



**UNIVERSIDADE FEDERAL DO AMAZONAS
PROGRAMA MULTI-INSTITUCIONAL DE PÓS-
GRADUAÇÃO EM BIOTECNOLOGIA - PPGBIOTEC**

**DESENVOLVIMENTO DE UM NOVO SISTEMA MULTIPLEX
DE MICROSSATÉLITES (STR) PARA ANÁLISE GENÉTICA
DE POPULAÇÕES HUMANAS**

ISABEL DA MOTA PONTES

**MANAUS
2009**

ISABEL DA MOTA PONTES

**DESENVOLVIMENTO DE UM NOVO SISTEMA MULTIPLEX
DE MICROSSATÉLITES (STR) PARA ANÁLISE GENÉTICA
DE POPULAÇÕES HUMANAS**

Tese apresentada ao Programa Multi-Institucional de Pós-Graduação em Biotecnologia da Universidade Federal do Amazonas, como requisito parcial para a obtenção do título de Doutor em Biotecnologia, área de concentração biotecnologia para a saúde.

Orientadores: *Dr. Spartaco Astolfi-Filho (UFAM)*
Dr. Elizeu Fagundes de Carvalho (UERJ)
Dra. Dayse Aparecida da Silva (UERJ)

**MANAUS
2009**

Ficha Catalográfica
(Catalogação realizada pela Biblioteca Central da UFAM)

Pontes, Isabel da Mota

P814d Desenvolvimento de um novo sistema multiplex de
microsatélites (STR) para análise genética de populações humanas
/ Isabel da Mota Pontes. – Manaus: UFAM, 2009.
156 f.; il. color.

Tese (Doutorado em Biotecnologia) — Universidade Federal
do Amazonas, 2009.

Orientadores: Dr. Spartaco Astolfi-Filho, Dr. Elizeu Fagundes
de Carvalho, Dra. Dayse Aparecida da Silva.

1. STR, 2. Multiplex, 3. Banco de dados, 4. Identificação
humana por DNA; 5. População humana 3. Marcadores
Moleculares, I. Universidade Federal do Amazonas II. Título

CDU 577.21(043.2)

ISABEL DA MOTA PONTES

**DESENVOLVIMENTO DE UM NOVO SISTEMA MULTIPLEX
DE MICROSSATÉLITES (STR) PARA ANÁLISE GENÉTICA
DE POPULAÇÕES HUMANAS**

Tese apresentada ao Programa Multi-Institucional de Pós-Graduação em Biotecnologia da Universidade Federal do Amazonas, como requisito parcial para a obtenção do título de Doutor em Biotecnologia, área de concentração biotecnologia para a saúde.

Aprovada em 14 de dezembro de 2009

BANCA EXAMINADORA

**Prof. Dr. Spartaco Astolfi Filho
Universidade Federal do Amazonas**

**Prof. Dr. Elizeu Fagundes de Carvalho
Universidade Estadual do Estado do Rio de Janeiro**

**Prof^ª. Dra. Dayse Aparecida da Silva
Universidade Estadual do Estado do Rio de Janeiro**

**Prof^ª. Dra. Paula Cristina Ângelo
Embrapa Amazônia Ocidental**

**Prof. Dr. Luiz Alberto dos Santos Monjeló
Universidade Federal do Amazonas**

**Prof^ª. Dra. Maria das Neves da Silva Viana
Universidade Federal do Amazonas**

**Prof^ª. Dra. Rozana de Medeiros Sousa Galvão
Universidade Federal do Amazonas**

Esta Tese foi desenvolvida nos Laboratórios de Diagnóstico por DNA do Instituto de Biologia Roberto Alcântara Gomes da Universidade Estadual do Estado do Rio de Janeiro – UERJ, sob orientação do Dr. Elizeu Fagundes de Carvalho e Dra. Dayse Aparecida da Silva, e no Laboratório de Diagnóstico Molecular do Centro de Apoio Multidisciplinar da Universidade Federal do Amazonas – UFAM, sob orientação do Dr. Spartaco Astolfi Filho recursos provenientes da UERJ e UFAM.



UNIVERSIDADE FEDERAL DO AMAZONAS
(Pista principal da entrada do Campus)



UNIVERSIDADE ESTADUAL DO ESTADO
DO RIO DE JANEIRO
(Vista dos prédios do Campus)

AGRADECIMENTOS

O DEUS pela infinita misericórdia de proporcionar a inteligência para que eu pudesse desenvolver este trabalho.

À Dra. Paula Ângelo, pela ajuda no desenvolvimento dos microssatélites, e pela amizade construída.

Ao Dr. Luiz Alberto dos Santos Monjeló, pelo grande incentivo na realização deste trabalho.

À MSc Fernanda Manta, esteve sempre me ajudando na realização das análises estatísticas e pela amizade.

Aos participantes como doadores voluntários que sem os mesmos não haveria o material biológico extremamente necessário para que este trabalho fosse desenvolvido.

Ao mestrando Roberto Alexandre Filho Alves Barbosa, companheiro fiel, dedicado aos ensinamentos, muito profissional, muito me ajudou nas análises da genética de população.

Ao grupo de trabalho do Laboratório de Diagnóstico por DNA da UERJ, sempre dando apoio na etapa da genotipagem.

Ao grupo de trabalho do Laboratório de Diagnóstico Molecular da UFAM, Francisca Dabella, Diego Bilby, André Santos, pela dedicação e entusiasmo na realização da prática deste projeto.

À Técnica Enedina Assunção, pela determinação e realização do sequenciamento dos primers utilizados no desenvolvimento do multiplex.

Aos amigos e colegas do Curso de Biotecnologia da Universidade Federal do Amazonas, pelos estudos compartilhados, pelos momentos agradáveis vividos.

Ao colega Roberto Lira e Josué, pela ajuda durante a revisão deste trabalho e a amizade que sempre esteve presente.

As amigas Doutoradas Dorian, Maria das Neves, Cristina Borborema, Cíntia Mara, que sempre me incentivaram e ajudaram nas conclusões finais deste trabalho.

A todos aqueles que de uma forma ou de outra colaboraram na realização deste trabalho.

AGRADECIMENTO ESPECIAL

Aos meus Orientadores

Prof. Dr. Elizeu Fagundes de Carvalho,
Prof. Dr. Spartaco Astolfi Filho,
Profa. Dra. Dayse Aparecida da Silva

Brilhantes mestres, por me proporcionarem um ambiente enriquecidos de conhecimentos, por terem disponibilizados recursos humanos e materiais necessários para o desenvolvimento deste trabalho, especialmente por acreditarem em mim e na conclusão deste projeto, e pela amizade conquistada.

Aos meus Familiares

Agradecimento especial ao meu esposo Jerildo Thomé, pelo incentivo, carinho, compreensão e paciência durante a realização desta tese.

Aos meus filhos Célio, Orbélio, Jerildo Jr, e meus netinhos queridos Yuri, Isabelle, Igo, Jerildo Neto, Nani Vitória e John “O cara especial”, pelo carinho, alegria que me proporcionaram.

Beijos a todos

Aos meus pais (*In memoriam*)

Dedico

RESUMO

O presente trabalho consiste no desenvolvimento de um novo sistema multiplex denominado PentaPlex-ISED, composto por 5 *loci* tipo microssatélites STR (D5S198498, D3S18773, GH15, D8S82535 e D11S118590). Esses microssatélites foram identificados no genoma humano, por meio do programa BLAST-N (GenBank). Primers específicos foram desenhados para amplificação via PCR. Primers fluorescentes foram sintetizados para permitir análise simultânea de vários amplicons (análise multiplex). Todos os *loci* microssatélites foram genotipados confirmando-se o alto grau de polimorfismo. Na validação do sistema, a concentração mínima de DNA para a amplificação dos *loci* incluídos nesse multiplex foi estimada em 0,25 ng. Diversos tipos de amostras biológicas (sangue, saliva, urina e sêmen) foram adquiridas satisfatoriamente. Foi construído um banco de dados de frequências dos alelos microssatélites para três populações brasileiras (Amazonas, Rio de Janeiro e Espírito Santo). Confirmou-se a independência dos *loci*. As distribuições das frequências alélicas apresentaram pouca variação entre os grupos populacionais. Por meio dos testes estatísticos de χ^2 , utilizando o programa Arlequin 3.1 verificou-se que a maioria dos *loci* não adere as condições do equilíbrio de Hardy-Weinberg se análise for feita separadamente em cada uma das populações. No entanto, na análise feita com o conjunto total das três populações, há aderência ao Equilíbrio de Hardy-Weinberg. As três populações foram agrupadas e comparadas par-a-par; mostrando-se que a variação genética é maior dentro das populações que entre elas. Os dados estatísticos sobre o poder de exclusão, poder de discriminação e o conteúdo de informação do polimorfismo, para os 5 *loci* foram determinados para cada STR e em conjunto para avaliar o potencial do sistema PentaPlex-ISED na identificação humana. Observou-se que este sistema multiplex é adequado para identificação humana e que pode ser usado de forma complementar com os outros sistemas de análises genéticas, por possuir um poder de discriminação acumulado em 99,996 % e o poder de exclusão acumulado em 93,436 %.

Palavras chave: 1. STR, 2. Multiplex, 3. Banco de dados, 4. Identificação humana por DNA; 5. População humana 6. Marcadores Moleculares.

ABSTRACT

The present work presents the development of a new system called multiplex Pentaplex-ised, composed of 5 microsatellite loci STR type (D5S198498, D3S18773, GH15, D8S82535 and D11S118590). Such microsatellites have been identified in the human genome, using the program BLAST-N (GenBank). Specific primers were designed for amplification via PCR. Fluorescent primers were synthesized to allow simultaneous analysis of multiple amplicons (multiplex analysis). All microsatellite loci were genotyped confirming the high degree of polymorphism. In the validation, the minimum DNA concentration of for amplification was estimated at 0.25 ng. Several types of biological samples (blood, saliva, urine and semen) were acquired satisfactorily. We built a database of microsatellite allele frequencies for three Brazilian populations -Amazonas, Rio de Janeiro and Espírito Santo. The distributions of allele frequencies showed little variation between population groups. By means of statistical tests of χ^2 using the program Arlequin 3.1 most alleles do not adhere to Hardy-Weinberg equilibrium if performed separately for each of the populations. However, in the analysis performed with the whole set of three populations, there is high adherence to the Hardy-Weinberg equilibrium. The three populations were grouped and compared pair-to-pair, showing that the genetic variation is greater within populations than between them. Statistical data about the power of exclusion, discrimination power and polymorphism information content for the five loci were determined for each STR and together to assess the potential of the system Pentaplex-ised in human identification. We observed that this multiplex system is suitable for human identification and that can be used in a complementary method with other systems for genetic analysis, by having a cumulative power of discrimination of 99.996% and 93.436% cumulative exclusion.

Keywords: 1. STR, 2. Multiplex 3. Database, 4. Human DNA identification; 5. human population; 6. Molecular Markers

LISTA DE FIGURAS

| | |
|---|-----|
| Figura 01 – Esquema dos cromossomos humanos (sexual feminino) | 23 |
| Figura 02 – Esquema da distribuição do genoma humano | 24 |
| Figura 03 – RFLP e VNTR (impressões digitais de DNA) | 31 |
| Figura 04 – STRs apresentando bandas <i>stutters</i> | 36 |
| Figura 05 – Microvariante do <i>locus</i> FGA | 38 |
| Figura 06 – Os 13 marcadores STRs definidos pelo FBI para compor o CODIS .. | 52 |
| Figura 07– Posição dos novos marcadores STRs..... | 82 |
| Figura 08 – Perfil eletroforético de DNA em gel de agarose 2,5% | 87 |
| Figura 09 – Perfil eletroforético em gel de poliacrilamida 0,6% | 88 |
| Figura 10 – Perfil eletroforético das amostras de DNA da população do (AM) | 91 |
| Figura 11 – Perfil eletroforético da escada alélica em gel de poliacrilamida | 93 |
| Figura 12 – Eletroferograma do (K562) utilizando o sistema PentaPlex-ISED | 96 |
| Figura 13 - Eletroferograma (K562) e Amelogenina utilizando PentaPlex-ISED .. | 99 |
| Figura 14 – Eletroferograma da amostra de sangue (58s_M) c/ PentaPlex-ISED .. | 99 |
| Figura 15 - Eletroferograma da amostra de sangue (50-7F) c/ PentaPlex-ISED .. | 100 |
| Figura 16 - Eletroferograma da amostra de <i>saliva</i> c/ PentaPlex-ISED | 101 |
| Figura 17 - Eletroferograma da amostra de sangue c/ PentaPlex-ISED | 101 |
| Figura 18 - Eletroferograma da amostra de sêmen c/ PentaPlex-ISED | 102 |
| Figura 19 - Eletroferograma da amostra de urina c/ PentaPlex-ISED | 102 |
| Figura 20 – Perfil eletroforético da banda <i>stutter</i> no locus GH15, alelo 12 | 104 |
| Figura 21 – Percentual de microvariância por <i>locus</i> | 106 |
| Figura 22 – Percentual de alelos tipados no <i>locus</i> D5S198498 | 110 |
| Figura 23 - Percentual de alelos tipados no <i>locus</i> D11S118590 | 112 |
| Figura 24 - Percentual de alelos tipados no <i>locus</i> GH15 | 114 |
| Figura 25 - Percentual de alelos tipados no <i>locus</i> D8S82535 | 116 |
| Figura 26 - Percentual de alelos tipados no <i>locus</i> D3S18773 | 119 |
| Figura 27 – Mapa do Brasil mostrando a distância geográfica entre 3 populações | 127 |

LISTA DE TABELAS

| | |
|--|-----|
| Tabela 01 – Sistema PowerPlex 16 (Promega) | 41 |
| Tabela 02 – Sistema Identifiler | 42 |
| Tabela 03 – Fórmulas utilizadas nos cálculos do Índice de Paternidade | 57 |
| Tabela 04 – Relação dos 5 marcadores STR e suas características..... | 81 |
| Tabela 05 – Seqüência dos <i>primers</i> e suas posições nos cromossomos humanos | 83 |
| Tabela 06 – Nomenclatura dos novos marcadores STR..... | 84 |
| Tabela 07 – Seqüências marcadas (NED, FAM e HEX)..... | 85 |
| Tabela 08 – Definição das condições da PCR para os marcadores monoplex..... | 86 |
| Tabela 09 – Condições da PCR para o sistema PentaPlex-ISED..... | 89 |
| Tabela 10 – Mostra a mistura para a composição da escada alélica..... | 92 |
| Tabela 11 – Frequências do <i>locus</i> D5S198498..... | 109 |
| Tabela 12 – Frequências do <i>locus</i> D11S118590..... | 111 |
| Tabela 13 – Frequências do <i>locus</i> GH15..... | 113 |
| Tabela 14 – Frequências do <i>locus</i> D8S82535..... | 115 |
| Tabela 15 – Frequências do <i>locus</i> D3S18773..... | 118 |
| Tabela 16 – Poder de Discriminação (PD), Poder de exclusão (PE) e o conteúdo de informação de Polimorfismo (PIC) para os 5 <i>loci</i> | 121 |
| Tabela 17 –Teste para o equilíbrio de Hardy-Weinberg nos 5 <i>loci</i> na população do Amazonas..... | 123 |
| Tabela 18 –Teste para o equilíbrio de Hardy-Weinberg nos 5 <i>loci</i> STR, na população do Rio de Janeiro..... | 124 |
| Tabela 19 –Teste para o equilíbrio de Hardy-Weinberg nos 5 <i>loci</i> STR, na população do Espírito Santo..... | 124 |
| Tabela 20 – Valores de R_{ST} e de p nas 3 populações (AM, RJ e ES) | 129 |
| Tabela 21– Estimativas de F_{ST} e Correspondentes de P dos resultados da Análise de Variância molecular (AMOVA) de amostras populacionais(AM, RJ e ES)..... | 129 |

LISTA DE ABREVIATURAS, SIGLAS, SÍMBOLOS E UNIDADES

A- Adenina

cm - Centímetro

Cl - Cloro

CODIS- Combined DNA Index System

°C - Graus Celsius

DNA - Ácido desoxirribonucléico

dNTP - Desoxirribonuceotídeo 5'-trifosfato

EDTA - Ácido etileno diamino tetracético

FBI- Federal Bureau of Investigation

g - Grama

GEP - Grupo Espanhol português

HLA - Antígeno leucocitário humano

HWE - Equilíbrio de Hardy-Weinberg

ICP - Índice Cumulativo de Paternidade

IP - Índice de Paternidade

ISFG - International Society for Forensic Genetics

Kb - kilobase

KCl - Cloreto de potássio

K562 - Linhagem celular humana

M - Molar

MgCl₂ - Cloreto de magnésio

mg - Miligrama

mL - Mililitro

mm- milímetro

mM - Milimolar

min - Minuto

µg - Micrograma

µL - Microlitro

µM -Micromolar

NaCl - Cloreto de sódio

NaOH - Hidróxido de sódio

ng - Nanograma

PAGE - Eletroforese em gel de poliacrilamida

pb - Pares de base

PC - Probabilidade de Coincidência

PCR - Reação em cadeia da polimerase

PD - Poder de Discriminação

PE - Poder de Exclusão

pH - Potencial de hidrogênio

PIC- Conteúdo de informação do polimorfismo

pm - Peso molecular

PP- Probabilidade de paternidade

QTL – Loci de caráter quantitativo

RFLP - Polimorfismo de tamanho do fragmento de restrição

RFU – Medidas de intensidade de fluorescência

rpm - rotações por minuto

RV - razão de Verossimilhança

SDS - dodecil sulfato de sódio

STR - Sequências curtas repetidas em tandem

Taq - *Thermus aquaticus*

TE - Tampão Tris-EDTA

TEMED – N,N, N', N' – Tetrametil-etilenodiamina

U - Unidade de enzima

UV- Ultra violeta

V- volt

VNTR – *Variable Number of Tandem Repeats* (Número variado de repetição em tandem)

∞ - infinito

SUMÁRIO

| | | |
|--------------|---|-----------|
| 1 | INTRODUÇÃO | 18 |
| 2 | REVISÃO BIBLIOGRAFICA..... | 22 |
| 2.1 | GENES E GENOMA..... | 22 |
| 2.1.1 | O genoma mitocondrial..... | 25 |
| 2.2 | MARCADORES GENETICOS MOLECULARES..... | 26 |
| 2.2.1 | A técnica da PCR..... | 28 |
| 2.2.2 | Marcadores RFLPs..... | 29 |
| 2.2.3 | Marcadores RAPD..... | 31 |
| 2.2.4 | Marcadores AFLP..... | 32 |
| 2.2.5 | Marcadores Microssatélites ou STRs..... | 33 |
| 2.2.5.1 | Características especiais das regiões microssatélites STRs..... | 35 |
| 2.3 | DESENVOLVIMENTO DE SISTEMA MULTIPLEX STR..... | 39 |
| 2.3.1 | Validação de sistemas multiplex de STR..... | 43 |
| 2.4 | GENÉTICA DE POPULAÇÕES..... | 45 |
| 2.4.1 | Formação da população do Brasil..... | 46 |
| 2.4.2 | Variabilidade Genética..... | 48 |
| 2.5 | IDENTIFICAÇÃO HUMANA..... | 49 |
| 2.5.1 | Normalização das análises de identificação humana..... | 52 |
| 2.5.2 | Parâmetros de interesse forense..... | 53 |
| 2.5.3 | Parâmetros de Paternidade..... | 54 |
| 3 | OBJETIVOS..... | 59 |
| 3.1 | OBJETIVO GERAL..... | 59 |
| 3.2 | OBJETIVOS ESPECIFICOS..... | 59 |
| 4 | MATERIAL E METODOS..... | 60 |
| 4.1 | DESENVOLVIMENTO DO SISTEMA MULTIPLEX..... | 61 |
| 4.1.1 | Amostras biológicas..... | 61 |
| 4.1.1.1 | Coleta das amostras..... | 61 |
| 4.1.2 | Extração e análise de DNA..... | 62 |
| 4.1.2.1 | Quantificação do DNA..... | 62 |
| 4.1.3 | Seleção dos <i>loci</i> STRs..... | 64 |
| 4.1.4 | Desenho e marcação dos <i>primers</i>..... | 64 |
| 4.1.4.1 | Nomenclatura do Sistema..... | 65 |
| 4.1.5 | Definição das condições da PCR do sistema multiplex..... | 66 |
| 4.1.6 | Definição das condições de amplificação via PCR..... | 67 |
| 4.1.6.1 | Construção da escada alélica..... | 68 |
| 4.1.6.2 | Eletroforese em gel de poliacrilamida (PAGE)..... | 68 |

| | | |
|--------------|---|------------|
| 4.1.6.2.1 | Preparo do gel de poliacrilamida..... | 69 |
| 4.1.6.2.2 | Preparo das amostras (PCR) para eletroforese..... | 69 |
| 4.1.6.2.3 | Coloração do PAGE com nitrato de prata..... | 70 |
| 4.1.6.3 | Condições eletroforéticas em sistema capilar..... | 71 |
| 4.2 | VALIDAÇÃO DO SISTEMA PENTAPLEX-ISED..... | 72 |
| 4.2.1 | Análise da sensibilidade do sistema PentaPlex-ISED..... | 72 |
| 4.2.2 | Análise da reprodutibilidade do sistema de tipagem..... | 73 |
| 4.2.3 | Determinação de alelos microvariantes..... | 73 |
| 4.2.3.1 | Sequenciamento dos <i>loci</i> STRs..... | 73 |
| 4.2.3.2 | Reação de Sequenciamento..... | 74 |
| 4.2.3.3 | Precipitação e leitura da reação de sequenciamento..... | 75 |
| 4.2.4 | Estudo das bandas <i>stutter</i>..... | 75 |
| 4.3 | CONSTRUÇÃO E ANÁLISE DO BANCO DE DADOS..... | 76 |
| 4.3.1 | Análise de parâmetros genéticos populacionais..... | 76 |
| 4.3.2 | Análises estatísticas..... | 77 |
| 4.3.2.1 | Cálculos das frequências dos alelos STR..... | 77 |
| 4.3.2.2 | Poder de discriminação (PD)..... | 77 |
| 4.3.2.3 | Probabilidade de coincidência..... | 78 |
| 4.3.3 | Estudo de Populações..... | 78 |
| 4.3.3.1 | Comparação populacional..... | 79 |
| 4.3.3.2 | Detalhamento das análises para o estudo de população..... | 80 |
| 5 | RESULTADOS E DISCUSSÃO..... | 81 |
| 5.1 | SELEÇÃO DOS <i>LOCI</i> STRS..... | 81 |
| 5.2 | DESENHO DOS <i>PRIMERS</i> | 82 |
| 5.2.1 | Nomenclatura dos STRs..... | 83 |
| 5.2.2 | Marcação dos <i>primers</i>..... | 84 |
| 5.3 | DEFINIÇÃO DAS CONDIÇÕES DA PCR..... | 85 |
| 5.4 | AMOSTRAS BIOLÓGICAS E ANÁLISE DE DNA..... | 90 |
| 5.5 | CONSTRUÇÃO DA ESCADA ALÉLICA..... | 91 |
| 5.6 | COMPOSIÇÃO DO SISTEMA PENTAPLEX-ISED..... | 94 |
| 5.7 | ESTUDOS DE VALIDAÇÃO DO SISTEMA PENTAPLEX-ISED..... | 97 |
| 5.7.1 | Análise da sensibilidade do sistema..... | 98 |
| 5.7.2 | Análise da reprodutibilidade do sistema de tipagem..... | 98 |
| 5.7.3 | Padronização da amplificação com diferentes enzimas..... | 103 |

| | | |
|--------------|---|------------|
| 5.7.4 | Análises das Bandas <i>Stutter</i>..... | 104 |
| 5.7.5 | Determinação de alelos microvariantes em cada <i>locus</i> STR..... | 105 |
| 5.8 | COSTRUÇÃO DO BANCO DE DADOS DE ALELOS DO SISTEMA PENTAPLEX- ISED..... | 108 |
| 5.8.1 | Frequências alélica do <i>locus</i> D5S198498..... | 108 |
| 5.8.2 | Frequências alélica do <i>locus</i> D11S118590..... | 111 |
| 5.8.3 | Frequências alélica do <i>locus</i> GH15..... | 113 |
| 5.8.4 | Frequências alélica do <i>locus</i> D8S82535..... | 115 |
| 5.8.5 | Frequência alélica do <i>locus</i> D3S18773..... | 117 |
| 5.9 | PARÂMETROS ESTATÍSTICOS DE INTERESSE FORENSE E DE VÍNCULO GENÉTICO COM BASE NOS 5 <i>LOCI</i> STR..... | 120 |
| 5.9.1 | Teste de independência em um <i>locus</i> (Equilíbrio de Hardy-Weinberg).. | 122 |
| 5.10 | COMPARAÇÃO POPULACIONAL..... | 126 |
| 6 | CONCLUSÕES..... | 131 |
| 7 | CONSIDERAÇÕES FINAIS..... | 132 |
| | REFERÊNCIAS..... | 133 |
| | APÊNDICE..... | 144 |

1 INTRODUÇÃO

O avanço do conhecimento técnico-científico sobre a estrutura da molécula do DNA resultou no desenvolvimento de ferramentas que se tornaram imprescindíveis na área biotecnológica. A descoberta da existência de genes, entidades capazes de transmitir os caracteres hereditários foram realizadas por Mendel (1865). Embora o DNA (ácido desoxirribonucléico) tenha sido descoberto por Friederich Miescher (1869), somente várias décadas depois, Avery e colaboradores (1944) demonstraram ser o material genético (Princípio Transformante) composto de DNA. A seguir Watson e Crick, em 1953, propuseram a estrutura tridimensional do DNA. Ainda na década de 50 e na seguinte, diversos experimentos foram realizados, para demonstrar como o DNA se replica como o gene é expresso e como é regulado, resultando no que chamamos de Dogma Central da biologia molecular. No início da década de 70, Boyer e Paul Berg dão início a uma nova fase da biologia molecular, a fase da Tecnologia do DNA Recombinante ou Engenharia Genética, que permitiu: - estudar o genoma e a expressão dos genes eucarióticos; - construir células recombinantes e seres transgênicos para aplicações biotecnológicas; - desenvolver diferentes tipos de marcadores genéticos moleculares para diferentes usos, inclusive para identificação humana.

No que se refere à identificação humana, até o início do século XIX, esta era realizada utilizando marcadores anatômicos, para os quais as análises eram processadas, por meio de observação e comparação de traços físicos. Em 1879, Alphonse Bertillon utilizou as dimensões ósseas como marcadores de individualização, baseando-se não só no conceito de que o tamanho dos ossos humanos permanece constante, a partir dos vinte anos de idade, mas também na pequena probabilidade de se encontrar dois indivíduos com o mesmo perfil de medição corpórea. Assim sendo, Bertillon (1879) propôs onze diferentes medições

anatômicas analisadas de um indivíduo, entre as quais foram incluídas: a altura do corpo, largura da cabeça e o comprimento do pé (BODY POLITIC, 2003).

Em 1909, Landsteiner descreveu o primeiro antígeno do grupo ABO, e Essen-Moller (1938) descreveu as fórmulas matemáticas que facilitaram os cálculos sobre a probabilidade de paternidade baseado no teorema de Bayer, publicado em 1763. Outros tipos de grupos sanguíneos foram descobertos e também passaram a ser utilizado em análise genética, juntamente com o ABO, como os sistemas MN e Rh (LANDSTEINER & WIENER, 1940; LEWIS MOURANT, 1946 e ANDRESEN, 1947).

Considerando-se o advento da biologia molecular e o desenvolvimento de técnicas para a análise do DNA, os testes baseados em marcadores protéicos (fenotípicos) tornaram-se obsoletos, possibilitando, tão somente, que os indivíduos fossem excluídos de paternidade ou maternidade. Tais sistemas apresentavam baixa capacidade de individualização, em razão de seu reduzido grau de variabilidade. A partir da década 80, entretanto, a identificação humana deu um salto, no que diz respeito às metodologias utilizadas, tendo sua capacidade de individualização ampliada. Este marco histórico proporcionado pelo trabalho de Jeffreys e seus colaboradores (1985a), fazendo uso de sondas multilocais (VNTR – *Variable Number of Tandem Repeats* - ou minissatélites), conseguiram produzir padrões de RFLP - *Restriction fragment length polymorphism* específicos para cada ser humano, do tipo *fingerprinting*. A mencionada metodologia permitiu a realização do primeiro teste de identificação humana por DNA, para fins criminalistas.

A invenção de uma nova técnica molecular desenvolvida por Mullis & Faloona (1987), chamada PCR - *Polymerase Chain Reaction*, que permite a amplificação de determinados *loci* de DNA, causou uma verdadeira revolução na biologia molecular e em suas aplicações. Configurando-se numa ferramenta poderosa, a técnica ganhou aceitação por toda a

comunidade científica, rapidamente, uma vez que a metodologia se faz extremamente eficiente, em casos nos qual a quantidade de DNA disponível é muito pequena. Atualmente, a PCR é usada para amplificar *loci* de microssatélites, que são altamente polimórficos e apresentam herança intermediária (sem dominância) permitindo a identificação de indivíduos heterozigotos. É um método poderoso na avaliação da individualização humana, sendo preferido na análise pericial de casos de identificação humana por DNA, no qual a amostra biológica seja mínima ou degradada.

Outra tecnologia para análise genética em seres humanos que vem sendo desenvolvida é aquela referente aos novos marcadores genéticos denominados SNPs - *Single Nucleotide Polymorphisms*. Os SNPs possuem taxas de mutações muito menores que os microssatélites, sendo utilizados, também, na investigação de origem étnica ou ancestralidade.

No Brasil, as técnicas da biologia molecular, aplicadas à identificação do material genético humano foram implantadas e consolidaram-se nas principais capitais do país, em diversos laboratórios das regiões Sul, Sudeste e Centro-Oeste. Em Manaus, desde 1998, a Universidade Federal do Amazonas, no Laboratório de Diagnóstico Molecular, vem desenvolvendo atividades de pesquisa em diagnóstico, implantando novas tecnologias para análises genéticas em seres humanos e desenvolvendo novos marcadores moleculares microssatélites - STRs (PONTES, 2003).

Neste trabalho, visando aprofundar os conhecimentos na área da genética molecular e genética forense, visando contribuir para a implantação dos testes de identificação humana em Manaus, foi desenvolvido um novo sistema multiplex composto por cinco marcadores moleculares do tipo microssatélite STR para a análise genética. Este sistema multiplex passou a ser chamado de PentaPlex-ISED, e foi validado pela análise das frequências dos alelos de cada marcador STR em uma amostra da população do Estado do Amazonas

comparativamente com as frequências alélicas de amostras das populações do Estado do Rio de Janeiro e do Estado do Espírito Santo.

2 REVISÃO BIBLIOGRÁFICA

2.1 GENES E GENOMA

Foram iniciados projetos, objetivando estudar o genoma de vários organismos, inclusive o genoma humano, após o desenvolvimento de várias técnicas baseadas na análise da estrutura molecular do DNA. Vale destacar que o corpo humano contém cerca de 100 trilhões de células, assim sendo, a identificação das seqüências e das funções dos genes abre perspectivas para o desenvolvimento de métodos para o diagnóstico e a cura de muitas doenças, como câncer, obesidade, diabetes, doenças autoimunes e hipertensão arterial (HARVERY *et al.*, 2005).

Considerado uma das maiores pesquisas da história da humanidade, o Projeto Genoma Humano (*Human Genome Project*, HGP) resultou num consórcio internacional que envolveu EUA, Europa, Japão, além de centros universitários de 18 países, inclusive do Brasil, com o objetivo de sequenciar, identificar e localizar os genes em cada um dos 23 pares de cromossomos humanos (Figura 01). O genoma humano haplóide é constituído, aproximadamente, de 3.2×10^9 pares de bases (pb) de DNA, e seu sequenciamento foi concluído e publicado em 2001 (LANDER *et al.*, 2001).

Buscando melhor esclarecer dados sobre o assunto acima abordado, diz-se que os genes são distribuídos nos cromossomos de forma linear. Cada gene tem uma posição definida, em um determinado cromossomo, que é chamada de *locus*. As formas alternativas de um gene podem ocorrer em um determinado *locus* cromossômico, sendo designadas por alelos. A seqüência específica das bases que compõem o gene, geralmente, corresponde a uma única proteína ou RNA complementares (HARVERY *et al.*, 2005).

