	12	0,312	Pinheiro <i>et al.</i> , 2005
	11	0,290	
Síria	12	0,279	Abdin <i>et al.</i> , 2003
	11	0,230	
Tanzânia	11	0,340	Forward <i>et al.</i> , 2008
	9	0,212	

Tabela 8.11: Alelos mais frequentes, no *loci* D16S539, em diferentes populações.

D18S51			
População	Alelos mais Frequentes	Frequência	Referência
Rio de Janeiro	15	0,170	População Estudada
	16	0,156	
Alemanha	15	0,177	Reichenpfader <i>et al.</i> , 2003
	14	0,153	
China	15	0,213	Gusmão <i>et al.</i> , 2010
	14	0,201	
Egito - Andaima	20	0,167	Coudray <i>et al.</i> , 2007a
	14	0,141	
Espanha - Andaluzia	14	0,180	Coudray <i>et al.</i> , 2007c
	12	0,171	
Sul da Espanha	12	0,167	Gaibar <i>et al.</i> , 2010
	17	0,150	
Indígenas da Amazônia Brasileira (Terenas)	14	0,278	Manta <i>et al.</i> , 2012
	17	0,183	
Itália	14	0,171	Brisighelli <i>et al.</i> , 2009
	13	0,154	
Japão	14	0,218	Okamoto <i>et al.</i> , 2003
	13	0,195	
Líbia	14	0,186	Immel <i>et al.</i> , 2006
	13	0,147	
Marrocos - Asni	16	0,147	Coudray <i>et al.</i> , 2007b
	16	0,181	
Marrocos - Bouhria	12	0,138	Coudray <i>et al.</i> , 2007b
	14	0,138	
Moçambique	12	0,173	Coudray <i>et al.</i> , 2007b
	14	0,149	
Moçambique	16	0,194	Alves <i>et al.</i> , 2004
	17	0,194	
Namíbia	17	0,238	Muro <i>et al.</i> , 2008
	16	0,205	
Síria	14	0,208	Abdin <i>et al.</i> , 2003
	12	0,159	
Tanzânia	16	0,196	Forward <i>et al.</i> , 2008
	17	0,187	
Centro Portugal	14	0,150	Lopes <i>et al.</i> , 2009
	15	0,149	
Nordeste Portugal	15	0,181	Gusmão <i>et al.</i> , 2000
	14	0,176	
Nordeste Portugal	16	0,180	Pinheiro <i>et al.</i> , 2005
	17	0,148	
Norte de Portugal	14	0,167	Amorim <i>et al.</i> , 2001
	16	0,164	

Tabela 8.12: Alelos mais frequentes, no *loci* D18S51, em diferentes populações.

D21S11			
População	Alelos mais Frequentes	Frequência	Referência
Rio de Janeiro	30	0,242	População Estudada
	29	0,216	
Alemanha	29	0,218	Reichenpfader <i>et al.</i> , 2003
	30	0,210	
China	29	0,328	Gusmão <i>et al.</i> , 2010
	30	0,282	
Egito - Andaima	29	0,359	Coudray <i>et al.</i> , 2007a
	30	0,307	
Espanha - Andaluzia	29	0,228	Coudray <i>et al.</i> , 2007c
	30	0,197	
Sul da Espanha	30	0,300	Gaibar <i>et al.</i> , 2010
	29	0,239	
Indígenas da Amazônia Brasileira (Terenas)	30	0,245	Manta <i>et al.</i> , 2012
	31	0,133	
Itália	29	0,244	Brisighelli <i>et al.</i> , 2009
	30	0,214	
Japão	30	0,384	Okamoto <i>et al.</i> , 2003
	29	0,226	
Líbia	30	0,286	Immel <i>et al.</i> , 2006
	29	0,233	
Marrocos - Asni	29	0,257	Coudray <i>et al.</i> , 2007b
	30	0,238	
Marrocos - Bouhria	29	0,303	Coudray <i>et al.</i> , 2007b
	30	0,231	
Moçambique	28	0,243	Alves <i>et al.</i> , 2004
	29	0,201	
Namíbia	28	0,238	Muro <i>et al.</i> , 2008
	30	0,192	
Centro Portugal	30	0,250	Lopes <i>et al.</i> , 2009
	29	0,231	
Nordeste Portugal	30	0,215	Gusmão <i>et al.</i> , 2000
	28	0,195	
Nordeste Portugal	30	0,252	Pinheiro <i>et al.</i> , 2005
	28	0,157	
Norte de Portugal	30	0,239	Amorim <i>et al.</i> , 2001
	29	0,191	
Síria	29	0,219	Abdin <i>et al.</i> , 2003
	32.2	0,127	
Tanzânia	28	0,257	Forward <i>et al.</i> , 2008
	29	0,192	

Tabela 8.13: Alelos mais frequentes, no *loci* D21S11, em diferentes populações.