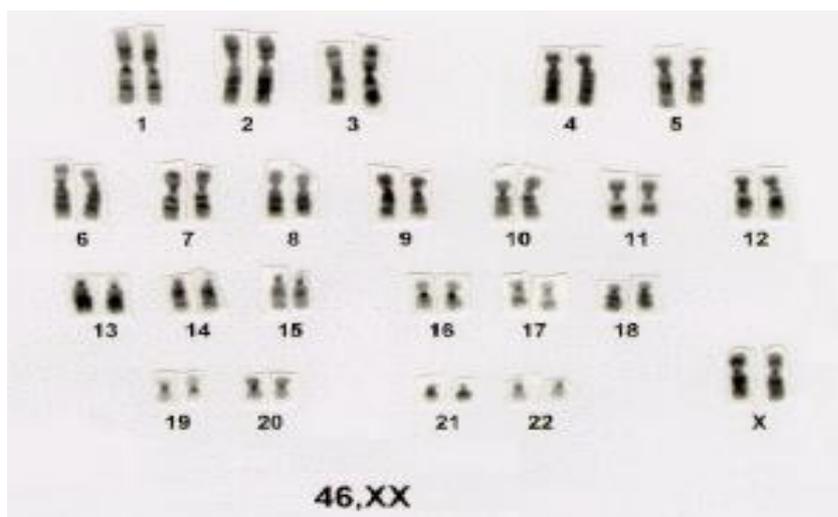


Figura 01: Esquema representando os 46 cromossomos do genoma humanos incluindo o cromossomo sexual feminino.

Lander e colaboradores, em 2001, estimaram que o genoma humano contém de 26.000 a 38.000 genes codificadores de proteínas.

O genoma humano compreende ao genoma nuclear e genoma mitocondrial. No genoma nuclear estão as regiões gênicas representando cerca de 25 % do genoma, e 75 % representando as regiões extragênicas. Nas regiões gênicas são constituídas por 10% de regiões codificantes (DNA codificador) e 90 % de regiões não codificantes, que são os pseudogenes, fragmentos de genes introns, sequências não traduzidas. Nas regiões extragênicas 40 % são de regiões moderadas e altamente repetitivas (repetições em tandem ou agrupadas e repetições dispersas), e 60 % são únicos ou baixo número de cópias (STRACHAN & READ, 2002) (Figura 02).

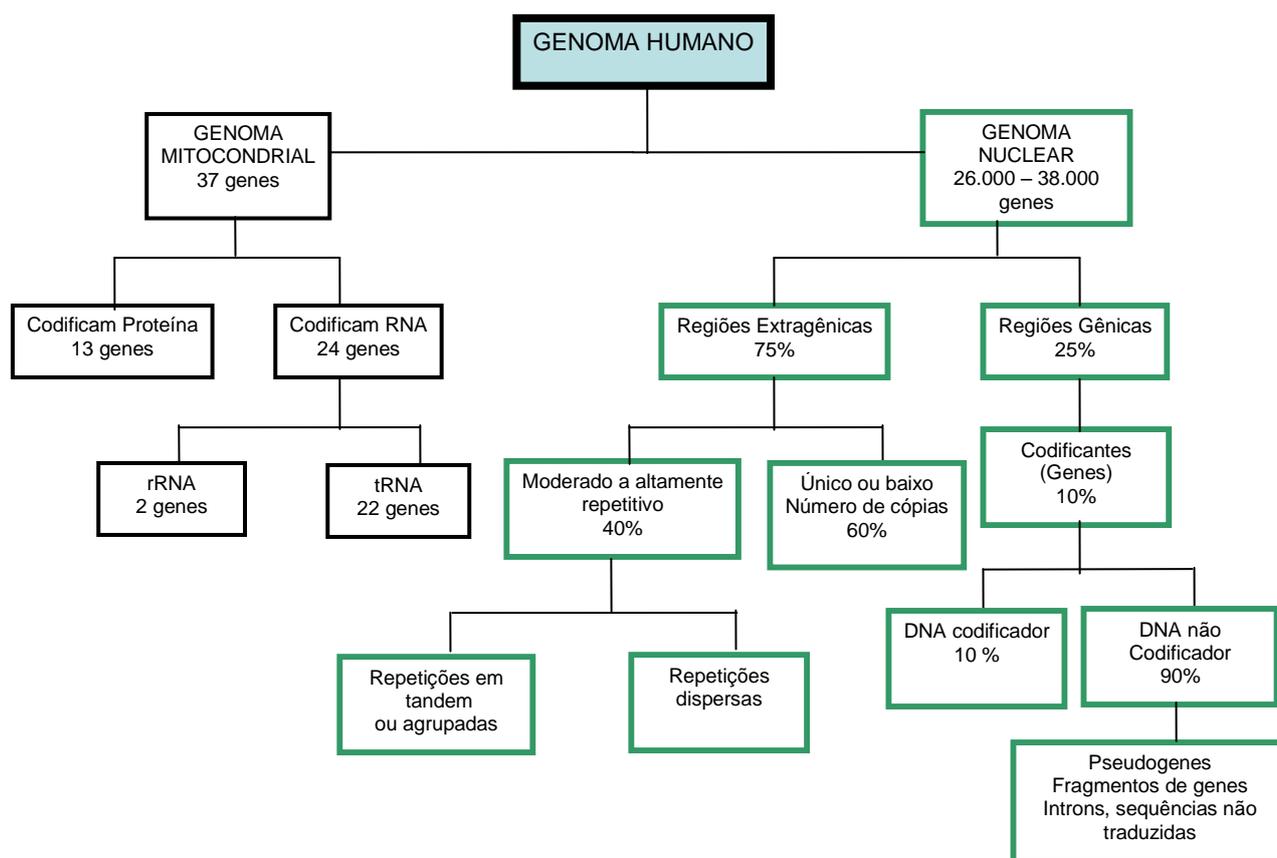


Figura 02: Esquema mostrando a distribuição do genoma entre o núcleo e a mitocôndria, e as porcentagens das sequências únicas e repetitivas. (LANDER *et al.*, 2005)

O DNA repetitivo divide-se em duas classes: sequência espaçada e sequência “em tandem”. As sequências espaçadas constituem a maior classe de DNA repetitivo, sendo distribuídas por todo o genoma. Os elementos espaçados compreendem 20 % do DNA total do ser humano. A repetição mais comum deste tipo é denominada de família *Alu*, pelo fato das sequências conterem sítios de clivagem para a enzima de restrição *Alu I*. Ressalte-se que estas são constituídas de, aproximadamente, 300 pb, compreendendo 5 % de todo o genoma. Sequências do tipo *Alu* e outras repetições espaçadas com menos de 500 pb são classificadas como elementos espaçados curtos (SINES – *short interspersed elements*). Os elementos espaçados propagam-se pelo genoma como elementos móveis ou transposons (FOWLER *et al.*, 1988).

2.1.1 O genoma mitocondrial

O genoma mitocondrial, constituído de 16.569 pb, é composto de apenas um cromossomo circular que pode se apresentar como um número de até 10 cópias por mitocôndria. Na matriz mitocondrial, o número de mitocôndrias varia de acordo com o tipo celular, podendo chegar a mais de 1000 organelas por células (ROBIN *et al.*, 1988). O mtDNA teve sua seqüência primária descrita por Anderson *et al.*, (1981) e, em seguida revisada por ANDREWS *et al.*, (1999).

Sendo o mtDNA humano de característica circular, este não apresenta introns, apresenta seus 37 genes dispostos em seqüência e compreendendo mais de 90 % do genoma. Na verdade, são 22 genes que codificam para RNAs transportadores, dois para RNAs ribossômicos e 13 para proteínas envolvidas na cadeia de transporte de elétrons e fosforilação oxidativa. Além da região codificante, há uma região com, aproximadamente, 1125 pb, conhecida por região controle (RC) – ou alça “D” (D-loop), devido às estruturas que se formam durante a replicação do genoma mitocondrial, podendo tais estruturas ser observadas por microscopia eletrônica. A taxa de mutação nesta região é, aproximadamente, dez vezes maior do que a da porção codificante (VIGILANT *et al.*, 1991).

A maior parte da variabilidade da região controle ocorre em duas regiões distintas, a primeira é conhecida como região hipervariável I (HVI), que compreende a seqüência entre as posições nucleotídicas 16.240 – 16.035, e a região hipervariável II (HVII), localizadas entre os nucleotídeos de 73 – 340. Há, também, uma região hipervariável menor (HVIII), localizada entre 16.024 a 16.569 e iniciais (1 a 576) VIGILANT *et al.*, (1989).

O nível de variabilidade das regiões hipervariáveis, aliado ao alto número de cópias de mtDNA em relação ao DNA autossômico, faz do mtDNA um alvo importante para análise de casos forenses, como a investigação de restos mortais de pessoas desaparecidas, desde que

um parente da linhagem materna disponibilize amostras para comparação com o mtDNA em questão (GINTHER *et al.*, 1992; HOLLAND *et al.*, 1993).

Apesar de o mtDNA ter sido amplamente utilizado em análise forense e em estudo evolutivo da população humana, trata-se de herança estritamente materna, não mendeliana com ausência total de recombinação. A herança mitocondrial é transmitida aos descendentes unicamente pela mãe, sendo retransmitida para as gerações posteriores somente por suas filhas (CONNOR *et al.*, 1994; TORRONI *et al.*, 1994; MACAULAY *et al.*, 1999; YAFFE *et al.*, 1999).

2.2 MARCADORES GENÉTICOS MOLECULARES

A partir do avanço nos estudos da genética molecular tem sido possível demonstrar, com sucesso, o emprego de diferentes marcadores moleculares em análise genética. Tais marcadores têm sido apontados como ferramentas eficientes para a identificação direta de genótipos, possibilitando determinar as frequências gênicas e a variabilidade genética das populações. A análise genética com marcadores moleculares tem criado abordagens inovadoras para identificação humana e para a determinação de vínculo genético entre indivíduos (FERREIRA *et al.*, 1998), assim sendo, considera-se “marcador molecular qualquer fenótipo molecular oriundo de um gene expresso, como no caso de isoenzimas, ou de um segmento específico de DNA que corresponde a regiões expressas ou não do genoma” (FERREIRA & GRATTAPLAGIA, 1998). Um marcador é definido como marcador genético, ao ser comprovado que seu comportamento está de acordo com as leis básicas da herança Mendeliana.

Atualmente, são utilizadas diversas metodologias para a análise genética com marcadores moleculares, que estão apresentadas no Quadro 01 (FERREIRA & GRATTAPAGLIA, 1998). Todos os 4 marcadores moleculares apresentados possibilitam a análise genética, diretamente a partir do DNA.

Quadro 01: As diferenças básicas entre Marcadores Genéticos Moleculares, baseado em Ferreira & Grattapaglia (1995), com modificações.

| Parâmetros | RFLP (<i>Restriction Fragment Length Polymorphism</i>) | RAPD (<i>Random Amplified Polymorphic DNA</i>) | STR / Microsatélites (<i>Short Tandem Repeats</i>) | AFLP (<i>Amplified Fragment Length Polymorphism</i>) |
|---|--|---|---|---|
| Forma de detecção do polimorfismo molecular. | Fragmentos de restrição de DNA detectados por hibridização com sonda de DNA genômico ou cDNA | Segmentos de DNA amplificados arbitrariamente | Sequências repetitivas em tandem amplificadas especificamente. | Segmentos amplificados por PCR após digestão de DNA com enzima de restrição. |
| Expressão genética | Co-dominante | Dominante | Co-dominante | Dominante |
| Número de alelos / <i>locus</i> | Multialélico | 2 alelos (binário) | Altamente multialélico | 2 alelos (binário) |
| Disponibilidade de marcadores no genoma | Praticamente ilimitada | Praticamente ilimitada | Praticamente ilimitada | Praticamente ilimitada |
| Distribuição no genoma | Distribuição aleatória no genoma, região de baixo número de cópias; medianamente repetitivas em telômeros e centrômeros (Minissatélites) | +/- ao acaso | +/- ao acaso | +/- ao acaso |
| Deteção | Extração de DNA (dezenas de microgramas); digestão c/ enzima envolvendo clivagem de DNA, hibridização de DNA. "Southern blotting". | Extração de DNA (dezenas de nanogramas); amplificação via PCR; eletroforese em agarose; coloração com EtBr | Extração de DNA (dezenas de nanogramas); amplificação especificamente via PCR; eletroforese em gel de agarose (EtBr) e de poliacrilamida coloração com prata; são automatizáveis em sistema multiplex | Extração de DNA (centenas de microramas); digestão c/endonucleases; ligação de adaptadores; amplificação por PCR eletroforese (alta resolução); auto-radiografia. |
| Informação ou desenvolvimento prévio necessário | Extração de DNA (centenas de microgramas); digestão com enzimas; eletroforese (agarose); transferência para membrana; clonagem de sonda; purificação do "insert"; marcação radioativa ou não; hibridização da sonda; lavagens de estringência; auto-radiografia. | Utilização imediatamente de <i>primers</i> arbitrários disponíveis no mercado; triagem de <i>primers</i> arbitrários adequados. | Extração de DNA, construção de biblioteca genômica enriquecida de microsatélites, sequenciamento do STR e de suas sequências flanqueadoras, desenho de <i>primers</i> para amplificação. | Triagem de combinações de enzimas e <i>primers</i> com bases arbitrárias adequadas. |
| Transferência da ocorrência de marcadores | Intra-específica, alta para espécies próximas | Intrapopulacional; média intra-específica | Pode ser transferido dentro de espécies de um mesmo gênero ou espécies próximas | Intrapopulacional; média intra-específica |
| Vantagens | Grande sensibilidade para detecção, alto grau de variabilidade, polimorfismo em populações diversas, aplicação na medicina forense. | Usado para identificar marcadores para diversas espécies, mesmo sem saber a sequência, uso de quantidade mínima de DNA. | Uso de quantidade mínima de DNA, altamente polimórfico, são automatizáveis em sistemas multiplex, permitindo avaliar rapidamente um grande número de indivíduos | Gera um grande número de fragmentos polimórficos, fornece mais informações que a técnica RAPD. |
| Desvantagens | Exigem quantidade maior de DNA, resultado com maior tempo, exigem DNA íntegro. | Baixa reprodutibilidade dos experimentos em função da troca de reagentes e termocicladores. | É limitada a amplificação de regiões genéticas. É menos informativo para identificação que o RFLP. É suscetível a contaminação. | Baixo conteúdo de informação por locus, envolve mais etapas na detecção, usa-se quantidade maior de DNA |

2.2.1 A técnica da PCR

A técnica da PCR (do inglês, *Polymerase Chain Reaction*, reação em cadeia da polimerase) foi descrita por Kary Mullis, em meados da década de 80 (MULLIS & FALOONA, 1987; SAIKI *et al.*, 1988). Esta consiste numa reação, catalisada por enzima DNA Polimerase, na qual uma pequena e específica região do genoma é amplificada “*in vitro*”, resultando em milhões de cópias de um único segmento do DNA. Kary Mullis recebeu o Prêmio Nobel, pelo desenvolvimento da técnica e suas inúmeras aplicações. Desde então, várias doenças, como a β -talassemias; distrofia muscular de Duchenne; síndrome de Lesch-Nyham; fibrose cística; doença de Tay-Sachs, entre outras doenças, várias delas têm sido diagnosticadas por PCR (DECORTE *et al.*, 1990). A sua rapidez e a sensibilidade das análises tornaram a PCR uma das ferramentas mais poderosas empregadas na área de diagnóstico molecular.

O processo da PCR consiste em três etapas principais: desnaturação, anelamento e extensão. O DNA, para ter um segmento amplificado, é desnaturado, por meio da elevação da temperatura de 92 °C para 95 °C. O anelamento dos *primers* ao DNA ocorre, devido à redução da temperatura de 60 °C para 35 °C. Esta temperatura depende da temperatura de desnaturação média (T_m) de cada iniciador com as sequências complementares que flanqueiam a região alvo. A extensão da nova fita de DNA é feita com a Taq DNA-polimerase e desoxirribonucleosídeos trifosfato, elevando-se a temperatura para 72 °C, a partir de cada terminal 3'OH dos iniciadores. A repetição destas etapas, por 20 a 40 ciclos, permite a amplificação de um segmento de DNA seguindo uma progressão geométrica. Inicia-se a reação, com quantidades mínimas de DNA (picogramas e nanogramas), e termina-se com grandes quantidades (microgramas) de DNA de uma sequência específica de interesse (FERREIRA & GATTAPAGLIA, 1998).

Um dos aspectos fundamentais da revolução causada pela PCR foi a possibilidade de serem geradas grandes quantidades de DNA de segmentos específicos do genoma. Os

fragmentos amplificados podem ser facilmente detectados, diretamente, por eletroforese em gel de agarose corado com brometo de etídio.

A reação de PCR é constituída pelos seguintes componentes: enzima *Taq* DNA-polimerase; desoxinucleotídeos trifosfatos (dNTPs: dATP, dCTP, dGTP, dTTP), *primers*, magnésio (MgCl), tampão da enzima e DNA (molde). A partir da PCR, tornou-se possível a identificação genética por DNA, em situações anteriormente impossíveis. Hoje, por exemplo, utilizam-se fios de cabelos sem bulbo; fragmentos de unhas; salivas usadas para colar selos; filtros de cigarros; goma de mascar, etc. (PENA, 2005).

2.2.2 Marcadores RFLPs

No final da década de 70, a utilização de enzimas de restrição permitiu a análise de polimorfismos de comprimento de fragmentos de restrição de DNA (*Restriction Fragment Length Polymorphism* – RFLP) (GRADZICKER *et al.*, 1974). Esta técnica baseia-se na possibilidade de detectar polimorfismos em sequências de DNA, pois mutações podem criar ou destruir sítios de enzimas de restrição, alterando o tamanho dos fragmentos produzidos pela ação de tais enzimas no DNA (FERREIRA & GRATTAPAGLIA, 1998).

Os RFLPs são suficientemente numerosos, permitindo uma cobertura adequada do genoma, proporcionando não só a construção de mapas genéticos de ligação, mas também a realização de análises genéticas, moleculares, além de várias aplicações como: clonagem de genes e mapeamento de QTLs (NODARI *et al.*, 1993).

Os fragmentos de DNA gerados por uma enzima de restrição podem variar de tamanho de um indivíduo para outro, devido, unicamente, a uma inserção ou deleção de nucleotídeos que amplifiquem na destruição ou criação de um sítio de enzima de restrição. Nas análises de DNA por RFLP, podem ser utilizadas sondas unilocais ou multilocais; nesse último caso, as sondas são de sequências minissatélites, as quais geram no processo muitas bandas, tornando um verdadeiro *fringerprinting* de cada indivíduo analisado. Tais sequências,

altamente repetidas, são referidas como VNTR - *Variable Number of Tandem Repeat* (JEFFREYS *et al.*, 1985b). O desenvolvimento de sondas de DNA para *loci* VNTRs possibilitou a primeira aplicação das análises de RFLP na medicina forense, por Sir Alec Jeffreys, em 1984. As sondas multilocais são mais informativas que as unilocais, pois permitem analisar várias regiões do genoma, em uma única hibridização. Na análise, quando são utilizadas sondas multilocais, ela vai ligar a várias sequências de DNA, produzindo inúmeras bandas, sendo a auto-radiografia resultante semelhante a um típico código de barras, como mostra Figura 3 (JEFFREYS *et al.*, 1985b).

Na década de 80, várias análises de regiões hipervariáveis do DNA foram utilizadas no Reino Unido, em casos forenses, como assaltos, estupros e etc. Tommie Lee Andrews em 1987, tornou-se o primeiro indivíduo a fazer referência a tipagem do DNA, e na corte dos Estados Unidos da América - EUA (Fourney, 1998), também o FBI (USA) usou sondas unilocais por bastante tempo.

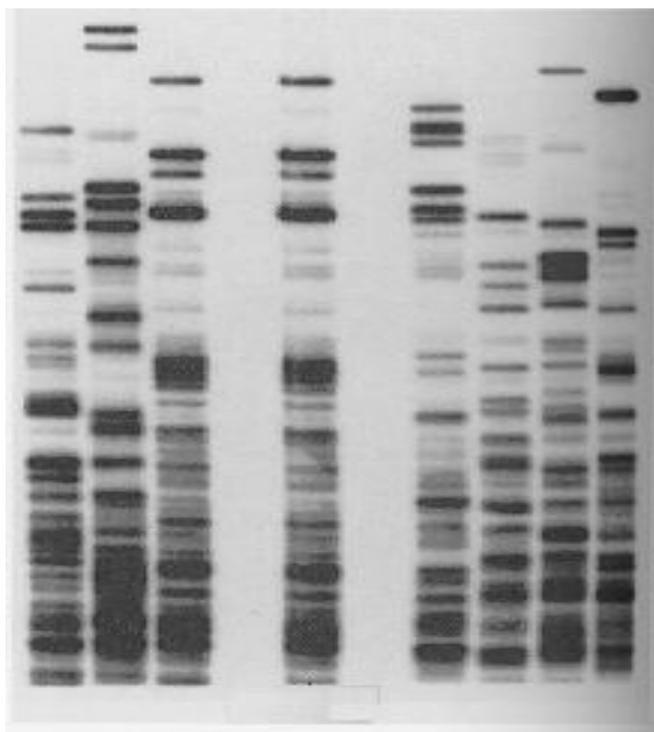


Figura 03: Análise de RFLP em humanos usando sondas multilocais minissatélites (VNTR), desenvolvidas por Jeffreys (1985b) (impressões digitais de DNA).

2.2.3 Marcadores RAPD

Segundo Williams *et al.*, (1990), a classe dos marcadores moleculares foi desenvolvida, independentemente, por dois grupos nos USA, sendo a tecnologia da mesma patenteada com o nome RAPD - *Random Amplified Polymorphic DNA*. Outro grupo propôs a denominação da técnica chamada AP-PCR - *Arbitrarily Primed-Polymerase Chain Reaction*, de Welsh & McClelland (1990). O desenvolvimento desses marcadores moleculares baseou-se na técnica da PCR, com a idéia de se utilizarem iniciadores mais curtos e de sequências arbitrárias para a reação de amplificação, eliminando-se a necessidade do conhecimento prévio de sequências (JEFFREYS, 1987; FERREIRA & GATTAPAGLIA, 1998). As aplicações da técnica RAPD incluem: a) obtenção de *fingerprint* genômico de indivíduos, variedades e populações; b) análise da estrutura e diversidade genética em populações; c)

estabelecimento de relacionamento filogenético entre diferentes espécies; d) construção de mapas genéticos.

Uma das características fundamentais dos marcadores RAPD é que eles se comportam como marcadores genéticos dominantes. Ao se observar uma banda RAPD no gel, não é possível distinguir se aquele segmento se originou a partir de uma ou de duas cópias da seqüência amplificada.

A diferença básica entre as técnicas de RFLP e RAPD é que enquanto os marcadores RFLP se baseiam na hibridização de DNA, o RAPD se baseia na amplificação de DNA, resultando numa série de vantagens práticas que podem ser resumidas em dois atributos: simplicidade e rapidez (PARAN *et al.*, 1991).

2.2.4 Marcadores AFLP

Os marcadores moleculares AFLP - *Amplified Fragment Length Polymorphism* ou polimorfismo de comprimento de fragmentos amplificados representa uma tecnologia capaz de obter um grande número de marcadores moleculares distribuídos em genomas de procariotos e eucariotos. Desde o desenvolvimento dessa técnica (Zebau, 1993), ela tem sido utilizada de forma crescente, com a finalidade de *fringerprinting*, mapeamento genético e construção de mapas genéticos (JEFFREYS *et al.*, 1988).

A análise do AFLP consiste em quatro etapas importantes que são: 1) o DNA genômico total do indivíduo é clivado com duas enzimas de restrição; 2) adaptadores específicos são ligados aos terminais dos fragmentos genômicos gerados pela clivagem; 3) uma fração dos fragmentos gerados é amplificada seletivamente por PCR utilizando *primers* específicos desenhados, para reconhecer seqüências dos adaptadores e de alguns nucleotídeos vizinhos; 4) a subpopulação de fragmentos amplificados é separada em gel de alta resolução (FERREIRA & GRATTAPAGLIA, 1998).

A vantagem do marcador AFLP é o fato de gerar um grande número de fragmentos que pode ser resolvido em um único gel. Outra vantagem é o poder de detecção de variabilidade genética, destacando-se, desta forma, por apresentar a maior robustez quando comparado com o ensaio RAPD. Este fato se deve, basicamente, por usar primers bem mais longos na etapa de amplificação por PCR, aumentando, significativamente a especificidade da amplificação, e evitando a competição durante a PCR.

As desvantagens do uso do marcador AFLP, por ser dominante e apresentar um baixo conteúdo de informação genética por *locus*, dizem respeito às seguintes ações: - envolve um número maior de etapas do que a análise de RAPD; - utiliza maior quantidade de reagentes; - necessitam de DNA em maior quantidade e melhor qualidade, para obter um bom resultado (FERREIRA & GRATTAPAGLIA, 1998).

2.2.5 Marcadores Microssatélites ou STRs

Os microssatélites, também denominados STRs (“Short Tandem Repeat” ou SSRs “Simple Sequence Repeats”) são regiões de DNA cujas unidades repetitivas possuem de 1 a 6 nucleotídeos repetidos em “tandem” (LITT & LUTY, 1989).

Os microssatélites se tornaram os marcadores de DNA repetitivos mais populares, porque são facilmente amplificados pela técnica da PCR, pois apresentam fragmentos que variam geralmente entre 100 a 300 pb (Schumn 1996). O número de repetições dos microssatélites pode ser muito variável entre indivíduos, o que torna estes marcadores muito eficientes para o uso na identificação humana (FRÉGEAU *et al.*, 1998).

As regiões STRs são encontradas tanto em regiões codificantes como em regiões não codificantes do genoma humano. Na região codificantes, usualmente, encontram-se localizadas em éxons e regiões flaqueadoras (EDWARDS *et al.*, 1991).

As seqüências repetitivas dos microssatélites são nomeadas pelo comprimento da unidade repetitiva, como di, tri, tetra, penta ou hexanucleotídica. As seqüências dos microssatélites variam, não somente no comprimento da unidade repetitiva e no número de repetições como também no padrão de disposição das repetições. São, frequentemente, divididos em várias categorias, tendo por base o padrão de repetições

Os *loci* STRs são de tamanhos pequeno, por isso permitem múltiplas amplificações simultâneas (Sistema multiplex de análise). Os produtos de PCR dos microssatélites STRs são, geralmente discretos, e possibilitam uma boa separação. A tipagem desses marcadores define o perfil genotípico de cada pessoa envolvida na análise. Após a amplificação, o produto pode ser analisado por eletroforese em gel de poliacrilamida e corado por prata (ALLEN *et al.*, 1989). O resultado de análises de regiões STR tem sido utilizado como prova em testes de vínculo genético, em substituição à análise de RFLP, devido à rapidez e automatização das técnicas envolvidas na tipagem destas regiões e, principalmente, por necessitar de muito menos DNA para análise (em torno de 1 – 30 ng).

As vantagens de marcadores microssatélites STRs são: a expressão co-dominante e o multialelismo. Esses marcadores possuem elevado grau de polimorfismo e são, geralmente, distribuídos ao acaso. Marcadores STRs permitem a mais completa cobertura de qualquer genoma eucarioto (FERREIRA & GATTAPAGLIA, 1998).

Outra vantagem dos STRs são facilmente analisados mesmo quando o DNA apresenta degradação ou está disponível em quantidade limitada. Os resultados são mais fáceis de interpretar do que os obtidos nas análises por hibridação ou pela análise do produto amplificado e digerido (NAKAMURA, *et al.*, 1987; KASAI, *et al.*, 1990).

A necessidade de análise de regiões curtas de DNA é recomendada, em certos casos, em razão do alto grau de fragmentação da amostra de DNA a ser analisado fato que tornou esta metodologia eleita quando da identificação de vítimas de acidentes, como por exemplo, o

desastre do ataque terrorista ao World Trade Center em Nova York, em 2001 e o atentado contra a sede da comunidade Judaica em Buenos Aires, em 1992 (CORACH *et al.*, 1995).

Outros marcadores polimórficos localizados tanto nos cromossomos autossômicos como nos sexuais já foram analisados – STR autossômicos e Y-STR (Kayser, *et al.*, 2001) e STR-X (Mingroni-Netto, *et al.*, 2002). Com as frequências alélicas de tais marcadores é possível caracterizar populações, através da estimativa de diversidade gênica e de mistura étnica.

Dentre os marcadores de DNA nuclear, encontra-se os *loci* ligados ao cromossomo Y, os quais possuem herança paterna e sua porção sexo-específica é transmitida sem mudanças dos pais para os filhos, exceto por ocorrência de mutações que se acumulam através de gerações. O valor dos polimorfismos ligados ao Y é reconhecido e usado para estudos populacionais e de evolução humana (HAMMER & HORAI, 1995; JOBLING & TYLER-SMITH, 1995; CARVALHO-SILVA, *et al.*, 2001)

Além do seu uso na identificação humana, um grande número de marcadores STR já foi caracterizado visando diagnóstico do câncer e de doenças hereditárias, tais como fibrose cística (MORRAL & ESTIVILL, 1992) e a doença de Huntington (WEBER *et al.*, 1992).

2.2.5.1 Características especiais das regiões microssatélites STRs

- Produto “Stutter”

Os produtos “stutter” são amplicons que apresentam, tipicamente, uma ou mais unidades repetitivas a menos que o alelo verdadeiro, e surgem durante a PCR, devido ao deslizamento do filamento durante a polimerização. Normalmente, estes produtos contêm uma unidade de repetição a menos, na proporção aproximadamente de 5 – 15 %, em relação ao alelo principal. Este fenômeno já foi extensamente descrito para o *locus* vWA, assim como a porcentagem prevista destes produtos para os *loci* STRs, comumente analisados (WALSH,

1996). Nestas circunstâncias, outro tipo de “stutter” pode ocorrer em decorrência da adenilação dos produtos da PCR pela enzima Taq DNA polimerase durante a reação. A adição de nucleotídeos adenosina pela enzima na extremidade 3` dos produtos da PCR resulta em produtos de PCR maiores que o tamanho real do alelo (SILVA, 2004).

O impacto da produção de “stutter” na análise dos produtos de PCR se reflete com maior intensidade, em casos contendo misturas de DNA. A tipagem do componente minoritário pode facilmente ser confundida com os artefatos. A tipagem de DNAs em baixa concentração, provenientes de vestígios e restos mortais, também pode ser dificultada pela presença destas bandas “fantasmas”, como mostra a Figura 4.

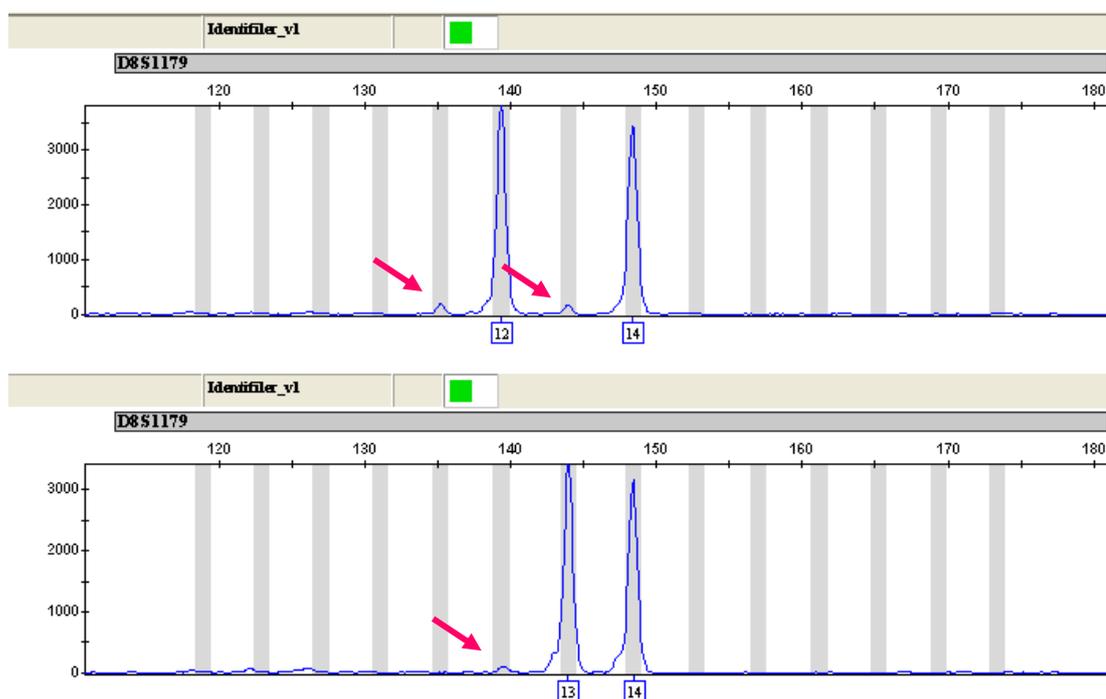


Figura 04: Mostra alelos STRs apresentando bandas *stutters*, indicadas pelas setas vermelhas no locus D8S1179 (Applied Biosystems. 2006) ABI Expert.

Uma das estratégias para minimizar o fenômeno é a inclusão de uma etapa adicional de extensão, após o término dos ciclos de amplificação por PCR. Esta etapa adicional, em

geral, é realizada à temperatura de 60 °C, por 30 minutos, visando à adenilação de todos os produtos de PCR (KIMPTON *et al.*, 1993).

- **Microvariância**

Outra característica especial dos microsatélites é a ocorrência de alelos que contém unidades incompletas chamadas microvariantes. Estes alelos são raros e podem diferir dos alelos comuns por uma ou mais bases, ocorrendo, devido a inserções, deleções ou troca de nucleotídeos. O locus TH01 é um microvariante, por apresentar repetições completas (AATG), e uma repetição parcial de três bases (ATG), diferenciando-se do alelo 10 por uma deleção de um único nucleotídeo, porque a sétima repetição perdeu uma única adenina da unidade repetitiva AATG (PUERS *et al.*, 1993).

Segundo Silva (2004), já foram observados microvariantes para a maioria dos *loci* STRs, eles são detectados à medida que aumenta o número de indivíduos analisados. Os *loci* mais polimórficos encontrados foram FGA, D18S51 e D21S11, já que são constituídas de unidades repetitivas maiores e mais complexa, como mostra a Figura 05.

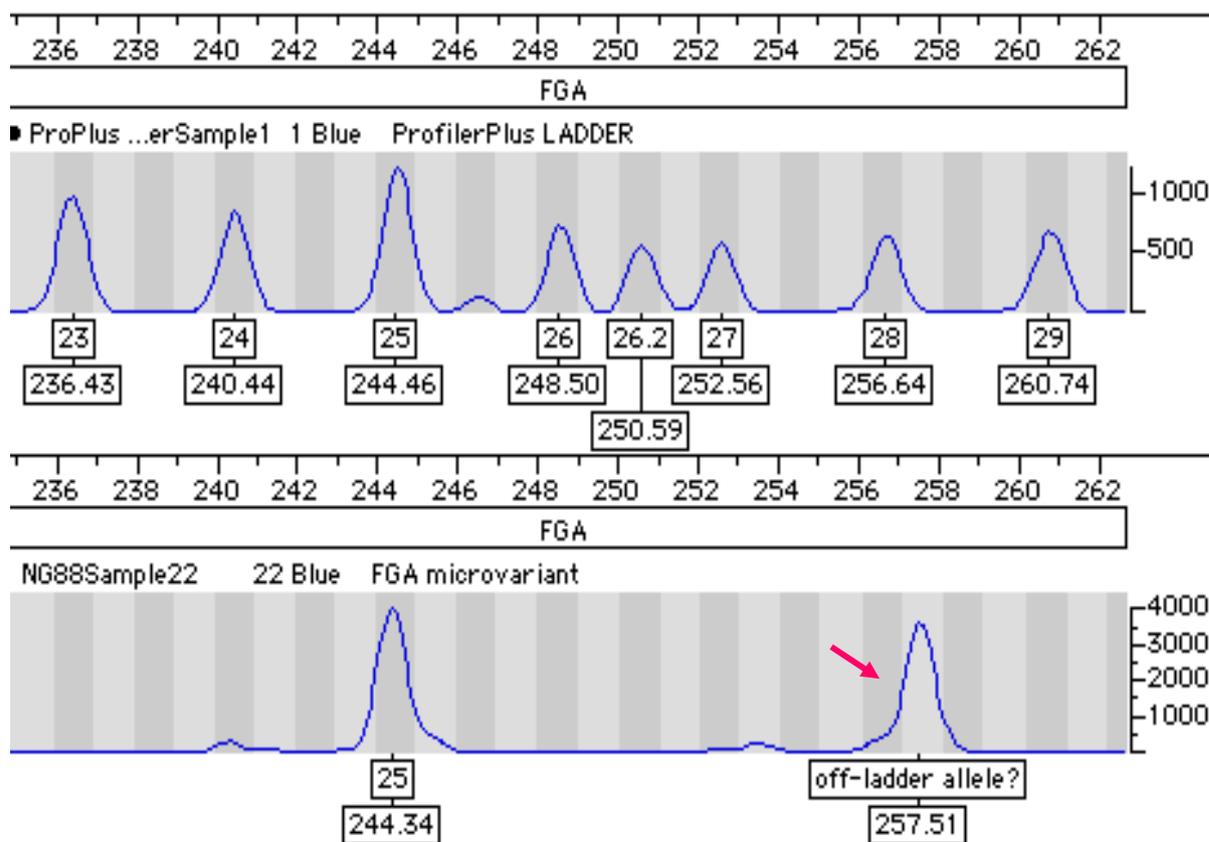


Figura 5: Microvariante do *loci* FGA mostrando com seta vermelha (Applied Biosystems 2006) ABI Expert.

- Mini STRs

O marcador mini STRs é uma nova tecnologia de amplificação, desenvolvida com o objetivo de possibilitar a produção de amplicons menores, utilizando-se iniciadores que flanqueiam uma região reveladora de alvo de menor tamanho (COBLE & BUTLER, 2005). Neste caso, utiliza-se o mencionado marcador, quando o DNA a ser analisado encontra-se extremamente degradado.

A observação de que produtos de PCR menores, obtidos pela amplificação de *loci* STRs, produziam uma melhor taxa de amplificação com amostras degradadas foi demonstrada pela primeira vez por Whitaker *et al.*, (1995), sendo seus resultados corroborados por diversos pesquisadores nos últimos anos (CHUNG *et al.*, 2004).

Uma das principais vantagens dos miniSTRs é a eficiência das análises destes marcadores, por serem compatíveis com a dos kits comerciais de STRs. Alguns *loci* do CODIS (banco de dados de DNA financiados pelos Estados Unidos, Federal Bureau of Investigation – FBI), não podem ser processados em menores *amplicons*, devido às regiões que flanqueiam as repetições não serem compatíveis com o desenho de bons iniciadores, ou devido à existência de alguns alelos muito grandes no *locus* em questão (COBLE & BUTLER, 2005).

2.3 DESENVOLVIMENTO DE SISTEMAS MULTIPLEX STR

Na década de 90, surgiram os primeiros “kits” para identificação humana, contendo amplificação simultânea de duas ou mais regiões microssatélites. O rápido desenvolvimento das novas metodologias para a tipagem de regiões microssatélites promoveu uma redução no tempo necessário para a investigação criminal e de vínculo genético por análise de DNA (BLAKE *et al.*, 1992).

O sistema multiplex relacionado à detecção da distrofia muscular Duchene foi o primeiro descrito por Chamberlain *et al.*, (1988). Nos Estados Unidos da América do Norte, os primeiros sistemas para tipagem de DNA humana capazes de amplificar, simultaneamente, duas ou mais regiões STR, foram desenvolvidos pela empresas Promega e Perkin-Elmer, em colaboração com a Roche Molecular Systems, no início da década de 90. Na Inglaterra, o primeiro sistema multiplex STR, desenvolvido pelo Serviço de Ciência Forense da Inglaterra, reunia iniciadores que amplificavam os *loci* THO1, FES/FPS, vWA e F13A01 (KIMPTON *et al.*, 1993). Outro sistema multiplex, contendo um conjunto de iniciadores para a amplificação dos *loci* CSF1PO, TPOX e THO1, frequentemente referido como CTT, foi disponibilizado pela empresa PROMEGA (BUDOWLE *et al.*, 1997). Posteriormente, outros sistemas comerciais, apresentando um número maior de marcadores foram lançados no mercado, com

o objetivo de aumentar, ainda mais, o poder de discriminação das análises em seres humanos (PROMEGA CORPORATION, 1999; 1997).

A *Applied Biosystems* lançou o sistema *16 plex Identifier kit*, que amplifica 13 *loci*, amelogenina e dois STRs tetranucleotídeos D2S1338 e D19S433 (COLLINS *et al.*, 2004). Esses *loci* extras foram identificados no final dos anos 90, na busca por novos tetranucleotídeos, proporcionando um aumento no poder de discriminação e melhoria na interpretação de misturas, nos casos em que há produção de *stutter* nos *loci* pentanucleotídeos.

Dentre as vantagens do sistema automatizado, destaca-se a possibilidade de análise simultânea de número maior de *loci* STR. Isto é possível, mesmo havendo sobreposição de alelos amplificados no gel de poliacrilamida, já que estes poderão ser discriminados, tendo em vista a faixa de emissão eletromagnética característica de cada fluoróforo que é utilizado para marcação dos iniciadores (WALSH *et al.*, 1996; SCHYMA *et al.*, 1999; BODOWLE *et al.*, 2000; EVETT *et al.*, 1991).

Uma característica importante incorporada aos sistemas automatizados é a detecção de somente uma das fitas do produto amplificado, já que somente um dos elementos de cada par dos iniciadores encontra-se marcados com fluoróforos. Desta forma, a precisão da leitura é aumentada, evitando-se possíveis distorções decorrentes da migração diferencial de DNA em dupla-fita. Os resultados da tipagem também são informatizados, podendo ser analisados os alelos com diferentes *softwares*. A análise automatizada de regiões STR é a metodologia eleita pelo FBI, a Polícia Federal dos Estados Unidos, para identificação de indivíduos e tipagem de criminosos, devido à rapidez, acuidade e alto poder de discriminação decorrentes da análise de múltiplas regiões STR (FRÉGEAU *et al.*, 1999). Na Tabela 01 estão descritas as principais características dos *loci* amplificados pelo sistema de tipagem multiplex STR, referente ao kit PowerPlex 16, que foi validado em 2000, e disponibilizado comercialmente

em 2001 (PROMEGA). Na Tabela 02 estão descritas as características dos *loci* amplificados pelo sistema de tipagem Identifiler (Applied Biosystems) (BUDOWLE *et al.*, 1998).

Tabela 01: Representa o Sistema PowerPlex 16, com os 16 *loci* STR (Promega).

| Locus | Definição do Locus | Localização Cromossômica | Tamanho (pb) | Seqüência Repetitiva |
|--------------|--|---------------------------------|---------------------|-----------------------------|
| Penta D | ND | 21q | 376-441 | AAAGA |
| FGA | Cadeia alfa do gene da fibrina | 4q28 | 326-444 | CTTT |
| TPOX | Gene humano tiróide Peroxidase | 2p23-pter | 262 | GAAT |
| D8S1179 | ND | 8 | 203-247 | (TCTA)(TCTG) |
| vWA | Gene humano fator Von Willebrand 28 | 12p12-pter | 123-171 | (TCTG)(TCTA) |
| Penta E | DN | 15q | 379-474 | GATA |
| D18S51 | HUMUT574 | 18q21.3 | 290-366 | AGAA |
| D21S11 | HUMD21LOC | 21q21 | 203-259 | (TCTA) (TCTG) |
| TH01 | Gene humano tirosina hidroxilase | 11p15.5 | 156-195 | TCAT |
| D3S1358 | NA | 3p | 115-147 | (TCTG)(TCTA) |
| CSF1PO | c-fms proto-oncogene para receptor CSF-1 | 5q33.3-34 | 295-327 | AGAT |
| Amelogenin | Cromossomo humano XY gene amelogenina – like | X: p22.1–22.3 Y: p11.2 | 106 (X), 112 (Y) | NA |
| D16S539 | ND | 16q24-qter | 264-304 | AGAT ^a |
| D7S820 | ND | 7q11.21-22 | 215-247 | AGAT ^a |
| D13S317 | ND | 13q22-31 | 165-197 | AGAT ^a |
| D5S818 | ND | 5q21-31 | 119-151 | AGAT ^a |

^(a) Unidade de repetição que representa todas as quatro possíveis permutações (AGAT, GATA, ATAG ou TAGA).

(ND) não definida

(NA) Não aplicável

Tabela 02: Mostra características do Sistema Identifiler, com os 16 *loci* STR

| Locus | Definição do Locus | Localização Cromossômica | Tamanho (pb) | Seqüência Repetitiva |
|--------------|--|---------------------------------|---------------------|-----------------------------|
| FGA | Cadeia alfa do gene da fibrina | 4q28 | 326-444 | CTTT |
| TPOX | Gene humano tiróide Peroxidase | 2p23-pter | 262 | GAAT |
| D8S1179 | ND | 8 | 203-247 | (TCTA)(TCTG) |
| vWA | Gene humano fator Von Willebrand 28 | 12p12-pter | 123-171 | (TCTG)(TCTA) |
| D18S51 | HUMUT574 | 18q21.3 | 290-366 | AGAA |
| D21S11 | HUMD21LOC | 21q11.2 – q21 | 203-259 | (TCTA) (TCTG) |
| TH01 | Gene humano tirosina hidroxilase | 11p15.5 | 156-195 | TCAT |
| D3S1358 | NA | 3p | 115-147 | (TCTG)(TCTA) |
| CSF1PO | c-fms proto-oncogene para receptor CSF-1 | 5q33.3-34 | 295-327 | AGAT |
| Amelogenin | Cromossomo humano XY gene amelogenina – like | X p22.1–22.3 e Y p11.2 | 106, 112 | NA |
| D16S539 | ND | 16q24-qter | 264-304 | AGAT ^a |
| D7S820 | ND | 7q11.21-22 | 215-247 | AGAT ^a |
| D13S317 | ND | 13q22-31 | 165-197 | AGAT ^a |
| D5S818 | ND | 5q21-31 | 119-151 | AGAT ^a |
| D2S1338 | ND | 2q35-37.1 | | |
| D19S433 | ND | 19q12-13.1 | | |

(a) Unidade de repetição que representa todas as quatro possíveis permutações (AGAT, GATA, ATAG ou TAGA).

(pb) Pares de bases (ND) não definida

2.3.1 Validação de sistemas multiplex de STR

No processo requerido para validação de qualquer sistema multiplex de identificação humana, com vistas à identificação e análise forense, parâmetros fundamentais devem ser avaliados. Para tanto, o processo de tipagem deve ser submetido a uma série de estudos (MICKA *et al.*, 1999).

Na estratégia de validação estão relacionados aspectos importantes dos procedimentos técnicos e metodológicos que são críticos e que precisam de cuidados, para poderem ser empregados na análise de identificação humana. Esses aspectos estão relacionados com o estudo da reprodutibilidade do método, especificidade e sensibilidade. Além destes aspectos, são realizados estudos populacionais, bem como efetuada a caracterização dos parâmetros da PCR (DANIELS *et al.*, 2004).

No estudo de reprodutibilidade, amostras de DNA oriundas de diferentes amostras biológicas devem ser amplamente analisadas, inclusive, por diferentes laboratórios. Neste contexto, amostras de DNA de diferentes espécies biológicas, como sangue, saliva, cabelo, urina e sêmen de um mesmo indivíduo devem ser tipados pelo mesmo sistema multiplex.

No que diz respeito aos estudos da especificidade, estes são requeridos, para se verificar o grau de especificidade do sistema de tipagem para a espécie humana. No contexto forense, isto é de fundamental importância, uma vez que amostras biológicas podem estar contaminadas por materiais biológicos provenientes de outros organismos (SILVA *et al.*, 2002)

Prosseguindo-se, o estudo da sensibilidade está relacionado com a determinação da quantidade mínima de DNA necessária para amplificação de todos os *loci* STR que constituem um sistema multiplex. Para a determinação desta etapa, é necessária a preparação da reação da PCR, contendo diferentes concentrações de DNA, e uma amostra de DNA padrão. Este fato se faz especialmente importante para a análise forense, uma vez que

amostras biológicas de diferentes fontes podem estar misturadas, em diferentes concentrações, em várias evidências (WILLOT & CROSSE, 1986; REYNOLDS *et al.*, 1991; COMEY, 1995; HAMM & SHUELER, 1995).

Vários fatores intrínsecos a um sistema multiplex podem afetar sua sensibilidade. Tais fatores podem estar relacionados com a seqüência dos iniciadores em relação às seqüências alvo, a natureza dos fluoróforos empregados na marcação dos iniciadores e a sensibilidade do equipamento utilizado na sua detecção. Estes fatores podem ser ajustados durante o processo de desenvolvimento do sistema multiplex (MICKA *et al.*, 1999).

No que tange à análise populacional, esta deve ser feita, a partir de uma base de dados de freqüências dos alelos na população. A freqüência de um determinado perfil genético observado para uma amostra de DNA pode ser calculada, por meio do produto das freqüências individuais dos alelos (MANTA, 2008).

Buscando-se a validação de um sistema multiplex, estudos relacionados com a tipagem de indivíduos de uma mesma família são, freqüentemente, realizados, com o intuito de detectar a ocorrência de desvio do modelo de herança genética para os *loci* analisados, devido à ocorrência de fenômenos mutacionais (NADIR *et al.*, 1996).

Pode-se observar outro tipo de evento mutacional em análise de microssatélites, ocorrendo na região de anelamento dos iniciadores, podendo resultar em uma não amplificação do alelo STR. Os alelos não detectados são denominados alelos nulos, fenômeno conhecido como “drop-out”, que tem sido observado, principalmente no *locus* Vwa (BUTLER, 2001). O percentual de alelos nulos na população pode ser estimado, levando-se em conta o número de indivíduos homozigotos observados e esperados, segundo a Lei de Hardy-Weinberg (HARDY, 1908).

2.4 GENÉTICA DE POPULAÇÕES

A genética de populações se baseia em descrever a variabilidade genética nas populações e na investigação dos mecanismos que a mantêm, tais como mutação, recombinação e fluxo gênico. Na grande maioria das populações naturais esta variabilidade é grande, fazendo com que os polimorfismos genéticos sejam comuns e amplamente estudados.

O estudo da genética de população tem como objetivo analisar as consequências das leis de herança mendeliana sobre a composição genética de uma população, e as forças que determinam e alteram esta composição. Em qualquer espécie, uma grande quantidade de variações genéticas, dentro e entre as populações, surge da existência de vários alelos em *loci* diferentes. A diversidade e a estrutura genética de populações podem ser investigadas, com o objetivo de se avaliar a variabilidade genética dentro e entre as populações. Além de indicar o grau de polimorfismo dos marcadores, esses estudos possibilitam avaliar o fluxo gênico, o grau de endocruzamento, extinção e migração das populações, possibilitando, portanto, recolher dados acerca dos fatores que podem estar relacionados com o desvio do Equilíbrio de Hardy-Weinberg (EHW) (ALLENDORF & LUIKART, 2007).

Toda população possui um conjunto gênico (gene pool ou pool gênico) que lhe é particular. Esse conjunto pode ser conhecido, por meio de suas frequências alélicas e genotípicas dos *loci* desta população (MATHER & JINKS, 1994).

A frequência de um determinado alelo em uma população pode ser alterada por mutações decorrentes de seleção, migração ou derivação genética, ou ainda fornecer valores diversos por efeitos de amostragem aleatória. Em uma população idealizada, na qual não está atuando forças modificantes, uma população entrecruzada, aleatoriamente, mostraria frequências genotípicas constantes para um determinado *locus* (GRIFFITHS, 1996). São necessárias cinco condições, para que uma população esteja em equilíbrio de Hardy-

Weinberg: tamanho populacional grande; cruzamentos produzidos randomicamente; ausência de mutação; ausência de migração; ausência de seleção.

As populações evoluem, ao longo das gerações, o que nos leva à conclusão óbvia de que a lei de Hardy-Weinberg não se aplica as situações reais, pois sempre existem fatores de evolução atuando. Os genes não se dividem na meiose sempre com exatidão (mutação), os genótipos não são transmitidos em taxas uniformes (seleção), as populações não são infinitamente grandes e os cruzamentos não são ao acaso (deriva genética), tampouco as populações estão isoladas (migração). Todos estes fatores tendem a alterar o equilíbrio das populações e, alteram, assim, as frequências gênicas, designando-se por fatores de evolução.

O padrão de heterozigidade na população é um dos parâmetros de diversidade genética mais estudada, e a sua diminuição pode ser provocada pelo aumento da taxa de endocruzamento. Este fenômeno é observado, principalmente, em populações pequenas e isoladas (DANIEL *et al.*, 1997).

Segundo Nei e colaboradores (1973), o grau de estruturação genética das populações permite verificar a ocorrência de populações subdivididas, fazendo referências às populações separadas por situações geográficas, lingüística, étnica e cultural. A verificação da existência de estruturação gênica, ou seja, de subdivisão das populações, pode ser feita, a partir de heterozigidade das subpopulações em relação à heterozigidade total.

2.4.1 Formação da população do Brasil

No período da colonização do Brasil, entre 1500 e 1850, chegaram aproximadamente 500.000 portugueses e 4 milhões de africanos que dividiram o espaço da nova terra com 2,4 milhões de ameríndios então existentes (Curtin 1969; IBGE 2000). No período seguinte, chegaram oficialmente no Brasil mais de 5 milhões de imigrantes, a maioria portugueses (espanhóis, alemães, sírios, libaneses, japoneses e outros não ultrapassaram 30% deste total)

IBGE 2000). Assim, a formação da população brasileira é resultado de mistura étnica entre populações de três diferentes continentes colonizadores europeus, escravos africanos e ameríndios. Em decorrência disso, a população brasileira atual é considerada tri-híbrida, embora as proporções dos três grandes componentes étnicos variem consideravelmente conforme a região geográfica considerada (BARRETO, 1959; SALZANO & FREIRE-MAIA, 1967).

Os negros foram trazidos para o continente Americano contra a sua vontade, para trabalharem como escravos, cerca de 6 milhões e 200 mil escravos somente no século XVIII (MATOSO, 1982). A distribuição do negro no território brasileiro foi bastante heterogênea, seguindo os seguintes percentuais segundo Viana (1988): Rio de Janeiro (38 %), Bahia (25 %), Pernambuco (13 %), São Paulo (12 %), Maranhão (7 %) e Pará (5 %).

O processo de miscigenação do Brasil, podem-se apontar o Estado de Minas Gerais e a região litorânea do Nordeste e do Leste como área em que verificou-se mais intensa a miscigenação branco-negra, no interior do Nordeste e o externo Norte (Amazonas, Pará e parte do Maranhão) houve o processo de mestiçagem branco-indígena, pode-se observar na área do Oeste (Mato Grosso e Goiás) (SAZANO & FREIRE-MAIA, 1967).

Segundo censo do IBGE do Estado do Espírito Santo revelou os seguintes números: 1.961.318 Pardos (53,7%), 1.638.236 Brancos (48,8%), 280.000 Negros (7,5%), 13.000 Asiáticos ou Índios (0,4%). Essa grande miscigenação, ocorreu nas décadas de 1940 até 1980, quando as grandes idades do estado começaram a se industrializar, isso atraiu muitas famílias do interior do estado (os italianos), para a Grande Vitória, onde havia grande concentração, sobretudo de pessoas de ascendência portuguesa e espanhola, além da boa parte da população local ser parda, esse êxodo rural, causou grande miscigenação de culturas no estado.

Na região Amazônica, onde está localizado o estado do Amazonas, constitui uma das áreas de mais antiga ocupação européia no Brasil (RIBEIRO, 1996), e os primeiros relatos sobre a ocupação dessa região datam de 1650 (TEIXEIRA; RIBEIRO DA FONSECA, 1998), instalando-se ali jesuítas, soldados espanhóis seguidos pelos colonos portugueses, formando núcleos de ocupação permanente (PINTO, 1993). Até meados do século XVIII, quase toda a Amazônia pertencia à Espanha. Nesse longo período permaneceu praticamente desconhecida, sendo visitada apenas por missionários e expedicionários enviados pela corte Portuguesa ou Espanhola, também se observam a presença de peruanos e bolivianos em toda região norte (BENCHIMOL, 1977).

Segundo a estimativa do IBGE (2008), o Amazonas é o segundo estado mais populoso da região Norte do Brasil com 3.341.096 habitantes. Destes, cerca de 103.066 são indígenas, dados apresentados pela FUNAI-AM, divididos em 65 etnias, que correspondem a 4,0% da população total do estado. No censo do IBGE (2000), 70,3 % dos amazonenses se declararam pardos; 21 % brancos; 4,3 % negros e apenas 4,2 % se definiram como amarelos ou indígenas. Os afros descendentes, no estado do Amazonas, vivem em sua maioria na cidade de Manaus. A sua capital, Manaus, está localizada no extremo norte do país, a 1932 km da capital federal, sendo a maior cidade da região Norte, com cerca de 1,7 milhões de habitantes, seguida por Belém do Pará com 1,4 milhões de habitantes. Manaus, uma das capitais que mais recebe migrantes no Brasil, cresce, desordenadamente, com muitas áreas ocupadas de forma ilegal por invasões. O município de São Gabriel da Cachoeira possui o maior número de indígenas, cerca de 23 mil.

2.4.2 Variabilidade Genética

A variabilidade genética de uma população é determinada por certos fatores como tamanho efetivo da população, deriva genética, mutação e fluxo gênico.

Os diferentes alelos de um mesmo gene que variam entre si, numa dada população (PRIMACK, 2006). A variabilidade genética pode ser caracterizada pelos seguintes índices de diversidade: número médio de alelos por *locus*; heterozigosidade média esperada (H_e); heterozigosidade média observada (H_o); frequências alélicas.

O número médio de alelos por *locus* é a soma do número total de alelos observados em todos os *loci* divididos pelo número de *locus* analisados.

A heterozigosidade esperada (H_e) estima a proporção esperada de indivíduos heterozigotos em um *locus*. H_e é dada por:

$$H_e = (1 - \sum p_i^2) (n/n-1)$$

onde p_i é a frequência do alelo i , e n é o número total de alelos observados.

A Heterozigosidade média observada (H_o) é a proporção observada de indivíduos heterozigotos em um dado loco.

$$H_o = \frac{\sum \text{número de indivíduos heterozigotos}}{\text{Número total de indivíduos analisados}}$$

2.5 IDENTIFICAÇÃO HUMANA

Os primeiros caracteres hereditários utilizados na identificação humana acerca de vinculação genética foram os relacionados com o grupamento sanguíneo do sistema ABO, que permitia a exclusão de paternidade. Outros sistemas dentro do grupo sanguíneo também passaram a serem utilizados, juntamente com o ABO, como os sistemas MNS e RH

(Landsteiner e Wiener, 1940). Um grupo de proteínas séricas e plasmáticas, também foram utilizados, em investigação de vínculo genético (JAYLE & GUILLARD, 1947).

A partir do avanço da tecnologia molecular por DNA, os testes que eram baseados em marcadores protéicos tornaram-se obsoletos. A molécula de DNA é mais estável e permite aplicação imediata na identificação humana de informações geradas por análises genômicas.

Em 1985, o grupo de Alec Jeffreys desenvolveu o primeiro estudo que permitiu a aplicação das regiões hipervariáveis na área forense. Buscando-se detectar os polimorfismos de tamanho de fragmentos de restrição (RFLP), foram utilizadas sondas com sequências de base complementar em especial as que apresentavam uma enorme variabilidade, o que dava um alto poder de resolução.

Na década de 90, os microssatélites (STRs) foram reconhecidos como ferramentas valiosas para a identificação humana, devido à utilização da PCR para sistema multiplex. A combinação do número elevado de STRs disponível, por estarem espalhados por todo o genoma, e em razão da elevada variabilidade e relativa facilidade de amplificação, STRs representam excelentes marcadores para análise genética (SCHAFER; HAWKINS, 1998).

O Departamento de Investigação Federal dos Estados Unidos (Federal Bureau of Investigation - FBI) foi o líder no desenvolvimento da tecnologia de tipagem do DNA para o uso na identificação de autores de crimes violentos, elucidando casos que antes eram impossíveis. Em 1997, o FBI propôs a seleção de 13 regiões STRs, localizados em cromossomos humanos diferentes, para constituir o núcleo do banco de dados nacional dos Estados Unidos, chamado de CODIS (*Combined DNA Index System*), é um banco de dados de DNA financiados pelos Estados Unidos, Federal Bureau of Investigation (FBI), que armazena perfis de DNA criados por federais, estaduais e laboratórios de criminalidade local nos Estados Unidos, tem a capacidade de auxiliar na identificação de suspeitos de crimes. O CODIS é formado pelos marcadores STR (TPOX, D3S1358, D5S818, FGA, CSF1PO,

D7S820, D8S1179, TH01, vWA, D13S317, D16S539, D18S51 e D21S11) e o *locus* da Amelogenina, localizado nos cromossomos Y e X, que permite a determinação do sexo, (Figura 06) BUTLER, 2001. Todos os STRs que compõem o “CODIS” são sequências repetitivas tetraméricas. Entre estes marcadores os mais polimórficos são FGA, D18S51 E D21S11. O TPOX foi o que demonstrou menor variação (BUTLER, 2005; LEITE *et al.*, 2005).

Inicialmente, o banco de dados “CODIS” era constituído de uma base de dados de alelos VNTR, gerada a partir da tipagem de amostras biológicas de criminosos condenados por homicídio e estupro. Em 1998, quando a comunidade científica forense comemorava, juntamente com o FBI, 10 anos de introdução da técnica de tipagem por DNA em investigações criminais, foi lançado nos EUA o banco de dados norte-americano, para identificação de criminosos. O CODIS deu ao FBI a autoridade de estabelecer um banco de dados em nível nacional para fins de investigação criminal. O projeto envolveu 22 laboratórios que avaliaram 17 candidatos a marcadores, disponíveis em kits preliminares ou comerciais pelas empresas *Promega Corporation* (Madison WI), e *Applied Biosystems* (Foster City, CA). O CODIS interliga informações geradas por unidades laboratoriais de todos os Estados dos USA proporcionando a comparação de perfis genéticos em nível local, Estadual e Federal (BUDOWLE *et al.*, 1998).

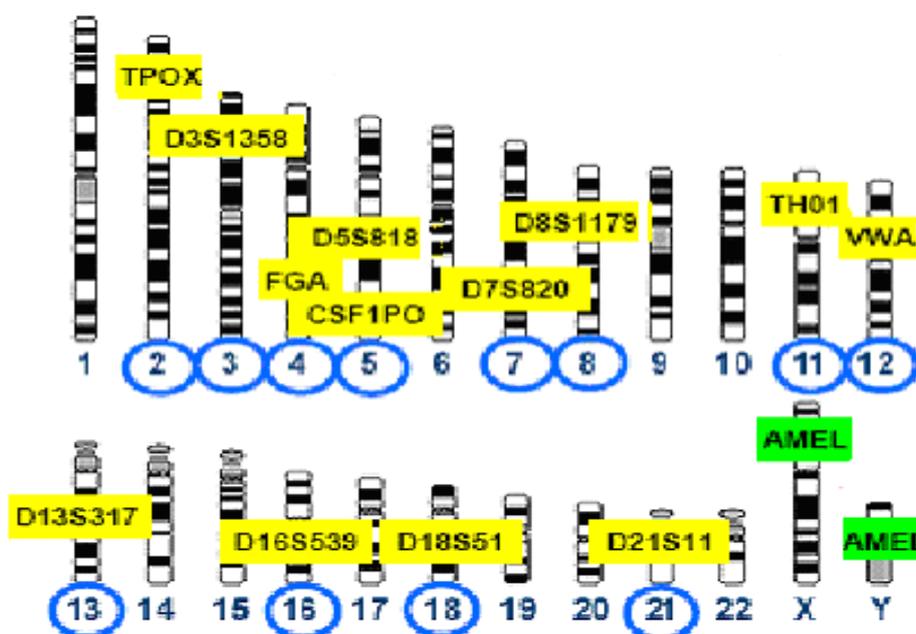


Figura 06: Os 13 marcadores STRs definidos pelo FBI para compor o CODIS, suas posições nos cromossomos e amelogenina nos cromossomos sexual (1997). Fonte: <http://www.fbi.gov/hq/lab/codis/index1.htm>

2.5.1 Normatização das análises de identificação humana

É importante ressaltar que, paralelamente à implantação e ao avanço das técnicas de identificação humana por DNA nas últimas décadas, foram criados diversos grupos técnicos internacionais, com a finalidade de assegurar a qualidade tanto das metodologias nascentes quanto da aplicabilidade das mesmas. Neste contexto, podem ser citadas as recomendações da Associação Americana de Bancos de Sangue - AABB, do inglês *American Association of Blood Banks*, criada nos Estados Unidos em 1978; o TWGDAM - *Technical Working Group on DNA Analysis Methods*, criado pela Polícia Federal Americana - FBI, do inglês *Federal Bureau of Investigation* – EUA, em 1988, para avaliar laboratórios forenses da América do Norte; a Sociedade Internacional para Genética Forense - ISFG, do inglês *International Society for Forensic Genetics*, criada em 1980, na Europa; o Grupo Espanhol e Português (GEP) da ISFG; a Academia Internacional de Criminalística e Estudos Forenses (AICEF),

além do Grupo Ibero Americano de Trabalho em Análise de DNA (GITAD) criado em 1998, em Belo Horizonte (LORENTE, 1999).