D2S1338			
População	Alelos mais Frequentes	Frequência	Referência
Rio de Janeiro	17	0,182	População Estudada
	19	0,152	
Alemanha	17	0,215	Reichenpfader <i>et al.</i> , 2003
	20	0,141	
Áustria	17	0,180	Reichenpfader <i>et al.</i> , 2003
	20	0,125	
Egito - Andaima	18	0,157	Coudray <i>et al.</i> , 2007a
	20	0,126	
Espanha - Andaluzia	17	0,268	Coudray <i>et al.</i> , 2007c
	20	0,149	
Sul da Espanha	17	0,278	Gaibar <i>et al.</i> , 2010
	20	0,139	
	25	0,139	
Marrocos - Asni	17	0,300	Coudray <i>et al.</i> , 2007b
	20	0,181	
Marrocos - Bouhria	17	0,346	Coudray <i>et al.</i> , 2007b
	20	0,144	
Moçambique	22	0,194	Alves <i>et al.</i> , 2004
	21	0,162	
Namíbia	19	0,218	Muro <i>et al.</i> , 2008
	21	0,136	
Centro Portugal	17	0,273	Lopes <i>et al.</i> , 2009
	20	0,147	
Tanzânia	21	0,186	Forward <i>et al.</i> , 2008
	19	0,177	

Tabela 8.14: Alelos mais frequentes, no *loci* D2S1338, em diferentes populações.

D19S433			
População	Alelos mais Frequentes	Frequência	Referência
Rio de Janeiro	13	0,258	População Estudada
	14	0,249	
Alemanha	14	0,293	Reichenpfader <i>et al.</i> , 2003
	13	0,245	
Áustria	14	0,340	Reichenpfader <i>et al.</i> , 2003
	13	0,210	
Egito - Andaima	13	0,247	Coudray <i>et al.</i> , 2007a
	14	0,197	
Espanha - Andaluzia	14	0,355	Coudray <i>et al.</i> , 2007c
	13	0,316	
Sul da Espanha	14	0,333	Gaibar <i>et al.</i> , 2010
	13	0,289	
Marrocos - Asni	13	0,290	Coudray <i>et al.</i> , 2007b
	14	0,200	
Marrocos - Bouhria	14	0,238	Coudray <i>et al.</i> , 2007b
	13	0,264	
Moçambique	13	0,324	Alves <i>et al.</i> , 2004
	14	0,229	
Namíbia	13	0,310	Muro <i>et al.</i> , 2008
	14	0,210	
Centro Portugal	14	0,313	Lopes <i>et al.</i> , 2009
	13	0,265	
Tanzânia	13	0,265	Forward <i>et al.</i> , 2008
	14	0,224	

Tabela 8.15: Alelos mais frequentes, no *loci* D19S433, em diferentes populações.

Penta E			
População	Alelos mais Frequentes	Frequência	Referência
Rio de Janeiro	12	0,159	População Estudada
	11	0,125	
Indígenas da Amazônia Brasileira (Terenas)	14	0,180	Manta <i>et al.</i> , 2012
	12	0,162	
Itália	12	0,172	Brisighelli <i>et al.</i> , 2009
	7	0,155	
Japão	15	0,150	Okamoto <i>et al.</i> , 2003
	11	0,121	
Líbia	16	0,121	Immel <i>et al.</i> , 2006
	11	0,165	
Moçambique	12	0,150	Alves <i>et al.</i> , 2004
	8	0,174	
Centro Portugal	13	0,130	Lopes <i>et al.</i> , 2009
	12	0,213	
Nordeste Portugal	7	0,153	Pinheiro <i>et al.</i> , 2005
	12	0,192	
Síria	7	0,140	Abdin <i>et al.</i> , 2003
	12	0,185	
	13	0,112	
	7	0,122	

Tabela 8.16: Alelos mais frequentes, no *loci* Penta E, em diferentes populações.

Penta D			
População	Alelos mais Frequentes	Frequência	Referência
Rio de Janeiro	9	0,182	População Estudada
	12	0,166	
Indígenas da Amazônia Brasileira (Terenas)	10	0,294	Manta <i>et al.</i> , 2012
	12	0,259	
Japão	9	0,297	Okamoto <i>et al.</i> , 2003
	10	0,220	
Líbia	13	0,241	Immel <i>et al.</i> , 2006
	9	0,168	
Moçambique	10	0,201	Alves <i>et al.</i> , 2004
	8	0,146	
Centro Portugal	13	0,205	Lopes <i>et al.</i> , 2009
	9	0,198	
Nordeste Portugal	13	0,235	Pinheiro <i>et al.</i> , 2005
	9	0,217	
Síria	13	0,200	Abdin <i>et al.</i> , 2003
	12	0,178	

Tabela 8.17: Alelos mais frequentes, no *loci* Penta D, em diferentes populações.