No Brasil, com iniciativa do Ministério da Justiça junto à Secretaria Nacional de Segurança Pública – SENASP elaborou-se uma proposta para um Programa de Ciência e Tecnologia na Segurança Pública, em 2003. Neste contexto estava a criação de uma Rede Nacional de DNA Forense com objetivo de montar um laboratório em cada região brasileira, promover a formação, capacitação dos profissionais na área periciais e padronização de exame de DNA, em perícias criminais no Brasil, tendo sido concluído em 2007. Em Manaus o grupo de pesquisadores da UFAM, apoiou a implantação do Laboratório de Genética Forense do Estado do Amazonas, atuou na formação nível de Especialização em Genética Forense dos Peritos Criminais da Polícia Civil.

2.5.2 Parâmetros de interesse forense

A *American Association of Blood Banks - AABB* (2000) desenvolveu normatização para Laboratórios nos testes de identificação humana, que consta de manual específico, onde são exigidos cálculos dos seguintes parâmetros:

- **Probabilidade de Coincidência (Pc)** - é a probabilidade de que duas pessoas escolhidas ao acaso tenham um dado genótipo, consiste do quadrado de sua frequência na população. Assim para um genótipo homozigoto é igual a $(p_i^2)^2$ e para um genótipo heterozigoto é igual a $(2 p_i p_j)^2$. A probabilidade de que duas pessoas tomadas ao acaso, tenham o mesmo genótipo não especificado é a soma dos quadrados das frequências de todos os genótipos possíveis.

- **Poder de Discriminação (Pd)** - é a probabilidade que teriam dois indivíduos, de uma mesma população, escolhidas ao acaso, de não apresentarem genótipos coincidentes em uma série de *loci* analisados. Este parâmetro é uma medida relativa da eficácia dos sistemas

de tipagem utilizados e dependerá da frequência dos genótipos na população e do polimorfismo do marcador genético. A probabilidade de discriminação pode ser calculada segundo Fisher, 1951, pela seguinte fórmula:

$$PD = 1 - \sum (\theta)^2$$

Sendo que θ é a frequência genotípica esperada. Para calcular o poder de discriminação total para a população estudada multiplicam-se os PD individuais.

- **Conteúdo de informação do polimorfismo (PIC)** - Mede o grau de polimorfismo de um marcador de DNA, usado principalmente no contexto de mapeamento gênico e diagnóstico genético. Depende da frequência dos alelos (BOTSTEIN *et al.*, 1980). É calculado como:

$$PIC = 1 - \left(\sum_{i=1}^n P_i^2 \right) - \sum_{i=1}^{n-1} \sum_{j=i+1}^n 2p_i^2 p_j^2$$

Onde: p = frequência alélica

$\sum P_i^2$ = Se refere ao genótipo homozigoto

$\sum 2p_i^2 p_j^2$ = Se refere ao genótipo heterozigoto

2.5.3 Parâmetros de Paternidade

As mais importantes vantagens na utilização da análise do DNA na determinação de paternidade em relação aos métodos clássicos, como tipagem sanguínea ou os marcadores de histocompatibilidade (HLA), estão no seu alto grau de certeza na inclusão de paternidade. Enquanto os métodos tradicionais podem estimar a inclusão com uma probabilidade entre 80 a 96 %, a análise do DNA permite estabelecer a paternidade com uma probabilidade maior ou

igual a 99,9999 %. O exame de DNA permite uma resposta definitiva que exclui a suposta paternidade ou comprova a mesma (inclusão de paternidade). Isto é possível porque cada pessoa possui um conjunto único de informações genéticas herdadas de seus pais biológicos. De acordo com a AABB (*American Association of Blood Banks*) (1989), deve-se considerar na análise do resultado de investigação de paternidade três principais parâmetros que são calculados com base nas frequências dos alelos que são: Probabilidade de Exclusão; Probabilidade de Inclusão; Índice de Paternidade.

- **Poder de Exclusão (PE)** - é o percentual de indivíduos cuja identidade ou paternidade biológica seria excluída com base nos marcadores estudados. Para calcular esse parâmetro, é considerada a ocorrência de incompatibilidade genética (CHAKRABORTY *et al*, 1974).

Este cálculo indica a probabilidade de não encontrarmos outro indivíduo, selecionado ao acaso, na população, que possua o mesmo perfil genético do pai biológico da criança.

$$\text{Fórmula: PE} = 1 - (a^2 + 2ab)$$

a = frequência do alelo paterno obrigatório herdado pela criança do pai.

b = frequência do alelo da criança

- **Probabilidade de Inclusão (PI)** - é igual à frequência dos alelos de todos os homens na população que coincide com o alelo paterno obrigatório da criança. Nos casos de inclusão de paternidade, a probabilidade de uma dada combinação de alelos ter ocorrido ao acaso seja menor que 0,00001, nos dando assim uma certeza maior que 99,99 %.

$$\text{Fórmula: PI} = (a^2 + 2ab)$$

- **Índice de Paternidade (IP)** – é analisado por meio da razão entre a probabilidade das bandas alélicas presentes ao perfil genético da criança e que estejam em coincidência com o perfil do suposto pai, e a probabilidade delas serem oriundas de qualquer indivíduo na população. Seu cálculo é baseado nos perfis genéticos apresentados pela mãe, criança e suposto pai. Nesse caso uma estimativa quantitativa da evidência em favor da paternidade por DNA do suposto pai, pode e deve ser calculada matematicamente utilizando a fórmula:

$$\mathbf{IP = X/Y}$$

onde **X** representa a probabilidade de que o homem analisado (suposto pai) realmente tenha contribuído com o alelo paterno obrigatório e **Y** representa a probabilidade de que outro homem qualquer tenha contribuído com o alelo paterno obrigatório, sendo que o valor de X e Y varia de acordo com a configuração dos genótipos da mãe, filho e do suposto pai e é determinado pelas leis de Mendel.

Num caso em que o suposto pai seja homozigoto para o alelo paterno obrigatório, ou tenha ambos os alelos obrigatórios, a probabilidade de que ele transmita o alelo paterno obrigatório é de 1 (2/2). Caso o suposto pai tenha somente uma cópia do alelo paterno obrigatório, ou somente um dos alelos paternos obrigatórios, a probabilidade de transmissão será de 0,5 (1/2). Assim X poderá ser 1 ou 0,5. O cálculo do índice de paternidade (IP) deve ser analisado para cada *locus*.

O IP pode ser convertido para uma probabilidade posterior de paternidade pelo Teorema de Bayes (LI & CHAKRAVARTI, 1988). Entretanto, o teorema nos força a ter de escolher uma probabilidade apriori, isto é, um valor subjetivo da chance de paternidade anterior ao teste, dado essencial para o cálculo de acordo com as regras do teorema. Probabilidades apriori constantes (0,5) têm sido consensualmente utilizadas em todo o mundo. Os resultados dos cálculos de probabilidade de paternidade e probabilidade de exclusão são normalmente expressos em valores percentuais, geralmente com índices próximos a 99,99%. A Probabilidade de Paternidade é calculada de acordo com o Teorema de

Bayes, a partir do Índice de Paternidade, e representa a probabilidade de o suposto pai ser o pai biológico. Na Tabela 3, encontram-se as fórmulas utilizadas para cálculo do IP.

Tabela 3: Fórmulas utilizadas para cálculos do índice de paternidade. As letras A,B,C,D representam os alelos do perfil genético da mãe, criança e suposto pai em cada *locus* analisado, onde (A = a) e (B = b).

| ALELOS | | | FÓRMULAS | |
|---------|---------|-------------|-----------------------|--------------------|
| Mãe | Criança | Suposto Pai | Índice de Paternidade | Índice de Exclusão |
| A | A | A | $1/a$ | $1-a(2-a)$ |
| A | A | AB | $1/2a$ | $1-a(2-a)$ |
| AB | A | A | $1/a$ | $1-a(2-a)$ |
| AB | AB | A | $1/(a+b)$ | $1-(a+b)(2-a-b)$ |
| AB | AB | AB | $1/(a+b)$ | $1-(a+b)(2-a-b)$ |
| AB | AB | AC | $1/[2(a+b)]$ | $1-(a+b)(2-a-b)$ |
| AB | A | AB | $1/2a$ | $1-a(2-a)$ |
| B | AB | A | $1/a$ | $1-a(2-a)$ |
| B | AB | AB | $1/2a$ | $1-a(2-a)$ |
| B | AB | AC | $1/2a$ | $1-a(2-a)$ |
| BD | AB | AC | $1/2a$ | $1-a(2-a)$ |
| BC | AB | AC | $1/2a$ | $1-a(2-a)$ |
| BC | AB | AB | $1/2a$ | $1-a(2-a)$ |
| BC | AB | A | $1/a$ | $1-a(2-a)$ |
| Ausente | AB | AC | $1/4a$ | $1-(a+b)(2-a-b)$ |
| Ausente | AB | AB | $(a+b)/4ab$ | $1-(a+b)(2-a-b)$ |
| Ausente | AB | A | $1/2a$ | $1-(a+b)(2-a-b)$ |
| Ausente | A | AC | $1/2a$ | $1-a(2-a)$ |
| Ausente | A | A | $1/a$ | $1-a(2-a)$ |

* Os alelos utilizados para o cálculo do índice de paternidade são os alelos paterno-obrigatórios. Quando não se pode determinar um alelo paterno-obrigatório, caso em que mãe e criança são heterozigotos e iguais, o cálculo é feito considerando-se a frequência dos dois alelos da criança.

No caso da mãe ausente as letras A,B,C representam os alelos do perfil genético da criança e suposto pai em cada *locus* analisado. O mesmo cálculo pode ser feito considerando-se o pai ausente (para comprovação de maternidade).

- **Índice de Paternidade Combinado (IPC)** - é o produto dos diferentes índices de paternidades para cada *locus* individualmente (Hartl &Clarck, 1997). O IPC varia de 0 a infinito, um IPC menor que 1 significa que o perfil genético é mais consistente com uma não paternidade que a paternidade. O IPC maior que um significa que o perfil genético suporta a hipótese de que o homem testado é o pai biológico, expressando uma probabilidade de 99,9999 % (AABB, 1989).

3 OBJETIVOS

3.1 OBJETIVO GERAL

Desenvolver um novo sistema multiplex de marcadores microssatélites (STRs) para análise genética em humanos.

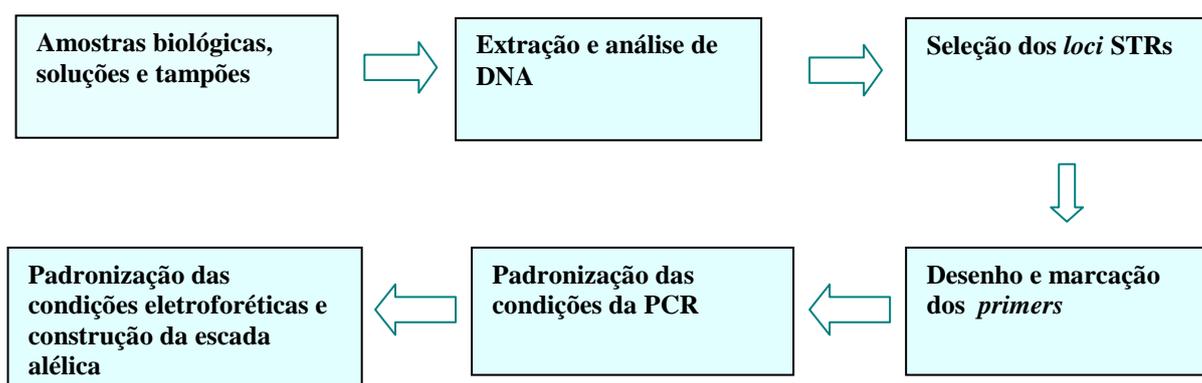
3.2 OBJETIVOS ESPECÍFICOS

- Selecionar por meio de ferramentas da bioinformática cinco marcadores STRs autossômicos polimórficos.
- Desenvolver escadas alélicas para os cinco *loci* STRs.
- Construir bancos de dados de frequências alélicas para os *loci* STR a partir de amostras da população do Estado do Amazonas e comparar com as populações do Estado do Rio de Janeiro e a do Estado do Espírito Santo.
- Avaliar os parâmetros genéticos populacionais individualmente para cada *loci*, baseados na frequência dos alelos STR, e verificar se estão de acordo com o Equilíbrio de Hardy-Weinberg.
- Avaliar a capacidade do sistema para identificação humana e estabelecimento de vínculos genéticos.

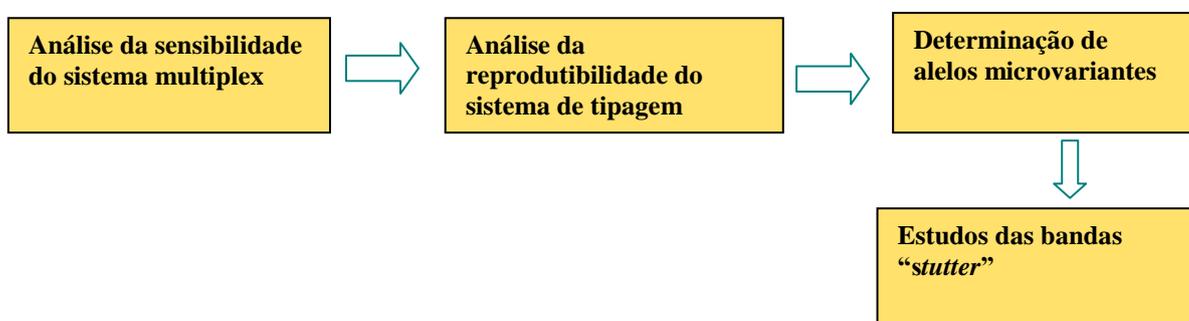
4 MATERIAL E MÉTODOS

Para a realização deste estudo foram seguidos os esquemas abaixo apresentados que indicam de forma sucinta as etapas cumpridas no desenvolvimento deste trabalho.

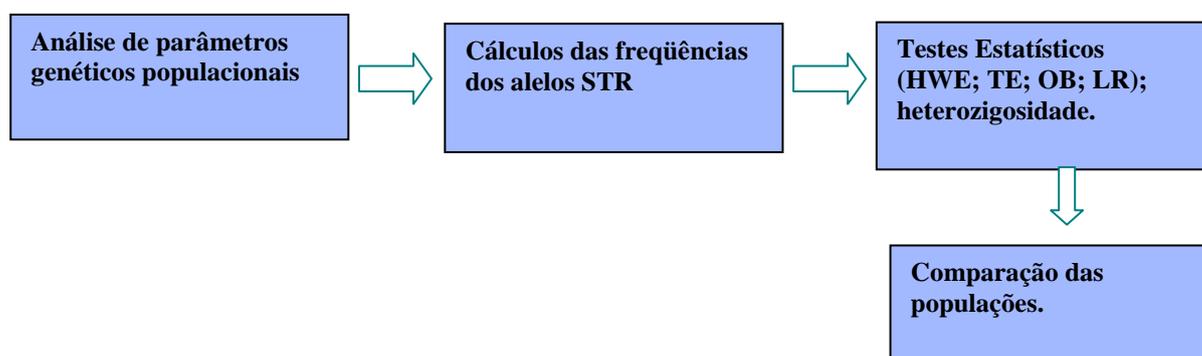
a) DESENVOLVIMENTO DO SISTEMA MULTIPLEX



b) VALIDAÇÃO DO SISTEMA



c) CONSTRUÇÃO E ANÁLISE DE BANCO DE DADOS



4.1 DESENVOLVIMENTO DO SISTEMA MULTIPLEX

4.1.1 Amostras biológicas

O projeto de pesquisa associado à parte experimental da tese de doutorado foi aprovado pelo Comitê de Ética em Pesquisa - CEP da Universidade Federal do Amazonas, sob o processo CAAE nº 0038/07. No presente estudo foram avaliadas 104 amostras biológicas (sangue) coletadas de homens e mulheres, geneticamente não relacionados, com idades iguais ou superiores a 18 anos, nascidos no Estado do Amazonas. Aos doadores foram entregues o Termo de Consentimento Livre e Esclarecimento (TCLE), no qual cada indivíduo expressou sua vontade em participar deste projeto, e também de preencher um questionário, de acordo com a Resolução CNS 196/96.

4.1.1.1 Coleta das amostras

Amostras com cerca de 1mL de sangue foram coletadas dos doadores, utilizando-se seringas e agulhas descartáveis, por meio de sistema vacuntainer contendo EDTA. Após a coleta, os tubos foram codificados e armazenados a 4 °C até a extração dos DNAs. A coleta do material biológico e a extração do DNA foram realizadas no Laboratório de Diagnóstico Molecular do Centro de Apoio Multidisciplinar da UFAM.

Outras amostras de DNA utilizadas neste trabalho fazem parte do acervo do Laboratório de Diagnósticos por DNA, do Instituto de Biologia Roberto Alcântara Gomes da Universidade Estadual do Estado do Rio de Janeiro, que foram gentilmente doadas pelo Coordenador do citado laboratório, sendo 112 amostras oriundas da população do Rio de Janeiro e 92 amostras da população do Espírito Santo. Tais amostras foram utilizadas em estudos relacionados com a validação do sistema multiplex, em desenvolvimento, reprodutibilidade dos resultados, construção de bancos de dados de frequências alélicas e comparações entre os parâmetros genéticos observados para as três populações.

Ainda foram utilizadas amostras DNA, preparadas de saliva, urina, bulbo capilar, sêmen e sangue, de um indivíduo saudável com a finalidade de confirmar a fidelidade do sistema multiplex na geração do perfil alélico a partir de diferentes fontes de material genético de um mesmo doador.

As amostras da mucosa oral (saliva) e sêmen foram coletados com *swabes* e mantidos a -20 °C até a extração do DNA. Para a produção de DNA 1 mL de urina foi coletado e transferido para um tubo de 1,5 mL, centrifugado, por 5 min, a 12000 rpm (centrífuga Centrimicro modelo 212, rotor). Ao sedimento foram adicionados 500 µL de tampão PBS, sendo a suspensão centrifugada nas mesmas condições. O procedimento foi repetido duas vezes, sendo o sedimento final, armazenado a -20 °C até a extração do DNA.

4.1.2 Extração e análise de DNA

As preparações de DNA referentes à população do Amazonas foram processadas no Laboratório de Diagnóstico Molecular da UFAM. O procedimento adotado na extração de DNA foi o mesmo descrito no *Kit Wizard Genomic DNA Purification*.

4.1.2.1 Quantificação do DNA

A concentração de DNA extraído de sangue da população do Amazonas era determinada por espectrofotometria ou/ estimada em gel de agarose 0,8 %, tomando-se por base o DNA K562, na concentração estoque de 10 ng/µL, para fazer a comparação.

- Por espectrofotometria

Para quantificação, o DNA foi diluído com água deionizada (1:500) e determinadas as absorbâncias a 260 nm e 280 nm determinadas utilizando-se um espectrofotômetro modelo BioMate 3 (*Thermo Electron Corporation*), usando uma cubeta de quartzo. A solução de DNA foi considerada pura quando a relação ABS_{260} / A_{280} foi igual ou superior de 1,8, sendo então a sua concentração determinada, considerando-se que 1U ABS_{260} corresponde a 50 µg de DNA (SAMBROOK *et al.*, 1982).

- Por eletroforese em gel de agarose

A análise semiquantitativa e de integridade do DNA foi realizadas por meio de eletroforese em gel de agarose a 0,8 %, em tampão TEB 1X. No início da corrida eram aplicados 80 volts até a entrada das amostras no gel. A seguir, a voltagem foi elevada para 100 volts até o término da eletroforese, cerca de 1h (SAMBROOK *et al.*, 1982). Para estimar a concentração do DNA, também foram aplicadas no gel amostras de DNA de concentrações conhecidas para fazer a comparação o gel era corado com brometo de etídeo (0,5 µg/mL), e visualizado no transiluminador de luz ultravioleta e fotografado no sistema IMAGE-MASTER da Amersham-Pharmacia Biotech.

Para verificar a eficiência da amplificação por PCR, foi realizada a detecção do DNA por eletroforese em gel de agarose a 2,5%, em tampão TEB 1X. Para isso amostras constituídas por 4µL do produto da PCR, 3µL de tampão de carregamento (TEB 5X), 3µL de água deionizada estéril (0,5µg/mL), sendo os produtos da PCR, visualizados em transiluminador de luz ultravioleta e fotografados.

4.1.3 Seleção dos *loci* STRs

Os marcadores STRs selecionados foram desenvolvidos por Pontes (2003), e identificadas diretamente no GeneBank, acessando por meio da página eletrônica da NCBI <http://www.ncbi.nlm.nih.gov>. – (AUSTCHUL *et al.*, 1990 e 1997).

Segundo Pontes, foram realizadas buscas de sequências pentanucleotídicas tais como (AGAT, GATA, TCTG, CAGA e GGCT), no banco de genes humano. Após a identificação dessas sequências microssatélite foram desenhados os iniciadores (*primers*) através da página eletrônica [http://www.wi.mit.edu/cgi-bin/primer/primer³_www.cgi](http://www.wi.mit.edu/cgi-bin/primer/primer3_www.cgi).

4.1.4 Desenho e marcação dos *primers*

No desenho dos iniciadores ou *primers* optou-se por três variações de tamanho dos amplicons, com o objetivo de elaborar um sistema multiplex que possibilitasse a amplificação simultânea de maior número de *loci* e detecção dos produtos da PCR tanto por eletroforese em capilar quanto em gel de poliacrilamida. Os iniciadores foram então denominados conforme o tamanho das sequências que amplificam, como: tamanho pequeno (P) variando em torno de 150-170 pb; o médio (M) variando em torno de 200-220 pb e o grande (G) em torno de 260-280 pb. Procurou-se desenhar os iniciadores com T_m entre 55 °C e 62 °C. Uma vez definidos, os iniciadores foram então sintetizados pela firma *Life Technologies* (Carlsbad, Califórnia).

Estudos preliminares foram feitos analisando-se os produtos de amplificação dos *loci* STR, por eletroforese em gel de agarose (baixa resolução), em seguida, por eletroforese em gel de poliacrilamida (média-alta resolução) e, posteriormente, por sistema de eletroforese em gel de poliacrilamida em sistema capilar automatizado (Equipamento MEGABace de seqüenciamento de DNA). Para análise automatizada, requereu-se a marcação dos iniciadores com fluoróforos, o que foi realizado pela empresa (Alfa-DNA), sendo marcado um iniciador de cada par (*Forward*), para facilitar a análise simultânea dos *loci* STR. As seqüências iniciadoras foram marcadas com os fluoróforos *Hexametildamina* (HEX), *Carboxi-*

fluoresceína (FAM) e *Dye yellow* (NED). As primeiras análises dos produtos amplificados foram realizadas no equipamento automático (MegaBace), que possibilita a identificação da origem dos alelos amplificados no gel de poliacrilamida em capilar, tenham eles uma mesma mobilidade eletroforética, ou não, uma vez que os produtos de cada um dos diferentes *loci* STR são diferenciados por apresentarem-se marcados com distintos fluoróforos que, portanto, emitindo luz em diferentes comprimentos de ondas, proporcionam uma perfeita discriminação dos produtos da PCR por cada iniciador condicionando uma visualização.

4.1.4.1 Nomenclatura do Sistema

Com o objetivo de padronizar a nomenclatura dos *loci* STRs e facilitar a comparação de dados entre laboratórios, o DNA, a Comissão da Sociedade Internacional de Medicina Legal *Haemogenetics* (ISFH) estabeleceu duas recomendações para a designação dos alelos STRs (BAR *et al.*, 1994 e 1997). Basicamente, uma seqüência repetitiva é denominada pela estrutura e número de sua unidade de repetição. Para a designação de cada marcador é utilizado o nome do gene do qual este marcador faz parte. Algumas vezes, utiliza-se o prefixo HUM para indicar que este *locus* pertence ao genoma humano. Por exemplo, o *locus* STR TH01 é proveniente do gene *tyrosine hydroxylase*.

Os marcadores que não fazem parte de uma região gênica são denominados pela sua posição no cromossomo. De modo geral, um locus de microsatélite é representado pela letra D, o cromossomo onde está o microsatélite, pela letra S e por um número que diz respeito à ordem de descoberta deste microsatélite. Por exemplo o *locus* STR D5S818, não está localizado dentro de uma região gênica. A letra D refere-se ao DNA. O número 5 refere-se ao número do cromossomo e a letra S refere-se ao fato de que este marcador é uma seqüência única (*single*). Os números finais indicam a ordem em que estes marcadores foram descobertos, considerando-se cada cromossomo em particular (BUTLER, 2006).

Em observação às normas de nomenclatura, o novo sistema desenvolvido neste trabalho, foram denominadas PentaPlex-ISED. Em decorrência de ser constituído por cinco marcadores microssatélites STR, utilizou-se o prefixo penta. As letras I, S, E e D representam as iniciais dos nomes dos pesquisadores que desenvolveram o sistema multiplex são: Isabel da Mota Pontes (I), Bióloga (UFAM); Spartaco Astolfi Filho (S), biólogo (UFAM); Elizeu Fagundes de Carvalho (E), biólogo (UERJ) e Dayse Aparecida da Silva (D), bióloga (UERJ).

O ajuste, montagem e a validação do sistema PentaPlex-ISED foram realizados nos Laboratórios de Diagnósticos Molecular da Universidade Federal do Amazonas - UFAM e no Laboratório de Diagnósticos por DNA da Universidade do Estado do Rio Janeiro - UERJ.

4.1.5 Definição das condições da PCR do sistema multiplex

Para definição das condições da PCR, cada componente não marcado dos diferentes pares de iniciadores foi dissolvido em Tris-HCl 2mM pH 8,0, para a concentração final de 100 picomoles/ μ L, solução estoque. Para uso na reação de PCR, cada primer foi diluído 1/20 em H₂O (PCR grande) ficando assim na concentração de trabalho, ou seja, 5 picomoles/ μ L.

Nos procedimentos para a reação da PCR foram utilizadas amostras de DNA de indivíduos que participam deste estudo e mais o DNA padrão K562. Em uma primeira etapa, as reações de amplificação foram realizadas nos termocicladores *Perkin Elmer 480* e *Mastercycler Gradiente da Eppendorf*, utilizando um mesmo programa de amplificação.

Na padronização das reações monoplexes foram usados diferentes enzimas *Taq* Polimerases: *Easy Taq* Polimerase da LAB TRADE do Brasil, *Platinum Taq* DNA Polimerase *High Fidelity* da INVITROGEN e *Go Tag Flex* da PROMEGA, utilizando-se o mesmo programa de amplificação.

Em relação aos componentes marcados dos diferentes pares de iniciadores as reações de amplificação por PCR foram ajustadas a partir da misturas dos diferentes pares de iniciadores nas proporções necessárias visando:

- a maior intensidade de fluorescência dos produtos na amplificação;
- o equilíbrio dos primers na amplificação;
- a T_m de cada *loci* STR.

Nesta etapa para amplificação com o sistema PentaPlex-ISED, foi aumentada a T_m de 60 °C para 62 °C, visando uma melhor definição dos produtos de amplificação de cada *locus* (alelos STR) e a minimização de produtos inespecíficos. Também foi aumentado o tempo da extensão de (60 °C p/10 min) para (60 °C p/ 60 min) para possibilitar a adenilação adequadas dos nucleotídeos. Nesta etapa, utilizou-se a enzima *Platinum* Taq DNA Polimerase (INVITROGEN).

4.1.6 Definição das condições de amplificação via PCR

Após a padronização dos sistemas de reação da PCR para cada *locus* separadamente (monoplex), uma das etapas importantes e desafiadoras desse processo foi a definição de um sistema multiplex que proporcionasse amplificação equilibrada de todos os *loci* STR. Em uma primeira etapa, foram otimizadas as reações de PCR, alterando-se as concentrações dos componentes: $MgCl_2$, dNTPs, Taq. DNA Polimerase e Tampão 10 X. A montagem do PentaPlex-ISED se deu com base nas informações a cerca das reações monoplexes, que permitiu então a mistura dos cinco iniciadores marcados em condições de possibilitar a amplificação simultânea, qualitativa e quantitativamente equilibrada dos cinco diferentes *loci*.

4.1.6.1 Construção da escada alélica

No contexto deste trabalho, uma escada alélica respectiva de um *locus* hipervariável deve ter o conjunto de alelos, ou os alelos majoritários, presentes naquele local genético, objetivando a tipagem fidedigna e conseqüente obtenção de perfis genéticos de indivíduos oriundos de diferentes populações, os quais poderão ser comparados com a finalidade de identificação humana.

Para a construção da escada alélica do sistema PentaPlex-ISED, foram realizadas amplificações a partir de amostras de DNA com perfis genéticos conhecidos. Para isso as amostras codificadas como 05.6, 11.7, 23.2, 40.6, 11.7, 58.2, 77.2, 65.2, 40.7, 20.7, foram individualmente amplificadas por PCR (reações monoplexes), utilizando-se iniciadores não marcados. Na preparação das reações de PCR para a construção da escada alélica, foram utilizadas quantidades de cada iniciador de acordo com as intensidades verificadas em gel de agarose 2,5 % corado com brometo de etídio

Os produtos da PCR foram analisados em gel PAGE a 6%, corado com nitrato de prata, em seguida aplicando no gel 2,5 μ L de tampão formamida + 2,5 μ L do Mix. Conforme procedimentos a seguir.

4.1.6.2 Eletroforese em gel de poliacrilamida (PAGE)

Os produtos amplificados a partir de cada um microsatélite foram analisados por PAGE, seguindo os procedimentos abaixo.

Nos procedimentos para a preparação das placas, seguiram-se rigorosamente as seguintes instruções: As placas de vidro de dimensões: 19,5cm x 33,5 cm e 19,5 cm x 36 cm, foram lavadas com detergente neutro e enxaguadas com água destilada. Após a lavagem, a placa maior foi tratada com 1mL da solução *REPEL*, espalhando-se de forma homogênea, com auxílio de papel toalha. Após 5 minutos do tratamento a placa foi lavada 3 vezes com

água destilada, para retirar o excesso da solução. A placa menor foi tratada com 1mL da solução *BIND*, esfregando-se rapidamente com papel toalha de forma homogênea; após 5 minutos foram lavadas três vezes com etanol absoluto secando-se bastante com papel toalha. Para retirar o excesso da solução. Para retirar o excesso da solução, procurou-se ter cuidado de não deixar as placas próximas uma das outra durante a preparação. Após as placas estarem secas, foram montadas em forma de sanduíche, colocando os espaçadores (0,4 mm de espessura) nas bordas laterais, em seguida foram colocados os prendedores para evitar vazamento no momento de verter o gel.

4.1.6.2.1 *Preparo do gel de poliacrilamida*

Na preparação do gel de poliacrilamida, os procedimentos foram os seguintes: misturou-se imediatamente antes 60mL da solução do gel de poliacrilamida 4%, com 40 μ L de TEMED e 400mL de persulfato de amônio (APS) a 10%. Após o preparo da solução de poliacrilamida, a mesma foi vertida lentamente entre as placas, evitando a formação de bolhas. Em seguida, foi inserido o pente (tipo dentes de tubarão), deixando-se o gel polimerizar por aproximadamente 1 hora. Após a polimerização do gel, as placas foram montadas na cuba de eletroforese e processou-se a pré-corrida a 40 watts, durante 30 minutos, para o pré-aquecimento das placas e estabilização do sistema em torno 40 - 50 °C.

4.1.6.2.2 *Preparo das amostras (PCR) para eletroforese*

Os procedimentos para o preparo das amostras contendo os produtos da PCR foram os seguintes: utilizou-se 2,5 μ L do sistema de reação mais 2,5 μ L de tampão de carregamento contendo formamida. De maneira similar, a escada alélica e o marcador de tamanho (*ladder* 100 bp) foram preparados. Anteriormente à aplicação nos géis, as amostras eram desnaturadas a 92 °C por 2 minutos, e transferida imediatamente para o banho de gelo a (4° C). De cada

amostra preparada a partir dos produtos de amplificação por PCR, eram aplicados 3,5 µL nos géis de poliacrilamida. A eletroforese era processada a 40 W (voltagem inicial 1.700 V, corrente inicial 27 mA), por uma hora e trinta minutos. Após o encerramento da corrida, as placas eram separadas, sendo então o gel fixado a placa *BIND* corado com nitrato de prata.

4.1.6.2.3 Coloração do PAGE com nitrato de prata

A coloração dos géis de poliacrilamida era realizada com nitrato de prata conforme descrito por CRESTE *et al*, (2001). Após a eletroforese, as placas eram cuidadosamente separadas e o gel aderido à placa menor fixado com 2 L de solução de fixação (etanol 10 % e ácido acético 1 %) e mantido sob lenta agitação, durante 10 minutos. Após a fixação, os géis foram lavados sob agitação lenta com 2L de água milli-Q, durante 1 minuto e, em seguida, submetido à solução de oxidação (ácido nítrico 1,5 %), durante 3 minutos. Então novamente lavado sob agitação lenta com 2 L de água milli-Q, durante 1 minuto e imerso em 2L de solução 0,2 % de AgNO₃ durante 20min sob agitação lenta. Em seguida, foram realizadas duas lavagens de 30seg. com água milli-Q, sob agitação lenta, e incubados em 1L de solução de revelação (Na₂CO₃ 3 % e formaldeído 0,02 %), sob agitação lenta e constante. Quando os produtos da PCR começavam ser visualizados, os géis eram transferidos para igual volume de solução de revelação fresca, agitando-se constantemente. Após obter-se o padrão de revelação desejado, os géis foram imersos em 2L de solução bloqueadora (ácido acético glacial 5 %) e mantidos por 5 minutos sob agitação lenta. Em seguida, os géis foram submetidos a uma lavagem final pela imersão em 2L de água ultrapura, sob agitação, durante 1min. Os géis foram mantidos a temperatura ambiente até que ficassem completamente secos e as imagens dos mesmos foram capturadas utilizando um *scanner*.

4.1.6.3 Condições eletroforéticas em sistema capilar

Para eletroforese em sistema capilar, foram utilizados dois equipamentos de sequenciamento automáticos de DNA. Os produtos das ampliações foram inicialmente separadas detectadas no sistema MegaBace. Nesta etapa, os produtos de PCR monoplexes com os iniciadores marcados com fluoróforos eram analisados individualmente. As amostras para as eletroforeses em capilares foram preparadas conforme descrito no protocolo.

Para a análise de produtos da PCR nesse sistema de eletroforeses, os capilares devem ser preenchidos com polímeros de poliacrilamida, sendo a injeção das amostras realizada eletrocineticamente. Com a aplicação de uma diferença de potencial nas extremidades dos capilares, ocorre somente a entrada de moléculas carregadas. A detecção dos fragmentos ocorre quando da passagem destes pela janela de detecção onde, após excitação por laser, os diferentes fluoróforos têm suas emissões captadas e convertidas em sinal elétrico, possibilitando assim a detecção dos fragmentos de DNA produzidos pela PCR.

A identificação dos alelos foi feita por meio da comparação entre os fragmentos de DNA gerados por PCR a partir da amostra a ser genotipada e da escada alélica produzida conforme anteriormente descrito, verificando-se ainda o tamanho em pares de bases de cada produto detectado, os alelos STR.

- Preparo das amostras para a leitura no sistema capilar

Após a PCR, alíquotas dos produtos da PCR foram analisadas no sistema capilar de acordo com seguindo os procedimentos: - preparação das amostras de carregamento contendo 8,8µL de Formamida + 0,2µL de LIZ (Size standard); - após o preparo do mix, foram distribuídos 8µL na microplaca, e em seguida foi adicionado 1 - 2µL do produto da PCR em cada poço; - a seguir foi realizada a desnaturação a 95 ° C, durante 3min, e colocando-se imediatamente a placa sobre o gelo, deixando-a pronta para ser injetada; - a injeção das

amostras nos capilares foi realizada a 3KV por 60-80 segundos e a corrida foi de 10 KV, por 80min. As interpretações dos resultados foram realizadas por meio dos programas computacionais utilizando *software* aplicado para a identificação humana.

4.2 VALIDAÇÃO DO SISTEMA PENTAPLEX-ISED

No processo de validação de um sistema multiplex, uma série de etapas tais como estudos a cerca da sensibilidade e reprodutibilidade do método, análise de amostras biológicas de diferentes fontes e de especificidade, sensibilidade do equipamento utilizado na detecção dos STR devem ser realizados visando à aplicação em identificação humana por DNA.

Nos estudos de validação do multiplex PentaPlex-ISED, foram realizados experimentos a partir de DNA extraídos de amostras de sangue e dos DNAs padrões (K562, 9947 e 9948).

4.2.1 Análise da sensibilidade do sistema PentaPlex-ISED

O estudo da sensibilidade do método em amostras biológicas provenientes de sangue está relacionado com a determinação da quantidade mínima de DNA necessária para a tipagem dos alelos dos *loci* STR reunidos no PentaPlex-ISED. Para verificar a quantidade mínima de DNA que se relacionava com intensidade de fluorescência detectável de todos os alelos relativos aos *loci* STRs do sistema PentaPlex-ISED. Na definição da quantidade mínima de DNA a serem amplificadas foram preparadas reações de PCR utilizando o DNA padrão K562. Foram feitas várias diluições do produto amplificado (1:10, 1:20 e 1:40), em seguida realizada a tipagem.

Na detecção do DNA por eletroforese em géis de poliacrilamida corado com nitrato de prata foram testadas em várias concentrações de DNA entre 0,5 – 1ng.

4.2.2 Análise da reprodutibilidade do sistema de tipagem

No estudo da reprodutibilidade foram realizadas algumas etapas importantes como: a reprodutibilidade das tipagens dos 5 *loci* STR, nas amostras de sangue da população do Amazonas – AM, na população do Rio de Janeiro e a população do Espírito Santo, usando também os DNAs padrões (9948, 9947 e K562).

Na análise da reprodutibilidade das tipagens do multiplex PentaPlex-ISED foi avaliada com base em dois estudos:

- a reprodutibilidade das tipagens de alelos STR de amostras de sangue e DNA padrão;
- a reprodutibilidade de tipagem do perfil de DNA proveniente de um indivíduo testado as diferentes amostras biológicas como: sangue, urina, saliva, sêmem e bulbo capilar.

4.2.3 Determinação de alelos microvariantes

Como parte da validação do multiplex PentaPlex-ISED, foram realizados estudos nos alelos que apresentam microvariâncias. Esses alelos são raros nas populações, caracterizando-se por apresentarem frações da unidade de repetição que caracteriza o microssatélite.

Nesta etapa, alíquotas de DNA que apresentavam alelos microvariantes foram seqüenciadas para confirmação da estrutura do alelo.

O percentual de microvariância verificado em cada *locus* foi calculado e comparado com aqueles observados em outras populações.

4.2.3.1 Sequenciamento dos *loci* STRs

O seqüenciamento processou-se a partir dos STR amplificado por PCR, submetidos à purificação em colunas “QIAquick[®] PCR Purification Kit – QIAGEN”, conforme o seguinte procedimento: em um tubo de 600µL, adicionou 35µL do produto amplificado, acrescentando

175 μ L de solução de “binding” (5 vezes o volume da amostra). Colocou-se esta mistura na coluna de purificação e centrifugou-se por 50seg. O tampão que passou pela coluna, contido no tubo suporte, foi descartado após a centrifugação. Nesta mesma coluna, adicionou 500 μ L do tampão de lavagem e centrifugou-se por 50seg. Novamente descartou-se o volume centrifugado, contido no tubo suporte da coluna. Centrifugou-se por 50seg., a fim de retirar o excesso de álcool. Trocou-se o tubo suporte da coluna por um tubo limpo, com capacidade para 1,5mL e adicionou 30 μ L do tampão de eluição no centro da coluna, e deixou-se embeber por 2 minutos. Centrifugou-se por 2 minutos. O material que eluiu da coluna constituiu-se no material purificado desejado. Nestes procedimentos de purificação utilizou-se uma microcentrífuga tipo “*Eppendorf-Centrifuge 5410*” e as centrifugações processaram-se a 12.000 g. Do volume de 30 μ L eluído, retirou-se 3 μ L para a análise por meio de eletroforese, em gel de agarose.

A amostra foi preparada adicionando-se aos 3 μ L, 5 μ L de água e 2 μ L de tampão de corrida TEB 5X. Armazenou-se o restante para fins de determinação de seqüência de nucleotídeos da amostra.

4.2.3.2 Reação de Seqüenciamento

O método de seqüenciamento usado foi o descrito por SANGER (1977). Para o seqüenciamento utilizou-se o seguinte sistema de reação: 2,5 μ L – 4 μ L do DNA; 2,0 μ L do “pré-mix DYEnamic ET Dye Terminator Cycle Sequencing Kit for MegaBACE DNA Analysis Systems”; 1,0 μ L do iniciador (5pmol/ μ L), e água Milli-Q para completar o volume final de 10 μ L. A reação preparada foi levada ao aparelho termociclador PXE 0.2 Thermal Cycler-Electron Corporation, programado para realizar o seguinte termociclo da seguinte forma: 95°C, por 25seg. para a desnaturação inicial e 30 vezes o ciclo: (95 °C, por 15seg.; 50 °C, por 20seg.; 60°C, por 1min.).

4.2.3.3 Precipitação e leitura da reação de seqüenciamento

Ao produto da reação de seqüenciamento foi adicionado 1 μ L de acetato de amônia 7,5M e 27,5 μ L de etanol absoluto. A preparação foi misturada utilizando-se um agitador magnético (vortex) e incubada, por 20min., à temperatura ambiente. A placa foi envolvida em alumínio laminado para evitar a incidência de luz. Seguindo-se a etapa de incubação, a placa foi centrifugada a 4.000 por 40 minutos, a 4 °C, em centrífuga refrigerada (Eppendorf 5804R). O sobrenadante foi descartado. Foram adicionados 120 μ L de etanol a 70%, seguido de agitação branda por 15seg. A placa era então centrifugada novamente a 4000g., por 10min, e o sobrenadante descartado, seguindo-se nova centrifugação, com a placa invertida, a 700g, por 15seg. A placa foi mantida em fluxo laminar, por 15min., para que o etanol fosse evaporado completamente. O DNA foi ressuspenso em 10 μ L de *loading buffer* (formamida a 70 % e EDTA 1mM). A placa era vedada e agitada vigorosamente em vortex, por 2 min, anteriormente a uma centrifugação a 700g., por 15seg. Em seguida, as amostras eram analisadas no seqüenciador automático “MegaBace 1000 (Amersham Biociences)” por meio de eletroforese capilar em gel de poliacrilamida. Para injeção da amostra no capilar, utilizou-se 3 kV, por 80 segundos, sendo a eletroforese procedida a 6kV, por 5h, na temperatura de 44 °C.

4.2.4 Estudo das bandas *stutter*

Os estudos dos artefatos denominados de *stutter* foram particularmente realizados ao longo das análises de DNA nas três populações estudadas. As amostras de DNA foram amplificadas por PCR, no sistema multiplex PentaPlex-ISED, procedeu-se à separação por eletroforese em capilar ABI 3100 (Applied Biosystems), no momento em que foram detectados os *loci*, também foi observada a presença de bandas *stutter*. A intensidade óptica de cada banda *stutter* era comparada àquela do alelo com tamanho imediatamente superior à banda *stutter*, muitas vezes podem ser confundidas como um alelo real na hora da leitura.

4.3 CONSTRUÇÃO E ANÁLISE DO BANCO DE DADOS

As amostras populacionais estudadas reuniram 308 indivíduos de três populações brasileiras. Os 616 alelos detectados foram importados para o programa Arlequin 3.1 (Excoffier *et al.*, 2005), visando à obtenção das frequências alélicas de cada marcador e outros parâmetros genéticos de interesse, e usou-se o *software PoweStar* para analisar dados estatística de interesse forense.

4.3.1 Análise de parâmetros genéticos populacionais

Os parâmetros analisados para o estudo da variabilidade genética foram: número médio de alelos por *locus*, frequência dos alelos STR, heterozigosidade média esperada (H_e), heterozigosidade média observada (H_o).

1) Para calcular o número médio de alelos por *locus*, faz-se a soma do número total de alelos observados em todos os *loci* divididos pelo número total de alelos de cada locus individualmente.

2) A heterozigosidade média esperada: mede a proporção esperada de indivíduos heterozigotos em um *locus* (HARTL & CLARCK, 1997);

3) Os cálculos da heterozigosidade média observada: equivale a proporção observada de indivíduos heterozigotos em um dado *locus* (HARTL & CLARCK, 1997);

4) Frequência alélica: o cálculo das frequências alélicas é realizado por contagem direta.

A frequência de determinado alelo na população é o número de vezes que o alelo foi encontrado dividido pelo número total de alelos do *locus* STR específico analisado.

4.3.2 Análises estatísticas

Neste estudo foram analisados os seguintes parâmetros estatísticos populacionais: equilíbrio de Hardy-Weinberg (HWE), heterozigosidade observada (H_o), heterozigosidade esperada (H_e), poder de discriminação (PD), poder de Exclusão (PE), conteúdo de informação polimórfica (PIC). Para este estudo, utilizou-se o programa Arlequin 3.1 (Excoffier *et al.*, 2005) e foi conduzido no programa PowerStats v1.2 (Promega Corporation).

4.3.2.1 Cálculos das frequências dos alelos STR

Neste trabalho, as frequências dos alelos tipados em cada um dos *loci* do multiplex PentaPlex-ISED, foram estimadas contando-se o número de vezes que um alelo foi encontrado em cada população dividido pelo número total de alelos do *locus*.

4.3.2.2 Poder de discriminação (PD)

A probabilidade de discriminação pode ser calculada segundo FISHER (1951).

$$PD = 1 - \sum (\theta)^2$$

θ = frequências genotípicas esperadas.

Para vários *loci* estudados, o poder de discriminação acumulado ($PD_{(A)}$) é dado por:

$$PD_{(A)} = 1 - [(1 - PD_1) (1 - PD_2) (1 - PD_3) \dots (1 - PD_n)]$$

n = *loci* estudado.

4.3.2.3 Probabilidade de coincidência

É a probabilidade de um mesmo genótipo ser originado por mais de uma pessoa. A probabilidade de que dois indivíduos escolhidos ao acaso tenham um dado genótipo é o quadrado de sua frequência na população. Assim, para um genótipo homocigoto a fórmula usada é: (p_i^2) e para um genótipo heterocigoto $(2 p_i p_j)^2$.

A probabilidade de que duas pessoas tomadas ao acaso tenham o mesmo genótipo não especificado é a soma dos quadrados das frequências de todos os genótipos possíveis (HARTL & CLARCK, 1997).

4.3.3 Estudo de Populações

Na realização dos estudos de populações, as amostras de DNA empregadas foram provenientes das populações do Estado do Amazonas, Estado do Rio de Janeiro e Estado do Espírito Santo. A análise populacional ocorreu a partir de uma base de dados das frequências alélicas das três populações em estudos reuniram 308 indivíduos. Para cada *locus*, as frequências dos alelos foram calculadas contando-se o número de vezes que um alelo é encontrado em cada população e dividindo-se esse número pelo total de alelos detectados na mesma população.

A frequência que determina o perfil genético observado para uma amostra de DNA pode ser calculada através do produto das frequências individuais dos alelos. Para tal é preciso ser demonstrado que os *loci* considerados estão em equilíbrio na população de acordo com a lei de Hardy-Weinberg.

4.3.3.1 Comparação populacional

Neste trabalho as populações foram comparadas com base no nível de diferenciação genética. As análises foram comparadas utilizando o software Arlequin 3.1, (Excoffier *et al.*, 2005) onde foram observados:

- níveis de diferenciação entre as populações estudadas, utilizando calculados da análise de variância molecular (AMOVA), (Excoffier *et al.*, 1992);

- os valores de Φ_{ST} que é análogo ao F_{ST} (*stepwised mutation model*) de Wright (1969) (Weir & Cockerham, 1984);

- o teste de Mantel (Mantel, 1967) a partir de ferramentas disponíveis no programa *Arlequin 3.1* (SCHNEIDER *et al.*, 2006). Foi também analisado o distanciamento geográfico, com o objetivo de verificar se este está interferindo no fluxo gênico entre as populações.

Na análise de Variância Molecular, do inglês *Analysis of Molecular Variance* (AMOVA), foi novamente utilizado o programa Arlequin versão 2000 (Schneider et al. 2000), para estimar a diferenciação genética das três populações pela análise de variância molecular (AMOVA), O que permite as análises hierárquicas de três componentes da variância genética, ou seja, aquele devido a diferenças genéticas entre indivíduos dentro das populações, entre populações, dentro dos grupos e entre grupos. O teste de significância dos valores de variância genética estimado com o uso de 10.000 permutações (EXCOFFIÉER *et al.*, 1992). Os valor de p extraído da AMOVA representa a probabilidade de se obter ao acaso valores maiores ou iguais àqueles de F_{ST} (déficit de heterozigotos entre populações) e dos componentes "reais" da variância. A rejeição da hipótese de ausência de diferenciação genética ocorre em um nível de significância de $p < 0,05$. Para o presente trabalho foram utilizadas 10.000 permutações aleatórias para calcular os valores de F_{ST} e de componentes da variância representante da distribuição nula.

Na análise estatística envolvendo múltiplos testes de significância independente, aplicou-se a correção de BONFERRONI (1936). Onde o valor de $p \leq 0,05$, dividiu-se o nível de significância de cada teste pelo número de populações analisadas menos um ($0,05 / n-1$), obtendo assim um valor de p corrigido.

4.3.3.2 Detalhamento das análises para o estudo de população

Para a realização das análises populacionais foram preparados arquivos em formatos de bloco de notas, contendo as bases de dados dos *loci* STRs: D5S198498, D8S82535, D11S118590, D3S18773 e GH15, componentes do sistema PentaPlex-ISED, provenientes das populações estudadas. Os dados foram importados para o *software* estatísticos Arlequin 3.1 (Excoffier *et al.*, 2005), visando calcular as frequências gênicas de cada *loci*, e verificar se eles estão em equilíbrio Hardy-Weinberg (HWE), em seguida construir um banco de dados. Onde foram analisados os seguintes parâmetros: Número médio de alelos por *locus*; Frequências alélicas; heterozigosidade observada (H_o); heterozigosidade esperada (H_e) e verificar se esses alelos estão em equilíbrio de Hardy-Weinberg (HWE).

5 RESULTADOS E DISCUSSÃO

5.1 SELEÇÃO DOS *LOCI* STRS

Dos 15 *loci* microssatélites desenvolvidos por Pontes 2003, apenas 5 *loci* foram selecionados por apresentarem maior grau de polimorfismo e melhor padrão de amplificação. As características relacionadas tipo de sequência de repetição, temperatura de anelamento, tamanho de pares de bases (pb), localização cromossomal e referências no GenBank estão demonstradas na Tabela 05.

TABELA 04: Relação dos 5 marcadores STR selecionados e suas características

| <i>Locus "Primers"</i> | <i>Seqüência Repetida</i> | <i>T_m (°C)</i> | <i>Tamanho pb</i> | <i>Localização Cromossomal</i> | <i>Referências GenBank</i> |
|------------------------|---------------------------|---------------------------|-------------------|--------------------------------|----------------------------|
| ISO3G | (GATA) ₁₀ | 60 | 200-270 | 11p11+2 | AC021573 |
| ISO2M | (GGCT) ₁₁ | 60 | 180 -220 | 5q35 | AC026689 |
| ISO2G | (TCTG) ₁₁ | 60 | 240-280 | 8p23 | AC022846 |
| ISO1M | (GATA) ₁₀ | 60 | 218-220 | 3p25 | AC090958.3 RP11-758G22 |
| ISO5P | (GATA) ₁₃ | 60 | 140-190 | 15p11+2 | AC073446 RPN1-757E13 |

Esses marcadores foram identificados no genoma humano, não haver até o momento desta um estudo mais detalhado destas regiões, somente a região 15p11+2 que encontra-se próximo a região da golgina no (RPN1-757E13).

Os 5 *loci* identificados nos cromossomos humanos estão apresentados na FIGURA 06 os marcadores ISO1M, ISO2M, ISO2G, ISO3G e ISO5P, além dos cromossomos sexuais (XY).

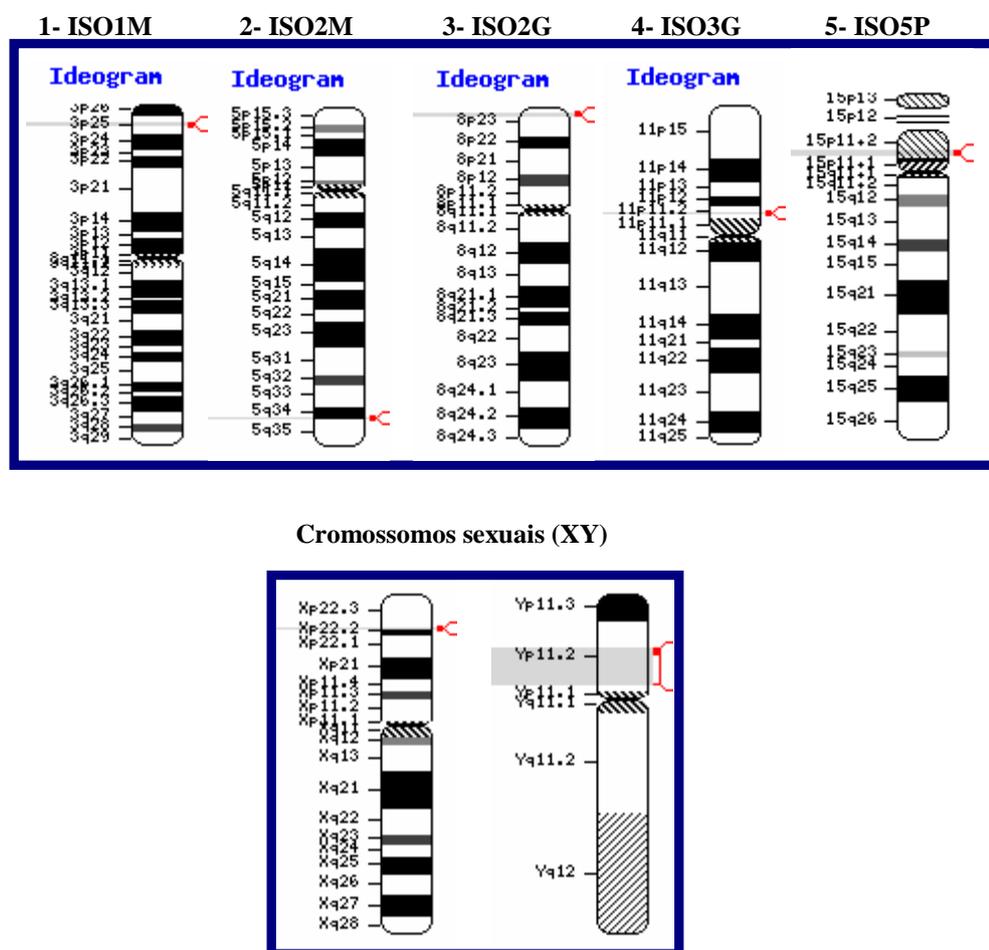


Figura 07: Posição dos novos marcadores STRs e também a localização do *locus* da amelogenina nos cromossomos sexuais.

5.2 DESENHO DOS *PRIMERS*

Os desenhos dos *primers* foram realizados após a identificação dos *loci* microssatélites STRs no genoma humano por meio do programa BLAST-N que analisa a homologia comparativa das sequências. Tanto o banco de genes quanto o programa de busca podem ser acessados por meio da página eletrônica da NCBI (<http://www.ncbi.nlm.nih.gov/BLAST>) (AUSTCHUL *et al.*, 1990 e 1997).

Após a identificação das sequências STR e suas sequências flanqueadoras, os primers foram desenhados com TMs de aproximadamente de 55°C a 62°C, veja TABELA 06.

Tabela 05: Sequência dos *primers* e suas posições nos cromossomos humanos. Os números indicam a base inicial e final de cada primer na montagem de referência de cada cromossoma no GenBank.

| <i>Locus</i> | <i>Cromosomas</i> | <i>Primer</i> | <i>Posição 5'-P</i> | <i>Posição 3'-OH</i> | <i>Seqüência dos primers 5'(-→)3'</i> |
|--------------|-------------------|---------------|---------------------|----------------------|---------------------------------------|
| ISO2M | 5 | R | 12668130 | 12668110 | CTCCTTCGAACCTTCTTCCTG |
| | | F | 12667927 | 12667948 | ATGTCCTATCCAGCCCTTCACC |
| ISO3G | 11 | R | 45748636 | 45748607 | AAGATGTATTAATGAAAAAGTATGCACTG |
| | | F | 45748372 | 45748395 | AGGAGTTCAAGACCAGACTGGATG |
| ISO5P | 15 | R | 164657 | 164683 | CTCTTGTTACAAAAATCAGTGTCTCC |
| | | F | 164828 | 164807 | CAGGTCTCCAGTACACAGGAC |
| ISO2G | 8 | R | 3557181 | 3557200 | TGGACACTTCTCCCCACACG |
| | | F | 3557444 | 3557423 | TATTGCAGGGATTGCCATAGG |
| ISO1M | 3 | R | 11948981 | 11949002 | AATTACGCCACCAGCTTTGTTG |
| | | F | 11949199 | 11949178 | GTGAACATCTGCCCTGCATTAG |

É importante ressaltar que os marcadores STR desenvolvidos neste trabalho identificados no genoma humano não foram ainda descritos para uso em análise genética.

5.2.1 Nomenclatura dos STRs

A nomenclatura destes marcadores STRs ficou definidos conforme mostra a nomenclatura utilizada por (Butler, 2006), para os novos microssatélites na Tabela 07. Ficaram os novos STRs com a nomenclatura seguinte forma: onde mostra a letra **D** refere-se ao DNA, a letra **S** e o número correspondem à localização no cromossomo humano, finalizando com a numeração da posição do *primers* no genoma humano. **GH** refere-se ao gene da golgina humana no cromossomo **15**.

Tabela 06: Nomenclatura dos novos marcadores STR.

| <i>STRs novos</i> | <i>Nomenclatura dos novos STRs</i> | <i>STRs do CODIS</i> | <i>Cromossomo</i> |
|-------------------|------------------------------------|----------------------|-------------------|
| IS02M | D5S198498 | D5S818 e CSF1PO | 5 |
| IS02G | D8S82535 | D8S1179 | 8 |
| IS03G | D11S118590 | TH01 | 11 |
| IS01M | D3S18773 | D3S1358 | 3 |
| IS05P | GH15 | | 15 |

Os resultados observados quanto à posição dos 5 *loci* nos cromossomos mostrou que quatro deles situam-se nos mesmos cromossomos de alguns dos *loci* que formam o CODIS. Assim temos o *locus* D3S1358 do CODIS, que está no mesmo cromossomo do *locus* D3S18773; o *locus* D8S1179 do CODIS está no mesmo cromossomo do *locus* D8S82535; os *loci* D5S818 e CSF1PO do CODIS estão no cromossomo do *locus* D5S198498; o *locus* TH01 do CODIS está no cromossomo do *locus* D11S118590, o *locus* GH15 é o único *locus* que não está inserido entre os cromossomos que fazem parte do CODIS.

5.2.2 Marcação dos *primers*

Para a detecção automatizada dos marcadores STRs foi necessário proceder à marcação dos *primers* utilizando fluoróforos HEX - *Hexametildamina* (no eletroferograma apresenta de cor verde), NED - *Dye yellow* (no eletroferograma apresenta preto) e FAM - *Carboxi-fluoresceína* (no eletroferograma apresenta na cor azul), na extremidade 5' como mostra a TABELA 08, deixando um intervalo de aproximadamente 20 - 40pb para que não houvesse sobreposição de alelos de um *locus* em outro marcado com o mesmo fluoróforo e interferência nas análises.

Tabela 07: Os iniciadores STR e as seqüências marcadas com fluoróforos (NED, FAM e HEX).

| <i>Locus</i> | <i>Seqüência dos iniciadores</i> (5' → 3') | <i>Fluoróforos</i> |
|--------------|--|---------------------------------|
| <i>Novo</i> | | |
| D3S18773 | Forward: GTGAACATCTGCCCTGCATTAG Reverse: AATTACGCCACCAGCTTTGTTG | NED Dye (yellow) |
| D5S198498 | Forward: ATGTCCTATCCAGCCCTTCACC Reverse: CTCCTTCGAACCTTCTTCCTG | FAM Carboxi- fluoresceína |
| D8S82535 | Forward: TATTGCAGGGATTGCCATAGG Reverse: TGGACACTTCTCCCCACACG | HEX Hexametilrodamin a |
| D11S118590 | Forward: AGGAGTTCAAGACCAGACTGGATG Reverse: AAGATGTATTAAATGAAAAAGTATG CACTG | FAM Carboxi- fluoresceína |
| GH15 | Forward: CAGGTCTCCCAGTACACAGGAC Reverse: CTCTTGTTACAAAATCAGTGTTCTCC | HEX Hexametilrodamin a |

A tipagem de marcadores STRs baseada na detecção por fluorescência é hoje praticamente universal, sendo usada em laboratórios de genética forense em todo o mundo (BRETTELL *et al.*, 2005).

5.3 DEFINIÇÃO DAS CONDIÇÕES DA PCR

Para definir as condições da PCR foram realizadas amplificações com diferentes enzimas baseando-se nas condições levantadas na literatura e nas sugestões dos fabricantes das enzimas. Os resultados desses estudos estão mostrados na Tabela 09. Diversas amplificações por PCR foram realizadas como sistema monoplex, a fim de avaliar a funcionalidade dos *primers*. Os produtos foram analisados em gel de agarose 2,5%, corado com brometo de etídio e em seguida através de gel de poliacrilamida (PAGE) 6%, corado com nitrato de prata.

Tabela 08: Definições das condições da PCR para os marcadores STRs monoplex

| Componentes da reação de PCR | Quantidade final para 20 µL |
|------------------------------|--|
| H ₂ O | 2,2 µL |
| Tampão 10X | 2,5 µL |
| MgCl ₂ (20 mM) | 2,5 µL |
| dNTP (2,5 mM) | 2,5 µL |
| Primer (R) (5 pmol/µL) | 2,5 µL |
| Primer (F) (5 pmol/µL) | 2,5 µL |
| Taq. DNA (5 U/µL) | 0,3 µL |
| Solução de DNA (5 ng/µL) | 5,0 µL |
| Vol. Final | 20,0 µL |
| Termociclo utilizado: | 95 °C – 2 min 40x (94 °C p/40 seg; 60 °C p/ 60 seg; 72 °C p/ 90 seg) 60 °C p/ 10min. 4 °C hold. |

Nesta etapa, as condições de amplificação utilizadas foram ajustadas conforme estudos relacionados com amplificações de cada *loci* STR, observados as condições individuais e observando estudos já realizados com outros STRs (PROMEGA, 1995; PROMEGA, 1998).

A FIGURA 07 mostra os perfis eletroforéticos dos amplicons em gel de agarose 2,5 %, corado com brometo de etídeo (1µg/mL), utilizando amostras de DNA de três indivíduos (3M, 3C, 3SP e controle negativo). Pode-se observar que utilizando esta metodologia não é possível obter a separação dos alelos microssatélites.

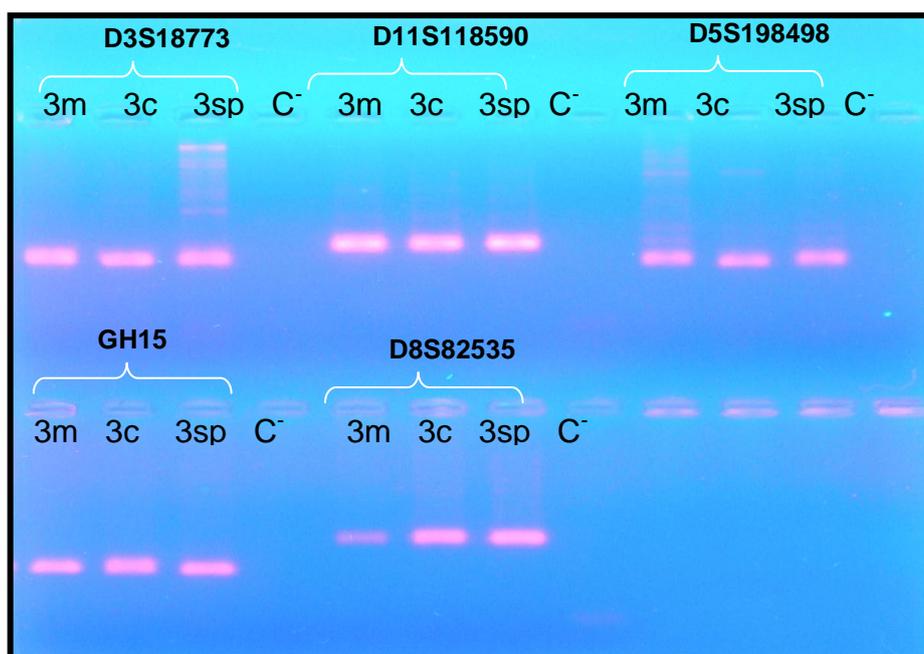


Figura 08: Mostra o perfil eletroforético de DNA dos 3 indivíduos (3m, 3c, e 3 sp), analisado em gel de agarose 2,5 % corado com brometo de etídeo (1 μ g/ μ L).

A seguir os amplicons para os monoplex (D3S18773, D5S198498, D8S82535, D11S118590 e GH15) foram analisados por eletroforese em gel de poliacrilamida como mostra a Figura 08. Pode-se observar em todos os casos a presença de um alelo nos indivíduos homozigotos de dois alelos nos indivíduos heterozigotos. A visualização de dois alelos microssatélites só é possível pelo alto poder de resolução desse tipo de eletroforese. Em alguns casos pode-se observar “sombras” que migram mais, estas correspondem aos *stuttes*.

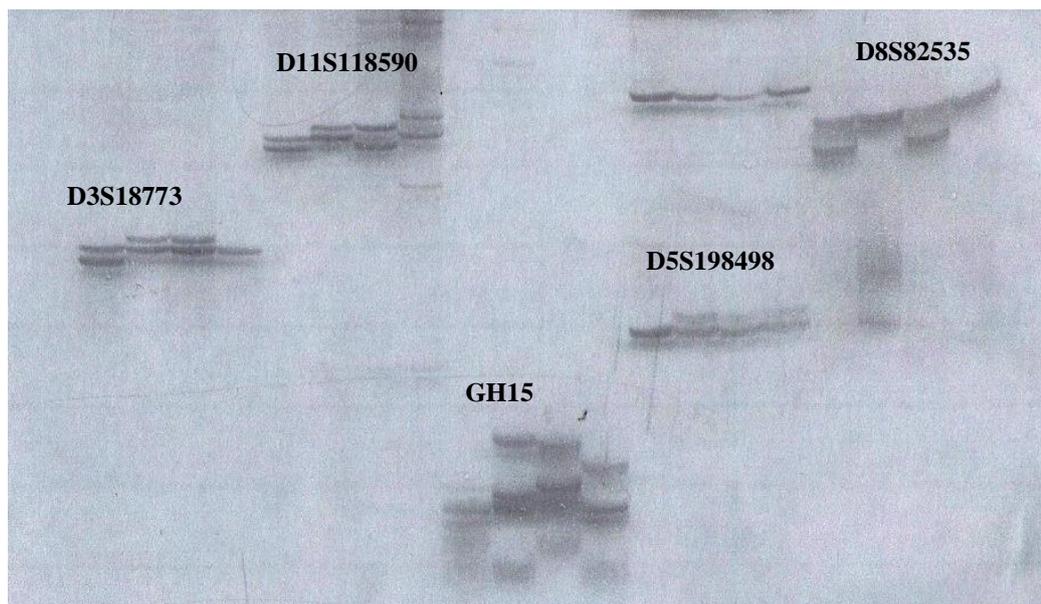


Figura 09: Perfil eletroforético em gel de poliacrilamida (PAGE) 0,6 % corado com nitrato de prata, evidenciando a amplificação por PCR dos sistemas monoplex “STRs”.

As condições empregadas para a amplificação de uma única região STR, nas (reações monoplex) diferem daquelas empregadas para a amplificação simultânea de um conjunto de regiões STRs, conhecida como reação multiplex. Para essas reações leva-se em conta a condição de amplificação, que devem ser ajustadas de forma que possibilitem uma geração de produtos simultânea e equitativamente a partir de cada uma das regiões STR que constituem os marcadores do sistema multiplex (BUTLER, 2003).

O complexo magnésio – dNTP é o principal substrato para a reação de alongamento da cadeia. O magnésio se liga preferencialmente às cargas negativas dos grupos fosfato, reduzindo assim as forças repulsivas entre esses grupos. A diminuição desses reagentes na reação da PCR pode promover a diminuição da quantidade dos produtos ou aparecimento de novos amplicons (LEVEDAKOU *et al.*, 2001).

A enzima AmpliTaq Gold DNA polimerase, por exemplo, sob condições ótimas, permanece ligada ao DNA molde até 40 a 60 nucleotídeos tenham sido replicados. Esta ligação, por sua vez pode ser afetada por uma série de condições como temperatura, força

iônica e interação com outras proteínas. Quando as condições de estringência são modificadas a taxa de processamento da enzima pode diminuir, afetando preferencialmente a polimerização dos fragmentos de maior tamanho (SILVA, 2004 & MANTA, 2008).

Na TABELA 10, está representada a mistura do sistema de reação da PCR para o Sistema PentaPlex-ISED, que pode ser utilizado com as enzimas DNA Polymerase HotStarTaq QIAGEN (Multiplex PCR Handbook); Platinum *Taq* DNA Polymerase (Invitrogen).

Tabela 09: Mostra as condições da PCR para reação do Sistema PentaPlex-ISED, que pode ser utilizado com diferentes enzimas na mesma concentração e com primers marcados com fluorocromos.

| Componentes da reação de PCR | Quantidade final para 10 µL |
|--|---|
| H ₂ O | 4,3 µL |
| Tampão Platinum Plus (4X) | 2,5 µL |
| <i>Primers</i> (PentaPlex-ISED) (100 pmol) | 1,0 µL |
| Taq DNA Polimerase Platinum (5U/µL) | 0,2 µL |
| Solução de DNA (1 ng/µL) | 2,0 µL |
| Vol. Final | 10,0 µL |
| Sistema de amplificação | 95 °C – 2 min 40x (94 °C p/40 seg; 62 °C p/ 60 seg.; 72 °C p/ 90 seg.) 60 °C p/ 60 min. 4 °C hold. |

O tampão Platinum Plus (4X) foi preparado a partir das soluções estoque (kit) (MgCl₂, tampão 10X, dNTPs) mostrado na TABELA 04 em Material e Métodos. As condições de amplificação foram ajustadas de forma que possibilitassem a amplificação simultânea dos 5 STRs escolhidos. Quando utilizamos enzima comum no sistema monoplex, os resultados foram ótimos como foi mostrado nas FIGURAS 07 e 08. No sistema multiplex, esse tipo de

enzima não foi eficiente, pois a sua processividade é muito baixa. Quando foram utilizadas outras enzimas com processividade maior os resultados foram excelentes.

As enzimas *Taq* DNA polimerase modificadas, como a *AmpliTaq Gold™* DNA polimerase (Applied Biosystem) e *Platinum® Taq* DNA polimerase (Invitrogen) possuem alta especificidade e sensibilidade.

5.4 AMOSTRAS BIOLÓGICAS E ANÁLISE DE DNA

No estudo das amostras biológicas foram arrolados inicialmente 104 indivíduos geneticamente não relacionados pertencentes à população de Manaus (AM). Amostras de sangue foram coletadas, os DNAs extraídos, conforme metodologia descrita em Material e Métodos. A seguir alíquotas das amostras de DNA foram analisadas por eletroforese em gel de agarose 0,8%, corado com brometo de etídio. As concentrações foram estimadas tomando-se como padrão o DNA humano K562. Os resultados estão mostrados na Figura 09. Foram aplicados no gel 5µL de cada amostra, além do padrão de DNA K562 na concentração 5ng/µL.

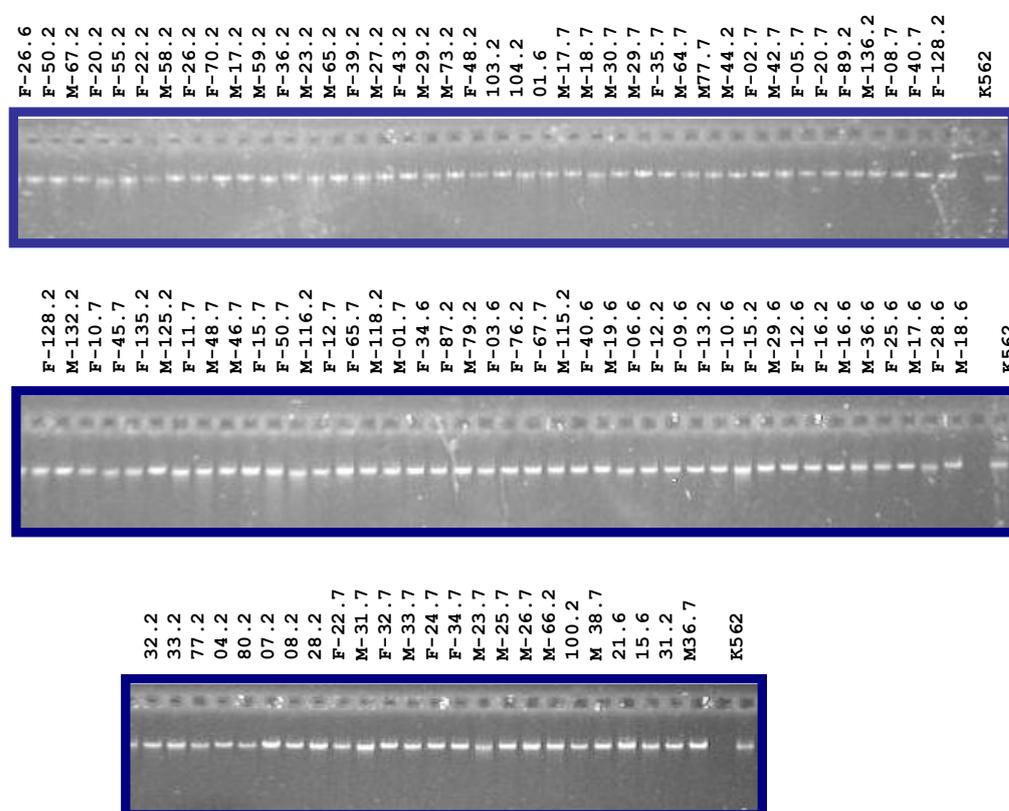


FIGURA 10: Perfis eletroforéticos das amostras de DNA extraídas da população do Amazonas. O DNA K562 que é DNA padrão de referência.

As análises por eletroforeses, em géis de agarose a 0,8%, permitiram conferir a integridade das amostras (ausência de rastro) de DNA e a comparação com a fluorescência do DNA padrão K562 (5ng/μL) permitiu estimar-se as concentrações do DNA das amostras em (5 – 10ng/μL).

5.5 CONSTRUÇÃO DA ESCADA ALÉLICA

Na construção da escada alélica, foram utilizadas algumas amostras de DNA da população do Estado do Amazonas, as quais estão mostradas na Tabela 11, e o sistema

monoplex para amplificação apresentado na TABELA 09. A TABELA 11, também mostra os volumes dos amplicons de cada STR que foram misturados para compor a escada alélica. Em seguida a escada alélica foi analisada por eletroforese em gel de poliacrilamida com nitrato de prata como mostra na FIGURA 10.

Tabela 10: Mostra a mistura para a composição da escada alélica, utilizando diversas amostras de DNA com perfis conhecidos.

| <i>Locus iniciadores</i> | <i>DNA utilizado na amplificação</i> | <i>Volume (μL)</i> |
|--------------------------|--------------------------------------|--|
| D3S18773 | 23.2 e 65.2 | 4,0 μL |
| D5S198498 | 40.6 e 40.7 | 4,0 μL |
| D8S82535 | 11.7 e 77.2 | 6,0 μL |
| D11S118590 | 05.6 e 58.2 | 4,0 μL |
| GH15 | 11.7 e 20.7 | 5,0 μL |

Nos resultados mostrados na FIGURA 10 indicam que os padrões de migração dos marcadores D8S82535 e D11S118590 foram similares, obviamente possuem o tamanho de pb muito próximo em torno de (260/280), todos migraram na mesma faixa do gel de poliacrilamida. Os fragmentos amplificados nos *loci* D3S18773 e D5S198498 também foram similares, obviamente porque possuem o número de pares de base (pb) muito próximo na faixa de (200/230). O *locus* DGH15 apresenta fragmentos em uma faixa de aproximadamente 100/180 pares de base (pb), pois seus produtos fizeram um percurso maior durante a corrida eletroforética tendo apresentado alelos bem definidos.

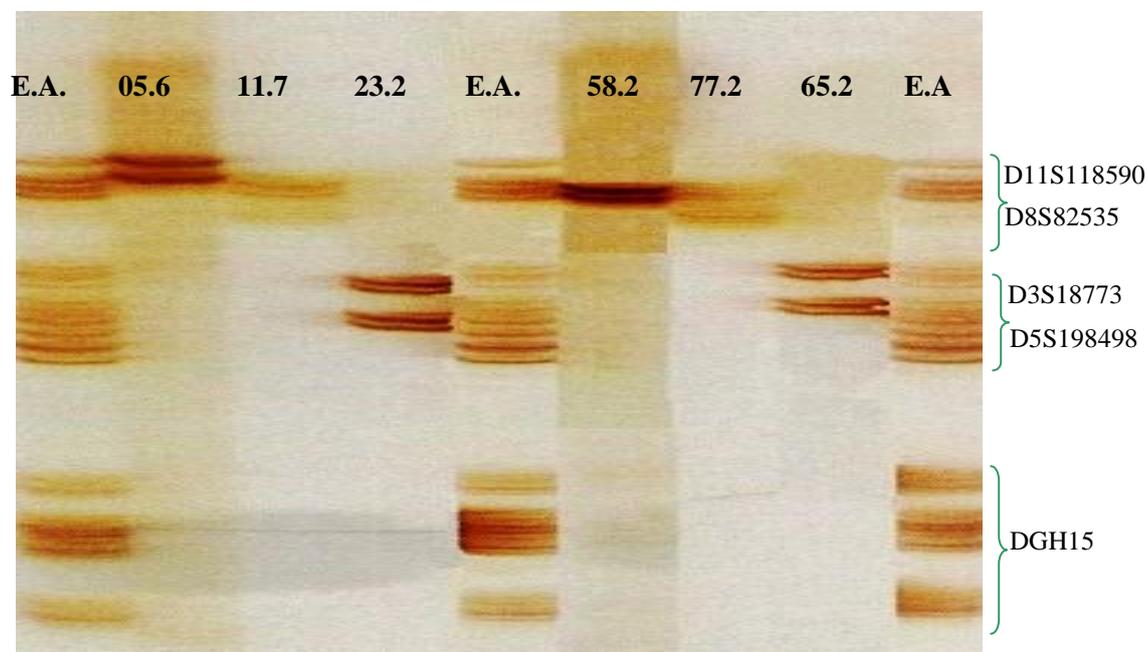


Figura 11 : Perfil eletroforese dos 5 monoplexes em gel de poliacrilamida corado por prata da Escada Alélica (E.A), utilizando DNAs (05.6, 11.7, 23.2, 58.2, 77.2, 65.2) da população do Amazonas.

Na escada alélica, que foi produzida para ser utilizada nos sistemas de eletroforese em gel de poliacrilamida vertical, foram incluídos os alelos mais freqüentes observados para cada *loci*. Os alelos amplificados tiveram suas concentrações estimadas e, então, misturados de forma a gerar uma escala alélica balanceada para cada *locus*.

No sistema eletroforético em capilar é utilizado um padrão de peso molecular, denominado LIZ, marcado com fluorescência de cor laranja. Esse padrão serve para comparar os fragmentos de DNA quanto ao peso molecular.

Para leitura no sistema automatizado, esses iniciadores STR, foram marcados com diferentes fluorocromos. Além disso, a padronização das condições eletroforéticas, conforme demonstrado neste trabalho, proporcionaram um bom padrão de separação dos alelos,

facilitando a detecção dos mesmos tanto visualmente quanto em sistemas automatizados, podendo ser detectado durante a genotipagem das amostras de DNA.

5.6 COMPOSIÇÃO DO SISTEMA PENTAPLEX-ISED

O protótipo do sistema PentaPlex-ISED apresentado neste trabalho foi preparado com cinco marcadores genéticos STRs previamente estudados e selecionados que se encontram localizados em diferentes cromossomos, apresentando repetições tetranucleotídeos, os quais foram desenvolvidos por Pontes (2003), como mostrado na Tabela 03, e que esse conjunto de marcadores apresentavam elevado grau de polimorfismo, o que justificava a utilização dos mesmos na análise genética com a finalidade de identificação humana com base nos elevados graus de polimorfismo por eles apresentados.

Para a composição do sistema PentaPlex-ISED procedeu-se à mistura de cada par de iniciadores em diferentes concentrações, a partir das soluções estoques dos iniciadores marcados na concentração de 100 pmoles/ μ L em volume de 20, 40 e 80 μ L, tendo o cuidado na hora de proceder à mistura para que tivéssemos um equilíbrio na intensidade dos cinco *loci* STRs. Na composição final do multiplex foi incluído a amelogenina, que é o único marcador de alossomos (cromossomos sexuais), que possui em torno de 106 pb no cromossomo X, e 112 pb no cromossomo Y. Este gene está presente na parte homóloga desses cromossomos e serve também para determinação sexual.

Após a amplificação dos alelos STR no sistema multiplex, com marcação com fluorocromos, os produtos foram genotipados por eletroforese capilar, no seqüenciador automático de DNA MegaBace GE Healthcare na UFAM e no ABI 3100 da Applied na UERJ.

Várias etapas foram realizadas para a montagem do multiplex com a amplificação de todos os *loci*, observando a intensidade, altura dos picos de cada alelo e a presença de artefatos que pudessem prejudicar a análise dos resultados. Foram também avaliadas a qualidade do produto amplificado e a quantidade de DNA necessária para uma boa detecção, a partir dos DNAs padrão.

Na padronização do sistema multiplex foram utilizadas alíquotas de DNA das linhagens celulares 9948, 9947 e K562, amplificadas por PCR e o produto analisado conforme descrito em Material e Métodos. Na FIGURA 11, são mostrados os alelos do DNA padrão (K562) utilizando os marcadores STR D3S18773, D5S198498, D8S82535, D11S118590 e GH15 marcados com fluorocromos.

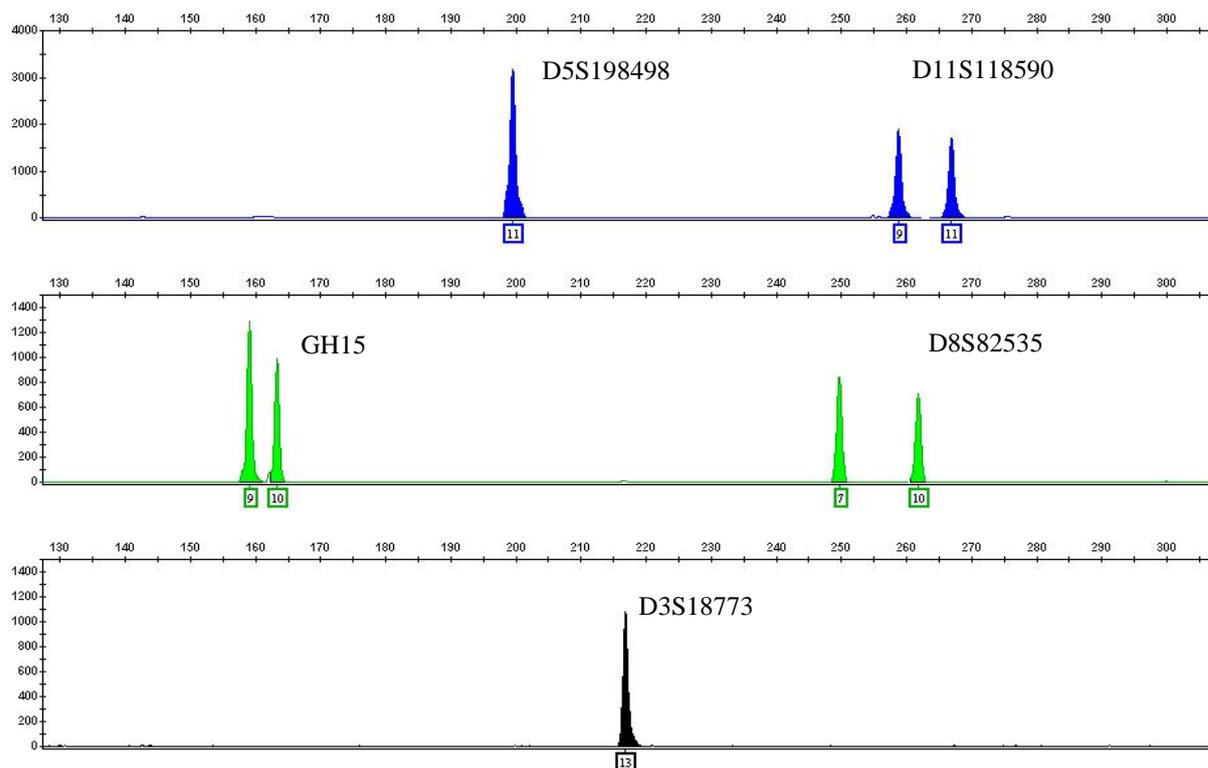


Figura 12: Alelos STR amplificados por PCR a partir dos DNA controle utilizando o sistema PentaPlex-ISED para os *loci* D5S198498 e D11S118590 marcados com fluoresceína, GH15 e D8S82535, marcados com hexametiltridamina e D3S18773, marcado com tetrametiltridamina. A amostra foi analisada após eletroferograma dos produtos amplificação DNA (K562) analisado por eletroforese em capilar no seqüenciador ABI 3100.

As análises dos produtos da amplificação com sistema PentaPlex-ISED revelaram um equilíbrio entre as quantidades dos produtos amplificados nos diferentes *loci*, e o poder da eletroforese capilar para separar os alelos STR (FIGURA 11).

Atualmente, os sistemas multiplexes vêm sendo utilizados com frequência para análise genética de marcadores STRs, com objetivo de promover a construção de bases de dados de frequências alélicas para vários *loci* simultaneamente e, a partir desses estudos dados estatísticos de interesse para vínculos genéticos, investigações criminais e análise evolutiva podem ser gerados (BUDOWLE *et al.*, 1998; FRÉGEAU *et al.*, 1999). Neste estudo podemos citar a construção de multiplex como o PentaPlex-ISED que pode ser uma alternativa para o

Laboratório de Diagnóstico Molecular da UFAM, para a diminuição dos custos das análises genéticas em relação aos sistemas comerciais de tipagem humana disponíveis no mercado.

A utilização de STRs tetra ou pentanucleotídeos em sistema multiplex na tipagem para identificação humana é devida ao fato desses serem menos sujeitos à falha no processo de amplificação, prevenindo-se entre outros eventos o aparecimento de *stutters* (VAN HOOFTAL *et al.*, 1998).

5.7 ESTUDOS DE VALIDAÇÃO DO SISTEMA PENTAPLEX-ISED

Nesta etapa foram desenvolvidos estudos com o propósito de validar o sistema PentaPlex-ISED. O processo de validação abrange aspectos técnicos e metodológicos que são críticos e que precisam ser observados. Esses aspectos são:

- 1 – Estudo da sensibilidade do sistema multiplex;
- 2 – Reprodutibilidade do método da tipagem de DNA em diferentes amostras de um mesmo indivíduo;
- 3 – Padronização da amplificação com diferentes enzimas;
- 5 – Análises das bandas *stutter* produzidas por cada *locus* STR;
- 6 – Determinação de alelos microvariantes em cada *locus* STR.

5.7.1 Análise da sensibilidade do sistema

Visando verificar a quantidade mínima de DNA que se relacionava com intensidade de fluorescência detectável de todos os alelos relativos aos *loci* STRs do sistema PentaPlex-ISED, para análise por eletroforese em poliacrilamida/prata, testando-se várias quantidades de DNA em torno de 1ng. a 50ng., verificou-se que 5ng./ μ L por sistema de amplificação era o ideal para as análises. Por outro lado, para as análises com o sistema PentaPlex-ISED por fluorescência, utilizando-se eletroforese capilar, essa quantidade de DNA por sistema de amplificação mostrou-se excessiva, por isso realizou-se diversas diluições e a menor quantidade de DNA que resultou em intensidade de fluorescência detectável para todos os *loci* foi correspondentes 1/20, ou seja, 0,25ng. de DNA.

Budowle e colaboradores (1997), trabalhando com o multiplex CTT, concluíram que o limite mínimo para amplificação dos alelos se estabelecia entre 0,20 e 0,25ng. Em 1996, Van Oorschot e colaboradores conseguiram sucesso na amplificação por PCR usando somente 0,05ng de DNA. No entanto, neste caso foi utilizado somente o iniciador para o *locus* TH01. Esta observação nos faz concluir que quanto maior o número de regiões analisadas, maior a quantidade de DNA requerida para a amplificação, devido provavelmente à competição dos iniciadores pelos respectivos amplicons. Isso significa que as concentrações dos DNAs utilizados neste estudo estão de acordo com os estudos realizados por vários estudiosos.

5.7.2 Análise da reprodutibilidade do sistema de tipagem

Neste estudo de reprodutibilidade da tipagem, com o sistema PentaPlex-ISED, foram avaliados com diferentes tipos de amostra de DNA. As análises realizadas foram:

1) Tipagem utilizando duas amostras de sangue (F50_7F e 58s_M), e o controle positivo (K562). Os resultados estão mostrados nas FIGURAS 12, 13 e 14, onde foi

observada a tipagem das três amostras amplificadas com o sistema PentaPlex-ISED e seus perfis eletroforéticos bem definidos.

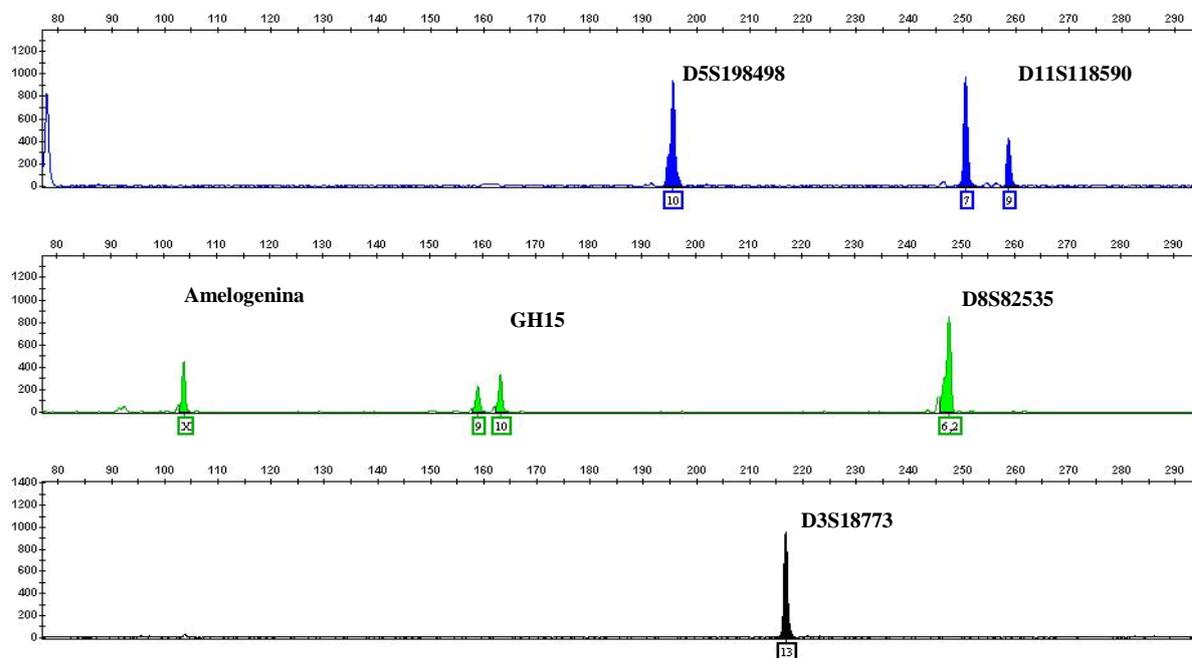


Figura 13: Alelos STR amplificados por PCR, a partir do DNA controle K562 utilizando o sistema PentaPlex-ISED para os *loci* D5S198498 e D11S118590 marcados com fluoresceína, GH15 e D88D8S82535, marcados com hexametiltródamina e D3S18773, marcado com Tetrametiltródamina. A amostra foi analisada por eletroforese em sistema capilar do seqüenciador ABI 3100.

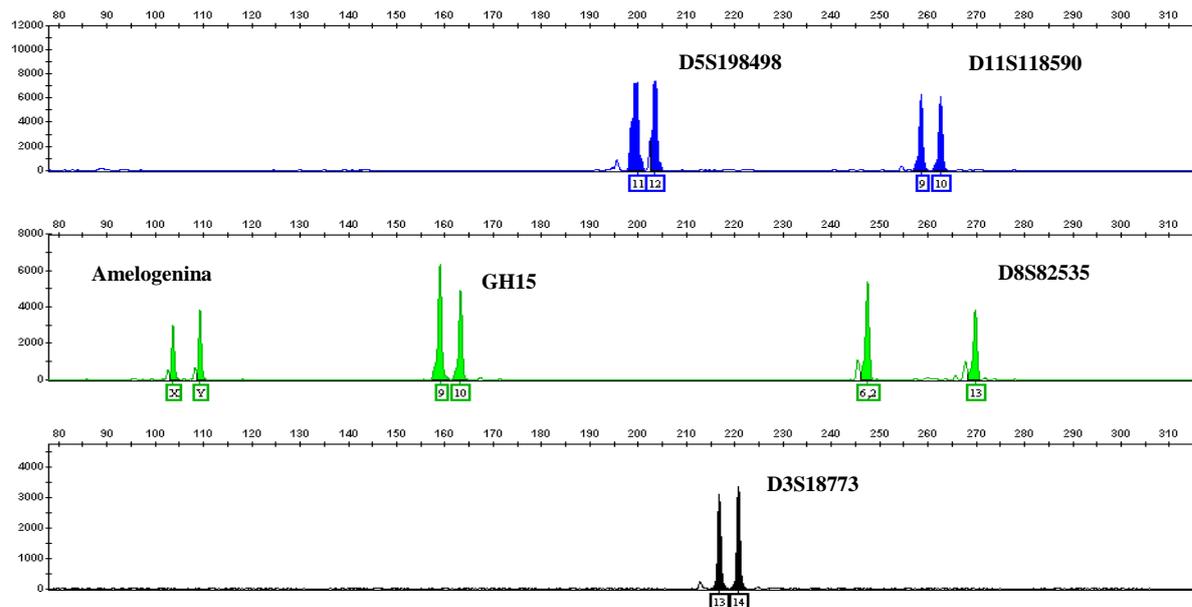


Figura 14: Alelos STR amplificados por PCR a partir de DNA de sangue do indivíduo (58s_M), utilizando o sistema PentaPlex-ISED para os *loci* D5S198498 e D11S118590 marcados com fluoresceína, e D8S82535, marcados com hexametiltródamina e D3S18773 marcado com tetrametiltródamina. A amostra foi analisada após eletroforese em sistema capilar do seqüenciador ABI 3100.

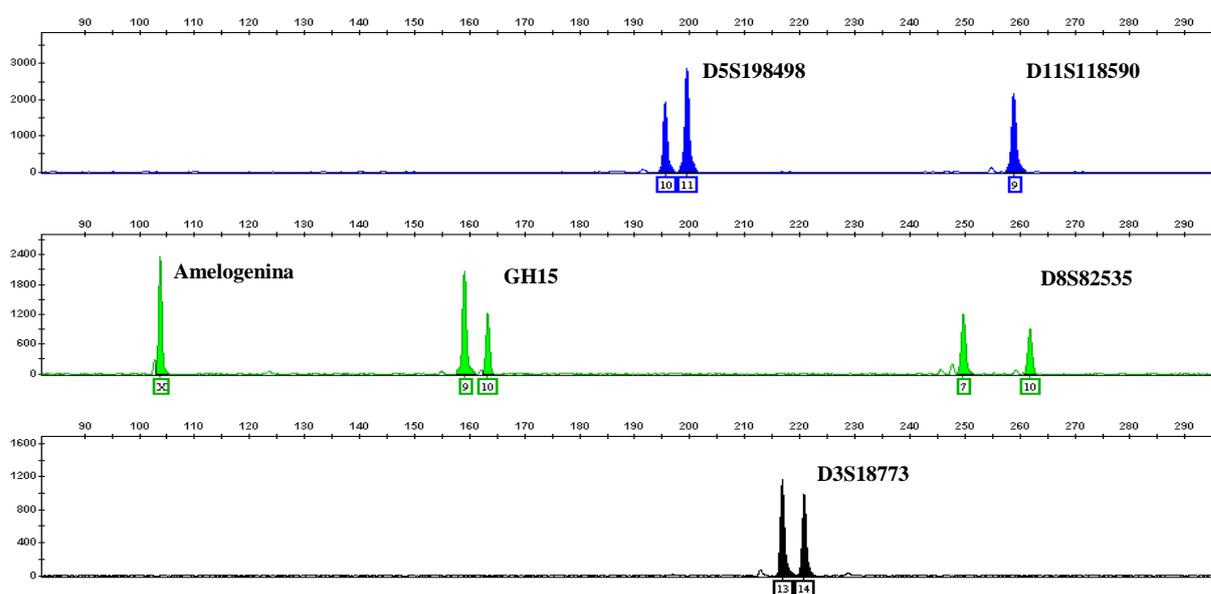


Figura 15: Alelos STR amplificados por PCR a partir dos DNA de sangue do indivíduo (50-7F), utilizando o sistema PentaPlex-ISED para os *loci* D5S198498 e D11S118590 marcados com fluoresceína, GH15 e D8S82535, marcados com hexametiltridamina e D3S18773 marcado com Tetrametiltridamina. A amostra foi analisada por eletroforese em sistema capilar do seqüenciador ABI 3100.

2) Nesta etapa, foram extraídos DNA de amostras biológicas de um mesmo indivíduo sangue, urina, sêmen e saliva. Após amplificação por PCR os produtos foram separados por eletroforese em capilar (ABI 3100) e os alelos detectados foram comparados com o objetivo de verificar se todos os alelos STR foram amplificados com o sistema PentaPlex-ISED. Os resultados estão mostrados nas Figuras 15 a 18. As amostras de DNA amplificadas e os alelos detectados foram comparados com os perfis alélicos das amostras analisadas anteriormente. O *locus* da amelogenina mostrou eficiência para a tipagem das diversas evidências biológicas analisadas.

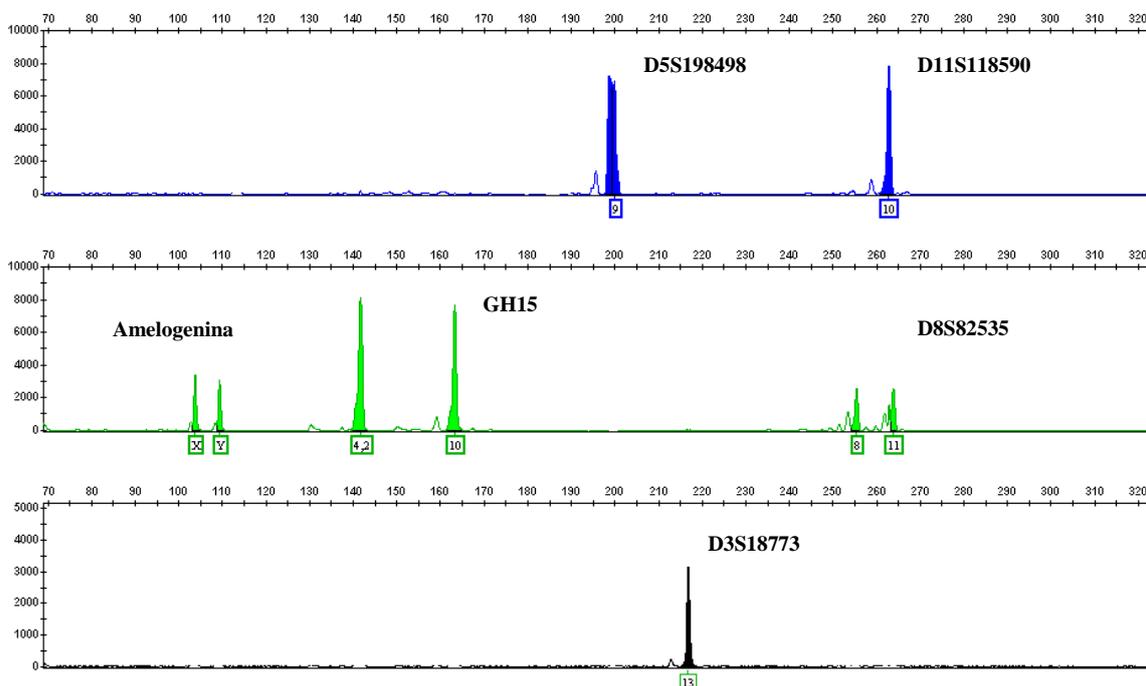


Figura 16 : Alelos amplificados por PCR a partir de amostra de DNA de **Saliva**, utilizando o sistema PentaPlex-ISED para os *loci* D5S198498 e D11S118590 marcados com fluoresceína, GH15 e D8S82535, marcados com hexametiltridamina e D3S18773, marcado com Tetrametiltridamina, + Amelogenina. As amostras foram analisadas por eletroforese em sistema capilar do seqüenciador ABI 3100.

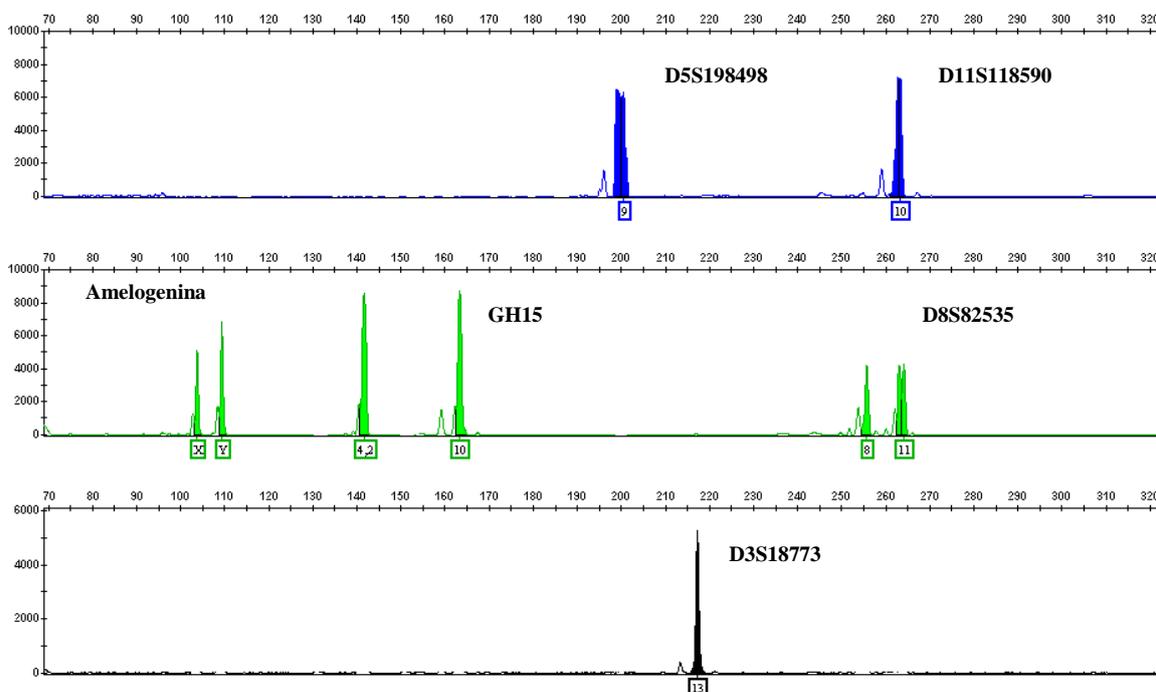


Figura 17: Alelos amplificados por PCR a partir de amostra de DNA de **Sangue**, utilizando o sistema PentaPlex-ISED para os *loci* D5S198498 e D11S118590 marcados com fluoresceína, GH15 e D8S82535, marcados com hexametiltridamina e D3S18773, marcado com Tetrametiltridamina, + Amelogenina. As amostras foram analisadas por eletroforese em sistema capilar do seqüenciador ABI 3100.

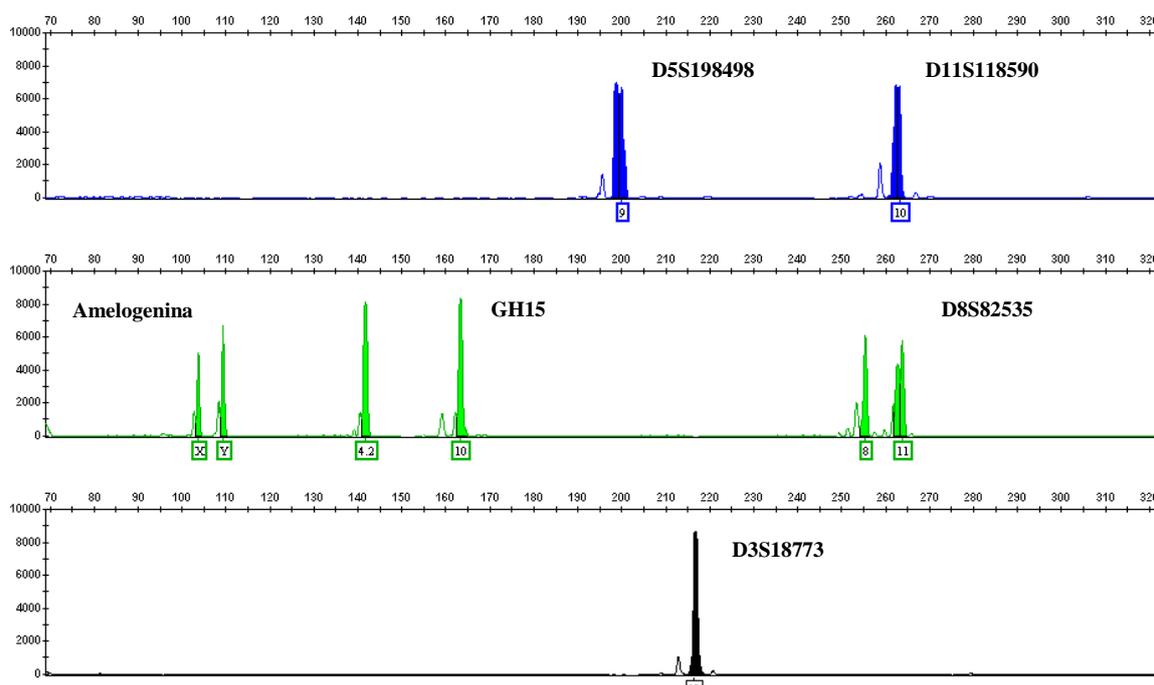


Figura 18: Alelos amplificados por PCR a partir de amostra de DNA de **Sêmen**, utilizando o sistema PentaPlex-I SED para os *loci* D5S198498 e D11S118590 marcados com fluoresceína, GH15 e D8S82535, marcados com hexametilrodamina e D3S18773, marcado com Tetrametilrodamina, + Amelogenina. As amostras foram analisadas por eletroforese em sistema capilar do seqüenciador ABI 3100.

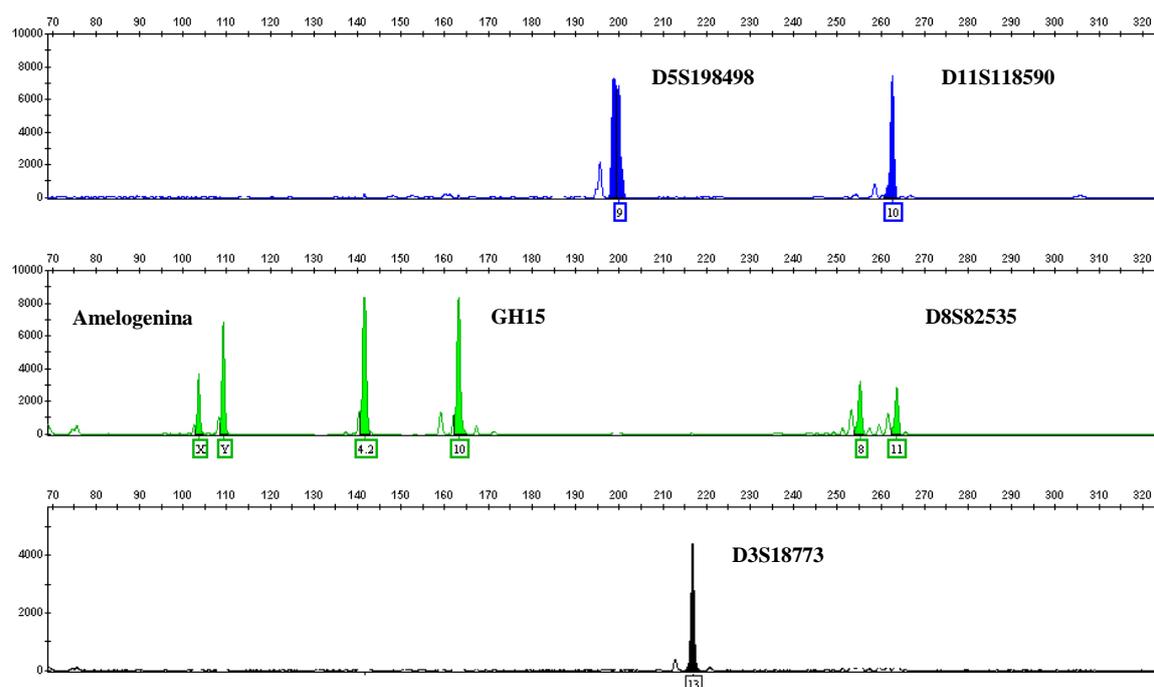


Figura 19: Alelos amplificados por PCR a partir de amostra de DNA de **Urina**, utilizando o sistema PentaPlex-I SED para os *loci* D5S198498 e D11S118590 marcados com fluoresceína, GH15 e D8S82535, marcados com hexametilrodamina e D3S18773, marcado com Tetrametilrodamina, + Amelogenina. As amostras foram analisadas por eletroforese em sistema capilar do seqüenciador ABI 3100.

Observou-se que os *loci* D5S198498, D11S118590 e GH15 apresentaram intensidade dos alelos em torno de 6000 a 9000 RFU em todas as amostras de DNA genotipadas (saliva, sangue, sêmen e urina). Outros *loci* amplificados com o sistema LDD10 (Manta, 2008) apresentaram intensidades mais baixas que os alelos amplificados com o PentaPlex-ISED. Porém, os *loci* D8S82535 e D3S18773 apresentaram intensidade dos picos muito parecidos em torno de 2000 a 5000 RFU, com as do o sistema LDD10. Em todas as amostras tipadas os alelos apresentaram altura bem definida, com exceção da amostra de DNA oriunda do bulbo capilar, que não amplificou, entendemos, seja por causa da concentração do DNA ter sido muito baixa.

3) Tipagem das amostras de DNA extraídas de sangue dos indivíduos das populações dos Estados do Amazonas, Rio de Janeiro, e Espírito Santo estão mostradas os resultados nas FIGURAS 21 a 25.

5.7.3 Padronização da amplificação com diferentes enzimas

Visando verificar a eficiência de amplificação de alelos STR, utilizando o sistema PentaPlex-ISED com diferentes enzimas, amostras extraídas pelo método Chelex e amostras de DNA padrão foram amplificadas com as enzimas HotStar *Taq* DNA Polimerase (QUIAGEN). Após amplificação por PCR, os produtos foram separados por eletroforese em capilar (ABI3100) e os alelos detectados foram analisados quanto às intensidades dos sinais de fluorescência (altura do pico) expressos em RFU. Também foram feitas amplificações com enzima *Platinum Taq* DNA polimerase (INVITROGEN).

Observou-se que as intensidades dos sinais de fluorescência (altura do pico) permitiram a detecção dos alelos com maior clareza comparando com os resultados das enzimas comuns.

5.7.4 Análises das Bandas *Stutter*

A detecção de bandas *stutter* foi realizada por meio da tipagem de 308 amostras de DNA de indivíduos não relacionados geneticamente, das três populações analisadas neste trabalho. As amostras de DNA foram amplificadas por PCR, em seguida separadas, por eletroforese no sistema capilar ABI 3100 momento em que era observada nos *loci* a presença de bandas *stutter*, como mostrado na FIGURA 19.

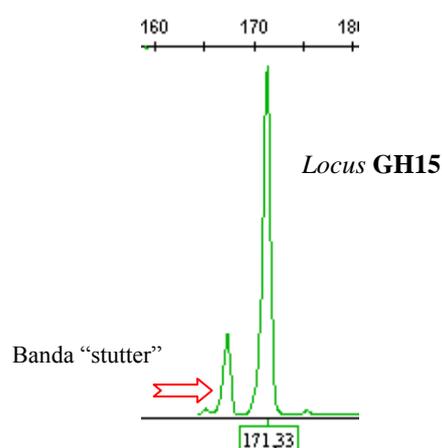


Figura 20: Perfil eletroforético de banda *stutter* no alelo 12 do locus GH15.

Neste estudo, foi determinado o percentual entre as intensidades de banda *stutter* e dos respectivos alelos incluídos no PentaPlex-ISED, quando apresentavam altura do pico *stutter* maior que 50 RFU (Medidas de intensidade de fluorescência).

Esses artefatos, na maioria das vezes são decorrentes de falhas no processo de amplificação, incluindo a adição ou a perda de uma unidade de repetição ao fragmento amplificado por PCR, gerando uma banda de baixa intensidade chamada de *stutter*

(DANIELS *et al.*, 2004). Essas bandas também podem representar sério problema em análises forenses que apresentem misturas de amostras biológicas, podendo confundir os resultados.

Sequências repetitivas ricas em bases complementares, tendem a formar uma estrutura secundária em forma de alça. Dependendo do tamanho dessa estrutura, podem ser induzidas pausas no processamento da DNA polimerase, devido à instabilidade do complexo DNA molde-DNA polimerase. Alternativamente, a enzima pode permanecer estável no complexo, ocorrendo o escorregamento na região da alça (FRESCO & ALBERT, 1960; LEVINSON & GUTMAN, 1987). Essas falhas de polimerização produzem amplicons menores que correspondem aos *stutter*.

Os percentuais de *stutter* observados neste trabalho estão em concordância com os verificados em outros estudos de validação envolvendo *loci* com quatro e cinco nucleotídeos constituindo as unidades de repetição (MICKA *et al.*, 1999; LEVEDAKOU *et al.*, 2002).

O uso de marcadores STRs tetra ou pentanucleotídico em sistema de tipagem é recomendado devido ao fato desses marcadores serem menos sujeitos a falhas no processo de amplificação, prevenindo entre outros eventos o aparecimento dos *stutter* (VAN HOOFFSTAT *et al.*, 1998). Van Hoofstat e colaboradores verificaram uma incidência menor de *stutter*, com o aumento do tamanho da unidade de repetição.

5.7.5 Determinação de alelos microvariantes em cada *locus* STR

No desenvolvimento deste trabalho, foram detectadas em alguns *loci* as ocorrências de alelos microvariantes, pois apresentavam migração eletroforética sem correspondência com os alelos que compunham a escada alélica. Nesta etapa foi determinada a taxa de ocorrência desses alelos nas três populações estudadas (Amazonas, Rio de Janeiro e Espírito Santo), somando um total de 308 amostras de DNA analisadas.

Os resultados observados nesta fase estão apresentados na FIGURA 20, que mostra o percentual de alelos microvariantes para cada *locus* analisado. O *locus* D3S18773 não apresentou microvariância. No *locus* D8S82535 foi observado um maior percentual de alelos microvariantes: 15 alelos (5.1, 6.1, 6.2, 6.3, 7.1, 7.2, 8.1, 8.2, 8.3, 9.1, 9.2, 9.3, 10.1, 10.2 e 11.2). O segundo que apresentou mais microvariância foi o *locus* D11S118590 em que foram observados 8 alelos microvariantes (4.2, 5.2, 7.3, 8.2, 8.3, 9.3, 10.3 e 11.3). No *locus* GH15, 5 alelos microvariantes foram observados (8.3, 9.1, 9.3, 10.1 e 10.3), e no *locus* D5S198498, 2 alelos microvariantes (9.3 e 10.3).

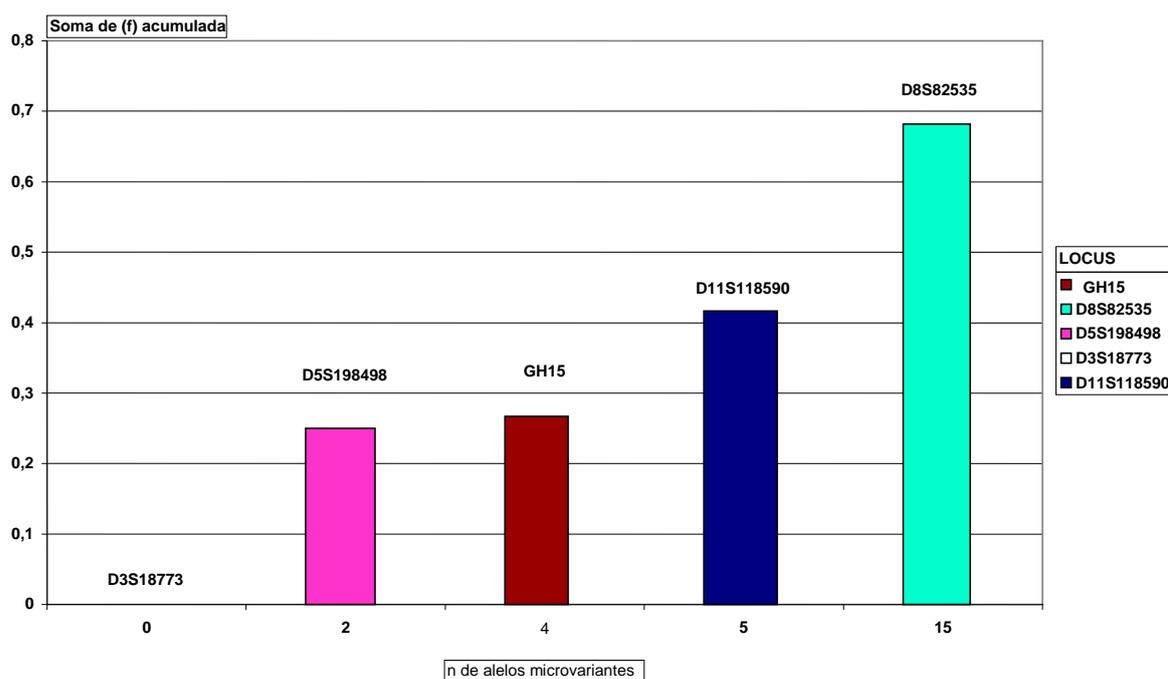


Figura 21: Percentual de microvariâncias por *locus*.

Para confirmar a existência dessas microvariâncias foram seqüenciados todos os *loci* utilizando o seqüenciador MegaBace da GE Healthcare, com amostras de DNA da população do Estado do Amazonas, os resultados confirmam as microvariâncias encontradas em cada *locus*, os seqüenciamentos encontram-se nos anexos deste trabalho.

A inclusão de alelos microvariantes na escada alélica torna-se especialmente importante quando os mesmos se apresentam com freqüências superiores à média das freqüências mínimas dos alelos observados nos estudos populacionais.

Alelos microvariantes se caracterizam por apresentar em frações de unidades das repetições diferentes em relação a um alelo. Esses alelos, com poucas exceções são raros nas populações. A possibilidade de erros de tipagem vem sendo diminuída pela inclusão de alelos microvariantes nas respectivas escadas alélicas, procedimentos feito após a verificação das freqüências alélicas nas populações (CHAKRABORTY, 1999).

Shriver e colaboradores observaram uma maior freqüência de alelos STR microvariantes em amostras de afro-descendentes. Estes resultados indicam que estes genomas teriam sofrido maior grau de mutabilidade devido ao longo processo de evolução, colaborando para a teoria da origem do homem a partir do continente africano (FRESCO e ALBERTS, 1960; LEVINSON & GUTMAN, 1987; SHRIVER *et al.*, 1997).

Livinson e Gutman sugerem que existam mecanismos de contração que garantam o controle sobre as expansões dos STR. Tais mecanismos se baseiam na ocorrência de deleções que produzem a retirada integral ou parcial de unidade de repetição. Tais mecanismos ocorrem na fase de meiose podendo resultar no aparecimento de alelos microvariantes os quais serão herdados e fixados na população (CHAKRABORTY, 1999).

5.8 COSTRUÇÃO DO BANCO DE DADOS DE ALELOS DO SISTEMA PENTAPLEX-ISED

Nesta fase foi construído um banco de dados das frequências alélicas dos 5 marcadores STR estudados, nas amostras biológicas de indivíduos das populações dos Estados do Amazonas, Rio de Janeiro e Espírito Santo, para os diferentes marcadores genéticos que compõem o sistema PentaPlex-ISED. Após a genotipagem das 308 amostras do DNA, foram estimadas as frequências alélicas de 5 *loci* STRs, das três populações estudadas, também foram calculadas a heterozigidade observada e esperada, o equilíbrio de Hardy-Weinberg e o teste exato, e outros parâmetros estatísticos comumente utilizados em teste de identificação humana. Os resultados observados estão mostrados nas Tabelas 13 a 17, onde também são apresentados dados relativos ao número de alelos observados em cada *locus*, suas frequências, número amostral (**N**), heterozigidade esperada (**He**), heterozigidade observada (**Ho**), poder de discriminação (**PD**), poder de exclusão (**PE**), conteúdo de informação de polimorfismo (**PIC**).

5.8.1 Frequências alélica do *locus* D5S198498

Os dados referentes ao *locus* D5S198498 estão apresentados na TABELA 13 e FIGURA 21, resultado da genotipagem de 308 amostras de indivíduos nas três populações estudadas.

Tabela 11: Freqüências dos alelos STR do *locus* **D5S198498**

| População – Brasil | | | |
|-----------------------------|---------------|---------------|---------------|
| Marcador : D5S198498 | | | |
| Alelos | AM | RJ | ES |
| 6 | 0,0048 | - | - |
| 9 | 0,0529 | 0,0313 | 0,0109 |
| 9.3 | 0,0577 | - | 0,0109 |
| 10 | 0,4375 | 0,5446 | 0,5000 |
| 10.3 | 0,0385 | - | - |
| 11 | 0,3798 | 0,4063 | 0,4511 |
| 12 | 0,0240 | 0,0179 | 0,0272 |
| 13 | 0,0048 | - | - |
| N | 104 | 112 | 92 |
| Intervalo | 6-13 | 9-12 | 9-12 |
| N alelos | 8 | 4 | 5 |
| He (*) | 0,673 | 0,446 | 0,511 |
| Ho (**) | 0,327 | 0,554 | 0,489 |
| PD | 0,769 | 0,715 | 0,694 |
| PE | 0,388 | 0,145 | 0,197 |
| PIC | 0,600 | 0,440 | 0,440 |

(*) Heterozigosidade esperada

(**) Heterozigosidade observada

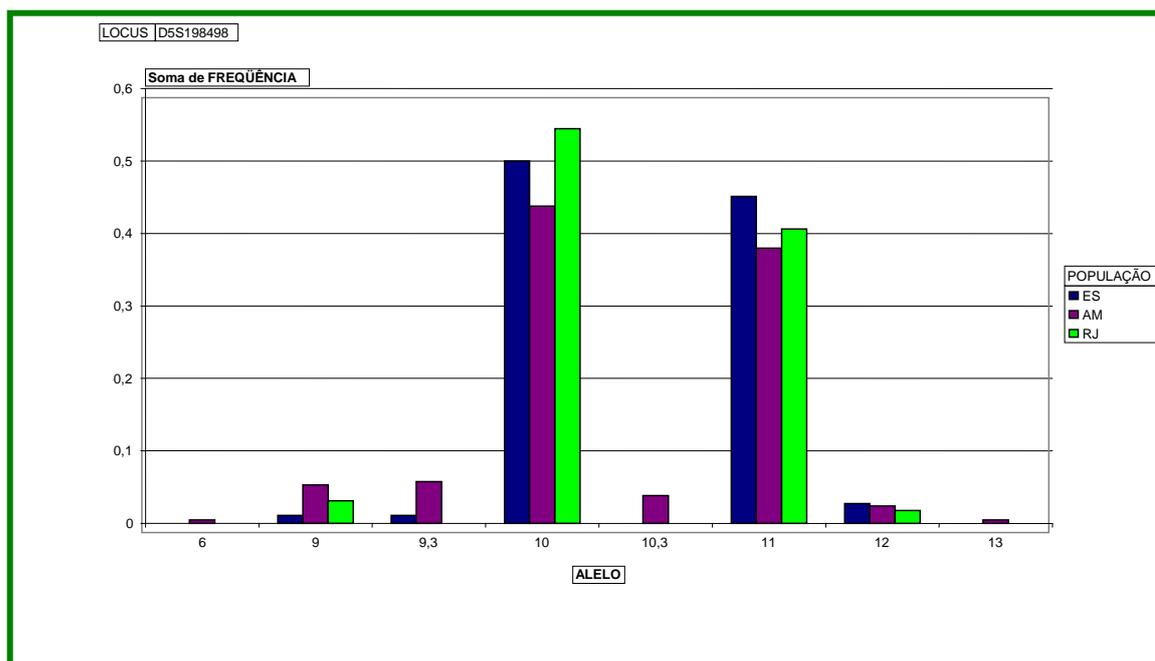


Figura 22: Gráfico das frequências dos alelos STR no locus **D5S198498** envolvendo as três populações brasileiras (AM, RJ, ES).

Foram observadas que as frequências alélicas variaram de 43,75% a 54,50%. Os valores da Heterozigosidade esperada (H_e) variam de 44,60% a 67,30% e a Heterozigosidade observada (H_o) variam de 32,70% a 55,40%. Os números de alelos segregantes foram 4, 5 e 8. O intervalo de variação do tamanho de alelos (número de repetição do STR) de 6 a 13; poder de discriminação (PD) varia de 0,694 a 0,769; poder de exclusão (PE) varia de 0,145 a 0,388; conteúdo de informação de polimorfismo (PIC) varia de 0,44 a 0,60. O alelo 10 apresentou maior frequência na população do Rio de Janeiro de (0,5446), seguido da população do Espírito Santo (0,5000) e na população do Amazonas teve a frequência menor (0,4375).

Também foram observados que os alelos (6, 10,3 e 13) não foram encontrados nas populações do: Rio de Janeiro e Espírito Santo. Supõem-se que esses alelos são raros encontrados só na população do Amazonas – AM (TABELA 13).

5.8.2 Frequências alélica do locus D11S118590

Os dados referentes ao locus D11S118590 estão apresentados na TABELA 14 e FIGURA 22, onde foram genotipadas 308 amostras de indivíduos das três populações estudadas.

Tabela 12: Frequência dos alelos STR do locus D11S118590 em três populações brasileiras.

| <i>População – Brasil</i> | | | |
|------------------------------|---------------|---------------|---------------|
| <i>Marcador : D11S118590</i> | | | |
| <i>Alelos</i> | <i>AM</i> | <i>RJ</i> | <i>ES</i> |
| 5 | 0,0048 | - | - |
| 7 | 0,0144 | 0,0402 | 0,0217 |
| 8 | 0,0385 | 0,0938 | 0,0489 |
| 8.3 | 0,0337 | 0,0134 | 0,0217 |
| 9 | 0,4183 | 0,3839 | 0,3967 |
| 9.1 | 0,0096 | - | 0,0109 |
| 9.3 | 0,0048 | - | 0,0272 |
| 10 | 0,3654 | 0,3214 | 0,3261 |
| 10.1 | - | - | 0,0109 |
| 10.3 | - | 0,0045 | 0,0272 |
| 11 | 0,0962 | 0,1295 | 0,1033 |
| 12 | 0,0144 | 0,0134 | 0,0054 |
| <i>N</i> | 104 | 112 | 92 |
| <i>Intervalo</i> | 5 – 12 | 7 – 12 | 7 -12 |
| <i>N alelos</i> | 10 | 8 | 11 |
| <i>He (*)</i> | 0,721 | 0,759 | 0,685 |
| <i>Ho (**)</i> | 0,279 | 0,241 | 0,315 |
| <i>PD</i> | 0,797 | 0,849 | 0,841 |
| <i>PE</i> | 0,462 | 0,525 | 0,405 |
| <i>PIC</i> | 0,62 | 0,68 | 0,68 |

(*) Heterozigosidade esperada

(**) Heterozigosidade observada

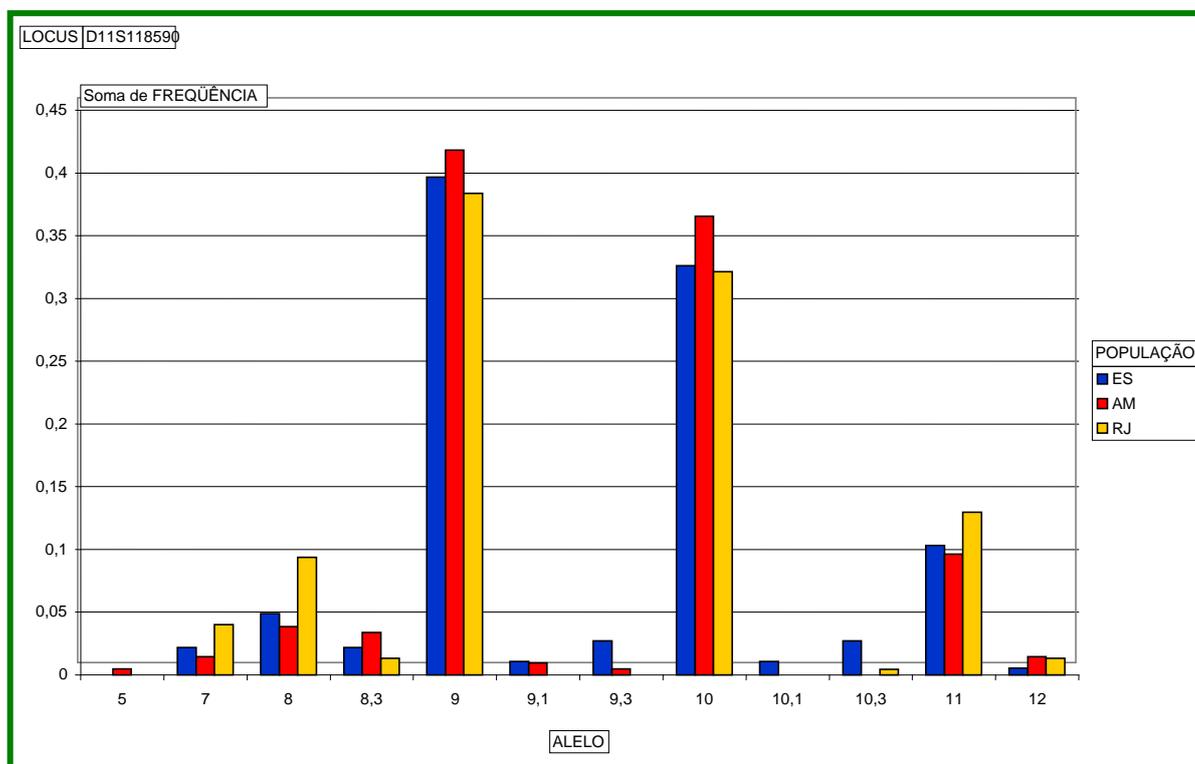


Figura 23: Gráfico da frequência dos alelos STR no locus **D11S118590** envolvendo as três populações brasileiras (AM, RJ, ES).

Pode-se observar que as frequências alélicas, variam de 38,40% a 41,80%. Os valores da Heterozigosidade esperada (H_e) variam 68,5% a 75,9% e a Heterozigosidade observada (H_o) variam 24,1% a 31,5%. Os números de alelos segregantes foram 8, 10 e 11.

O intervalo de variação do tamanho de alelos (número de repetição do STR) de 5 a 12. O poder de discriminação (PD) varia de 0,797 a 0,849; poder de exclusão (PE) varia de 0,405 a 0,525; conteúdo de informação de polimorfismo (PIC) varia de 0,62 a 0,68.

O alelo 9 apresenta maior frequência (0,4183) na população do Amazonas. Também foi observado que o alelo 5 não foi encontrado nas populações do Rio de Janeiro e Espírito Santo, supõem-se que eles são alelos raros encontrados na população do Amazonas. Como o alelo 10.1 que só foi encontrado na população do Espírito Santo.

5.8.3 Frequências alélica do *locus* GH15

Os dados referentes ao *locus* GH15 estão apresentados na TABELA 15 e FIGURA 23, onde foram genotipadas 308 amostras de indivíduos das três populações estudadas.

Tabela 13: Frequência de alelos STR do *locus* GH15 em três populações brasileiras

| População – Brasil | | | |
|--------------------|---------------|---------------|---------------|
| Marcador : GH15 | | | |
| Alelos | AM | RJ | ES |
| 4 | - | 0,0040 | - |
| 4,2 | 0,0720 | 0,0080 | 0,0430 |
| 5,2 | - | - | 0,0050 |
| 8 | 0,0050 | - | 0,0110 |
| 8,2 | 0,0050 | - | 0,0110 |
| 8,3 | - | - | 0,0050 |
| 9 | 0,4130 | 0,3750 | 0,2880 |
| 10 | 0,2400 | 0,1830 | 0,3420 |
| 11 | 0,0900 | 0,1650 | 0,1470 |
| 12 | 0,1110 | 0,1560 | 0,0870 |
| 13 | 0,0380 | 0,0710 | 0,0270 |
| 14 | 0,0100 | 0,0180 | 0,0270 |
| 15 | 0,0050 | 0,0090 | - |
| 18 | 0,0100 | - | - |
| 19 | - | - | 0,0050 |
| N | 208 | 224 | 184 |
| Intervalo | 4.2 – 18 | 4 – 15 | 4.2 – 14 |
| N alelos | 11 | 9 | 11 |
| He (*) | 0,750 | 0,679 | 0,739 |
| Ho (**) | 0,250 | 0,321 | 0,261 |
| PD | 0,899 | 0,914 | 0,905 |
| PE | 0,510 | 0,396 | 0,491 |
| PIC | 0,71 | 0,74 | 0,73 |

(*) Heterozigosidade esperada

(**) Heterozigosidade observada

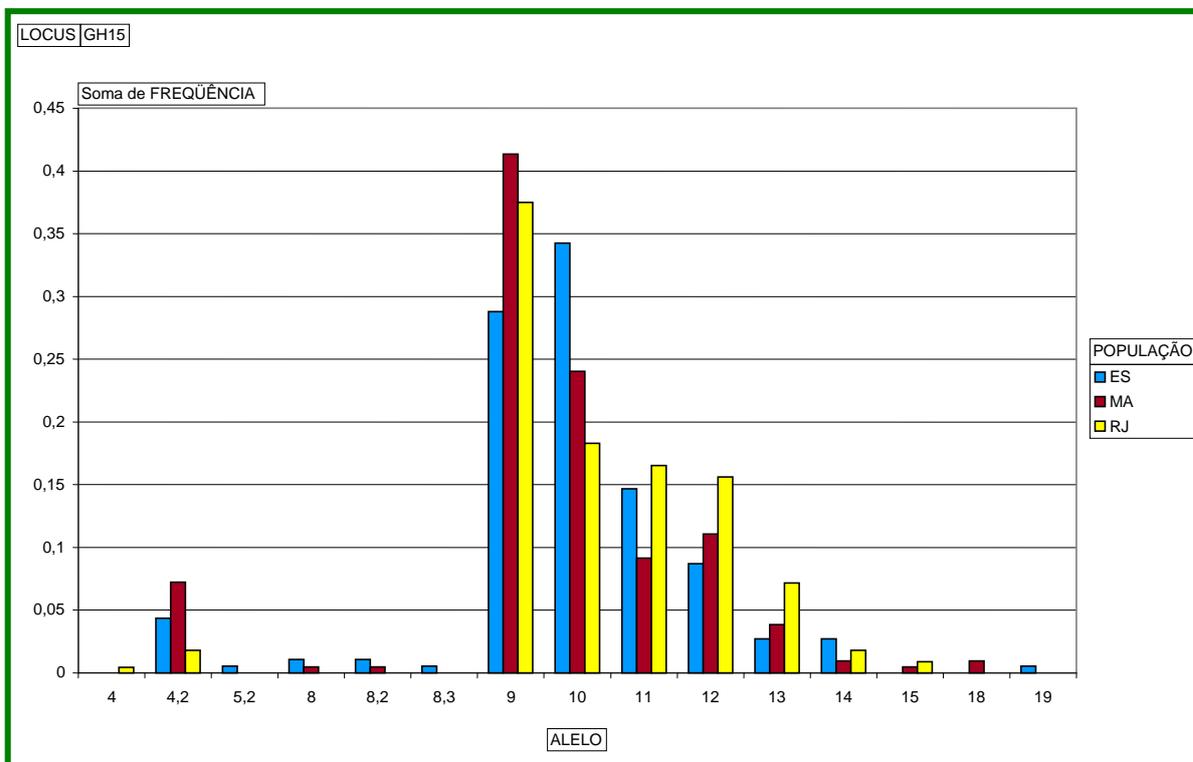


Figura 24: Gráfico das frequências dos alelos STR do *locus* **GH15** envolvendo as três populações brasileiras (AM, RJ, ES).

Verificou-se que neste *locus* as frequências alélicas variam de 0,4% a 41,30%. Os valores da Heterozigidade esperada (H_e) variam 67,9% a 75,0% e a Heterozigidade observada (H_o) variam 25,0% a 32,1%. Os números de alelos segregantes foram 9 e 11. O intervalo de variação do tamanho de alelos (número de repetição do STR) de 4 a 19.

O poder de discriminação (PD) varia de 0,899 a 0,914; poder de exclusão (PE) varia de 0,396 a 0,510; o conteúdo de informação de polimorfismo (PIC) varia de 0,73 a 0,74. O alelo 9 apresenta maior frequência (0,4130) na população do Amazonas. E o alelo 10 teve maior frequência (0,3420) na população do Espírito Santo.

Observou-se ainda que os alelos (5,2, 8,3 e 19) só foram encontrados na população do Espírito Santo, com uma frequência de 0,0050. Os alelos raros encontrados foram: 4 na

freqüência (0,0040) na população do Rio de Janeiro, o alelo 18 com a freqüência (0,0100) na população do Amazonas.

5.8.4 Freqüências alélica do *locus* D8S82535

Os dados referentes ao *locus* D8S82535 estão apresentados na TABELA 16 e FIGURA 24, onde foram genotipadas 308 amostras de indivíduos das três populações estudadas.

Tabela 14: Mostra freqüências dos alelos STR do *locus* D8S82535 em três populações brasileiras

| População – Brasil | | | |
|---------------------|---------------|---------------|---------------|
| Marcador : D8S82535 | | | |
| Alelos | AM | RJ | ES |
| 5.1 | - | - | 0,0054 |
| 6 | 0,0096 | - | - |
| 6.1 | 0,0144 | 0,0134 | 0,0217 |
| 6.2 | 0,1587 | 0,1518 | 0,1957 |
| 6.3 | 0,0048 | 0,0089 | 0,0109 |
| 7 | 0,1154 | 0,1295 | 0,1576 |
| 7.1 | - | 0,0045 | 0,0109 |
| 7.2 | 0,0096 | 0,0313 | 0,0217 |
| 8 | - | 0,0045 | 0,0054 |
| 8.1 | - | - | 0,0054 |
| 8.2 | 0,0048 | 0,0089 | 0,0109 |
| 8.3 | - | 0,0089 | - |
| 9 | 0,0433 | 0,0625 | 0,0054 |
| 9.1 | 0,0385 | 0,0223 | 0,0163 |
| 9.2 | 0,1490 | 0,1295 | 0,1141 |
| 9.3 | 0,0096 | 0,0223 | 0,0109 |
| 10 | 0,2644 | 0,2768 | 0,2989 |
| 10.1 | 0,0288 | 0,0089 | 0,0054 |
| 10.2 | 0,1010 | 0,0580 | 0,0543 |
| 11 | 0,0433 | 0,0536 | 0,0326 |
| 11.2 | 0,0048 | 0,0045 | 0,0054 |
| 12 | - | - | 0,0109 |

| N | 104 | 122 | 92 |
|------------------|----------|------------|----------|
| Intervalo | 6 – 11.2 | 6.1 – 11.2 | 5.1 – 12 |
| N alelos | 16 | 18 | 20 |
| He (*) | 0,885 | 0,786 | 0,685 |
| Ho (**) | 0,108 | 0,214 | 0,315 |
| PD | 0,950 | 0,951 | 0,940 |
| PE | 0,764 | 0,573 | 0,405 |
| PIC | 0,84 | 0,84 | 0,81 |

(*) Heterozigosidade esperada

(**) Heterozigosidade observada

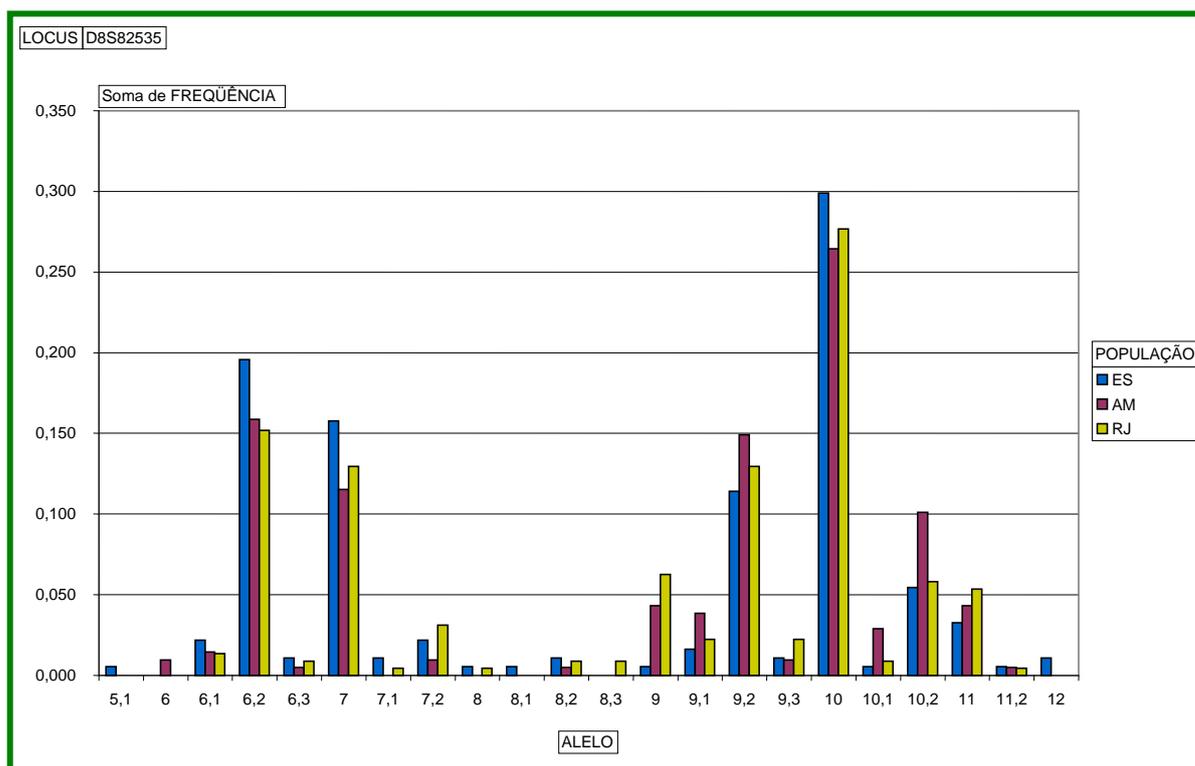


Figura 25: Mostra gráfico das frequências dos alelos STR do locus **D8S82535** envolvendo as três populações brasileiras (AM, RJ, ES).

As frequências alélicas no *locus* **D8S82535**, variam de 26,40% a 29,90%. Os valores da Heterozigosidade esperada (He) variam 68,50% a 88,50% e a Heterozigosidade observada (Ho) variam 11,50% a 31,50%. Os números de alelos segregantes foram (16, 18, 20). O intervalo de variação do tamanho de alelos (número de repetição do STR) de 5,1 a 12.

O poder de discriminação (PD) varia de 0,940 a 0,951; poder de exclusão (PE) varia de 0,405 a 0,764; o conteúdo de informação de polimorfismo (PIC) varia de 0,81 a 0,84.

O alelo 10 apresenta maior frequência (0,2989) na população do Espírito Santo. Também foi observado que os alelos (5.1, 8.1 e 12) não foram tipados nas populações do Rio de Janeiro e do Amazonas, portanto eles são alelos raros encontrados somente na população de Espírito Santo. Assim como o alelo (8.3) considerado raro por ser tipado só na população do Rio de Janeiro. Este *locus* foi o que apresentou maior número de microvariancia, sendo muito polimórfico, em relação aos outros estudados neste trabalho.

5.8.5 Frequências alélica do *locus* D3S18773

Os dados referentes ao *locus* D3S18773 estão apresentados na TABELA 17 e FIGURA 25, onde foram genotipadas 308 amostras de indivíduos das três populações estudadas.

Tabela 15: Mostra frequência de alelos STR do *locus* **D3S18773** em três populações brasileiras:

| População – Brasil | | | |
|----------------------------|---------------|---------------|---------------|
| Marcador : D3S18773 | | | |
| Alelos | MA | RJ | ES |
| 11 | 0,0481 | 0,0625 | 0,0326 |
| 12 | 0,2308 | 0,1116 | 0,1250 |
| 13 | 0,4327 | 0,4509 | 0,5326 |
| 14 | 0,2596 | 0,3527 | 0,2935 |
| 15 | 0,0240 | 0,0223 | 0,0054 |
| 16 | 0,0048 | - | 0,0109 |
| N | 104 | 122 | 92 |
| Intervalo | 11 – 16 | 11 – 15 | 11 – 16 |
| N alelos | 6 | 5 | 6 |
| He (*) | 0,625 | 0,625 | 0,511 |
| Ho (**) | 0,375 | 0,375 | 0,489 |
| PD | 0,841 | 0,822 | 0,801 |
| PE | 0,322 | 0,322 | 0,197 |
| PIC | 0,64 | 0,59 | 0,55 |

(*) Heterozigosidade esperada

(**) Heterozigosidade observada

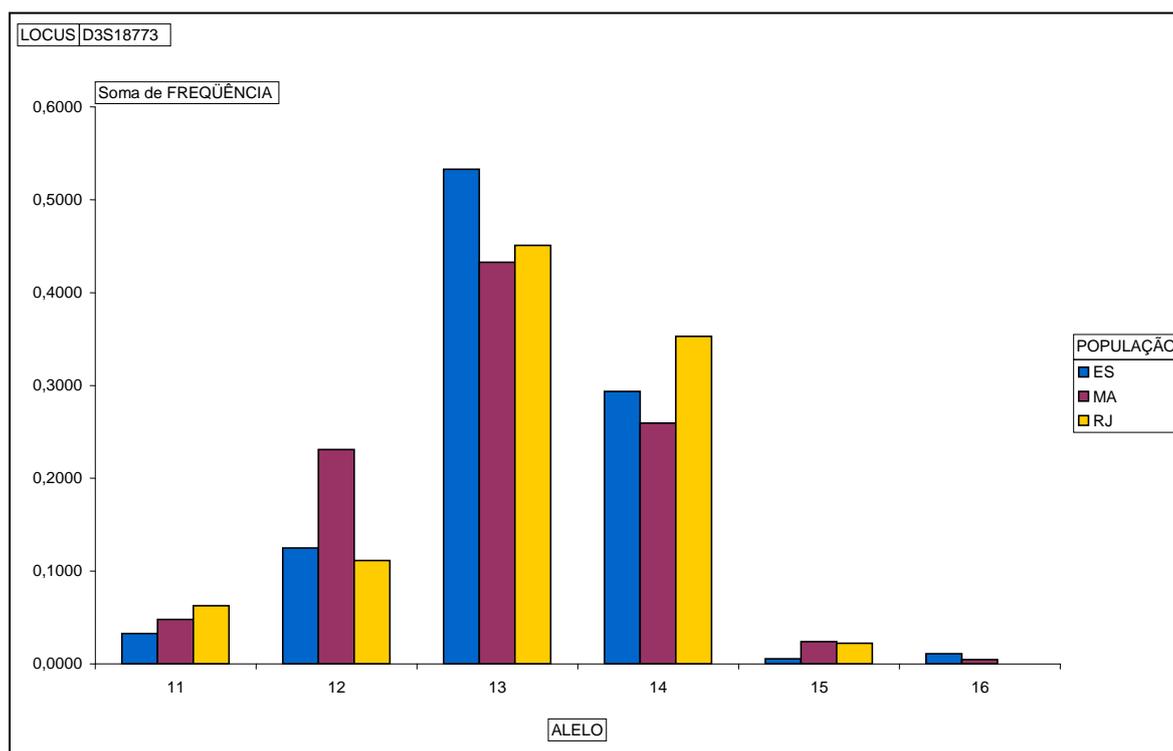


Figura 26: Mostra gráfico da frequência dos alelos STR do locus **D3S18773** envolvendo as três populações brasileiras (AM, RJ, ES).

Neste *locus* observou-se que as frequências alélicas variam de 43,30% a 53,30%. Os valores da heterozigosidade esperada (H_e) variam 68,50% a 75,90% e a heterozigosidade observada (H_o) variam 24,19% a 31,50%. Os números de alelos segregantes foram 6 e 5. O intervalo de variação do tamanho de alelos (número de repetição do STR) de 11 a 16.

O poder de discriminação (PD) varia de 0,801 a 0,841, o poder de exclusão (PE) varia de 0,197 a 0,322. O conteúdo de informação de polimorfismo (PIC) varia de 0,55 a 0,64. Observou-se também neste marcador que o alelo 13 apresentou maior frequência de (0,5326) na população de Espírito Santo. E o alelo 16 só foi encontrado nas populações: Amazonas e Espírito Santo. Este *locus* (D3S18773) foi o único que não apresentou alelos microvariantes, e possui o menor grau de polimorfismo.

As análises das frequências alélicas dos 5 *loci* STR, observou-se existir total independências entre a transmissão dos alelos. Na maioria dos alelos apresentaram frequências próximas ou abaixo de 25%, evidenciando que estes *loci* são polimórficos e bem distribuídos. Segundo Miscika-Sliwka (1997), esta é uma tendência geral nas regiões repetitivas STRs, nas quais o acúmulo de deleções ou substituições é considerado um mecanismo para prevenir a expansão ou contração externa, funcionando como um controle do polimorfismo dessas regiões.

No nosso banco de dados, os *loci* D11S118590, GH15, D8S82535, foram vistos como mais polimórficos apresentam variação do tamanho de alelos nas populações de 5 – 12, 4 – 19, 5.1 - 12, respectivamente. Também verificou-se que o *locus* D3S18773, apresentou o menor tamanho de alelos nas populações de 11 - 16.

5.9 PARÂMETROS ESTATÍSTICOS DE INTERESSE FORENSE E DE VÍNCULO GENÉTICO COM BASE NOS 5 *LOCI* STR

Os parâmetros estatísticos forenses foram calculados com base nas frequências alélica dos 5 *loci* STR que estão incluídos no sistema PentaPlex-ISED, utilizando-se como base de dados as frequências alélicas das três populações estudadas.

Os resultados estão mostrados na TABELA 18, onde foram estimados os valores do poder de discriminação (PD) e o poder de exclusão (PE) para os 5 *loci* STR.

Foram observados que os parâmetros estatísticos para *locus* D5S198498, apresentaram valores memores como: poder de discriminação (PD) igual 72,60% e o poder de exclusão (PE) em torno de 25%. Enquanto que para o *locus* D8S82535, os valores estatísticos foram mais altos, onde poder de exclusão (PE) igual a 58,1% e poder de discriminação (PD) 94,7%.

Tabela 16: Poder de Discriminação (PD), Poder de Exclusão (PE) e o Conteúdo de Informação de Polimorfismo (PIC) para os 5 *loci* STR que compõem o sistema multiplex PentaPlex-ISED.

| <i>Loci</i> | PD | PE | PIC |
|-------------------|-------|-------|-------|
| D5S198498 | 0,726 | 0,243 | 0,493 |
| D11S118590 | 0,829 | 0,464 | 0,660 |
| GH15 | 0,906 | 0,465 | 0,726 |
| D8S82535 | 0,947 | 0,580 | 0,830 |
| D3S18773 | 0,821 | 0,280 | 0,593 |

Os dados estatísticos realizados tais como poder de discriminação (PD), poder de exclusão (PE) e o conteúdo de informação polimórfica (PIC) foram analisados no *software* “PowerStar”. Os resultados observados foram: o *locus* D8S82535 apresentou o maior poder de discriminação (0,951), maior poder de exclusão (0,764) e o maior conteúdo de informação do polimorfismo (0,840) para as três populações analisadas. O *locus* D5S198498 apresentou o menor poder de exclusão (0,145), menor poder de discriminação (0,694) e o menor conteúdo de informação do polimorfismo (0,440).

O Poder de Discriminação é obtido a partir dos quadrados das frequências esperadas, assim como para o Poder de Exclusão, conclui-se que o Poder de Discriminação é proporcional às frequências alélicas. Logo, os *loci* que apresentaram alelos com baixa frequência possuem os maiores valores de Poder de Discriminação (GH15 – 90%; D8S82535 – 94%. Os *loci* que apresentaram alelos com alta frequência possuem os menores valores de Discriminação (D5S198498 – 72,6%; D3S18773 – 82,1%; D11S118590 – 82,9%. Observou-

se que os valores para cada locus são bem próximo nas três diferentes populações, donde podemos concluir que as frequências alélicas também são muito semelhantes.

Os valores do Poder de Exclusão Acumulado foi 93,436%, e o Poder de Discriminação 99,996%, para os 5 *loci*. Esses resultados mostram que esses *loci* STR são passíveis de serem utilizados, em conjunto com outros microssatélites a serem selecionados, em análise genética forense, comparando com outros estudos realizados utilizando marcadores STRs, com amplificação simultânea com até 15 *loci* verificada pela Applied Biosystems, trabalhos de validação de 14 *loci* STR realizado por SILVA (2004), e também trabalho de validação do sistema LDD 10 por MANTA (2008).

5.9.1 Teste de independência em um *locus* (Equilíbrio de Hardy-Weinberg)

Para que um marcador genético possa ser utilizado na identificação humana, além do polimorfismo, é necessário avaliar o nível de transmissão de seus alelos na população. No caso de marcadores autossômicos, se pressupõe transmissão independente, permitindo que a frequência de um perfil genético seja calculada segundo a lei de Hardy-Weinberg.

Nesta etapa, com objetivo de verificar a aderência das frequências genótípicas observadas às proporções teóricas de Hardy-Weinberg foi verificada com o emprego do programa (Guo, S. & Thompson, e. 1992) observados nas três populações brasileiras estão de acordo com as proporções genótípicas esperadas para uma população no equilíbrio de Hardy-Weinberg. O nível de significância para cada teste foi determinado pela proporção de vezes que o genótipo esperado excedia o genótipo observado. O teste definido segundo a simulação em 10.000 permutações dos genótipos encontrados na população executadas por software Arlequin 3.1 (Excoffier et al., 2005) onde os resultados estão apresentados nas TABELAS 19, 20, 21.

Observou-se que os testes realizados com as populações alguns *loci* apresentaram valores de p maiores que 0,05, mostram que as proporções alélicas e genotípicas observadas para cada *locus* analisado encontram-se em conformidade com as proporções genotípicas estabelecidas no equilíbrio de Hardy-Weinberg.

Tabela 17: Teste para o equilíbrio de Hardy-Weinberg, nos 5 *loci* STR na população do Amazonas.

| POPULAÇÃO DO AMAZONAS (AM) | | | | |
|-----------------------------------|-----------|----------------------------|---------------------------|-------------------------|
| <i>Loci</i> | Genótipos | Heterozigosidade Observada | Heterozigosidade Esperada | Valor de ($P < 0,05$) |
| D5S198498 | 104 | 0.67308 | 0.65928 | 0.00000 |
| D11S118590 | 104 | 0.72115 | 0.68241 | 0.00273 |
| GH15 | 104 | 0.75000 | 0.74735 | 0.73147 |
| D8S82535 | 104 | 0.88462 | 0.85670 | 0.07491 |
| D3S18773 | 104 | 0.62500 | 0.69254 | 0.00003 |

De acordo com os resultados mostrados na TABELA 19, a população do Amazonas em dois *loci* GH15 e D8S82535, apresentou valor de p (0.73147; 0.07491), respectivamente, indicando que esses dois *loci* estão de acordo com as proporções genotípicas esperadas para uma população no equilíbrio de Hardy-Weinberg, caracterizando independência dos alelos para os *loci* analisados. Porém os valores de p para os *loci* D5S198498, D11S118590 e D3S18773, revelaram não aderir o equilíbrio de HWE.

Tabela 18: Teste para o equilíbrio de Hardy-Weinberg, nos 5 *loci* STR, na população do Rio de Janeiro.

| POPULAÇÃO DO RIO DE JANEIRO (RJ) | | | | |
|---|-----------|----------------------------|---------------------------|-------------------------|
| <i>Loci</i> | Genótipos | Heterozigosidade Observada | Heterozigosidade Esperada | Valor de ($P < 0,05$) |
| D5S198498 | 112 | 0.44643 | 0.53944 | 0.26853 |
| D11S118590 | 112 | 0.75893 | 0.72498 | 0.02220 |
| GH15 | 112 | 0.67857 | 0.77178 | 0.02994 |
| D8S82535 | 112 | 0.78571 | 0.85798 | 0.00000 |
| D3S18773 | 112 | 0.62500 | 0.65839 | 0.00125 |

Os resultados observados na tabela 20, a população do Rio de Janeiro apenas no *locus* D5S198498 está de acordo com as proporções genotípicas esperadas para uma população no equilíbrio de Hardy-Weinberg, os outros *loci* não aderir o equilíbrio de HWE. Para a população do Rio de Janeiro embora existam desvios do equilíbrio de Hardy-Weinberg em quatro *loci*, como já verificado, os mesmos não parecem ser significativos. Além disso, os desvios observados encontram se próximos dos limites de significância de 5% e de 1%.

Tabela 19: Teste para o equilíbrio de Hardy-Weinberg, nos 5 *loci* STR, na população do Espírito Santo.

| POPULAÇÃO DO ESPÍRITO SANTO (ES) | | | | |
|---|-----------|----------------------------|---------------------------|-------------------------|
| <i>Loci</i> | Genótipos | Heterozigosidade Observada | Heterozigosidade Esperada | Valor de ($P < 0,05$) |
| D5S198498 | 92 | 0.51087 | 0.54853 | 0.08422 |
| D11S118590 | 92 | 0.68478 | 0.72446 | 0.00000 |
| GH15 | 92 | 0.73913 | 0.77120 | 0.05983 |
| D8S82535 | 92 | 0.68478 | 0.83304 | 0.00000 |
| D3S18773 | 92 | 0.51087 | 0.61671 | 0.23348 |

Nas amostras da população do Espírito Santo (TABELA 21), a condição de equilíbrio HWE foi observada em três *loci* D5S198498, GH15 e D3S18773, com os valores de p (0.08422; 0.05983; 0.23348), respectivamente, enquanto os outros *loci* D11S118590 e D8S82535 observaram o desequilíbrio segundo o teorema de HWE.

Podem-se observar com essas análises os *loci* GH15 e D5S198498 estão em equilíbrio na população do Espírito Santo correspondem também equilíbrio na fração amostral do Amazonas o *locus* (GH15) e em equilíbrio na fração amostral do Rio de Janeiro o *locus* (D5S198498), e entre as duas frações amostral Rio de Janeiro e Amazonas não apresentaram nenhum *locus* correspondente isto é, em equilíbrio de HWE, sugerindo que estão geneticamente distantes. O *locus* D11S118590 apresentou desequilíbrio de HWE nas três populações em estudo, com valores de $p=0.00273$, $p=0.02220$ e $p=0.00000$, respectivamente (AM, RJ e ES). Essas diferenças podem ser levadas em conta ao fluxo gênico, o grau de extinção de alelos e migração das populações, que pode estar inferindo a cerca dos fatores que podem estar relacionados com o desvio do Equilíbrio de Hardy-Weinberg (EHW) (ALLENDORF & LUIKART, 2007).

Neste estudo foi possível observar que a maioria dos *loci* STR do sistema PentaPlex-ISED, apresentados nos 5 *loci*, podem ser atribuídos às premissas não obedecidas numa estabilidade genética, isto acontece quando a população está sujeita aos fatores evolutivos, que são capazes de alterar as frequências gênicas como (mutação, seleção natural, fluxo gênico e deriva genética). Apesar dessa não estar em equilíbrio, às populações se beneficiam dos efeitos dos fatores evolutivos, podendo buscar novos pontos de equilíbrio gênico, a fim de manter a estabilidade das frequências genotípicas (BEIGUELMAN, 1994).

5.10 COMPARAÇÃO POPULACIONAL

Os historiadores explicam que a população brasileira é uma mistura étnica entre três diferentes continentes (europeus, africanos, e ameríndios). De acordo com os nossos resultados observados a população do Amazonas apresentou alguns alelos diferentes (raros) em relação às outras populações estudadas destacando-se das outras etnias, isso confirma que os índios que habitavam as terras brasileiras e a contribuição dos imigrantes foi gerando um “pool” genético na população brasileira, que ao longo de mais de 500 anos de povoamento, que se espalhou entre cidades e regiões (IBGE, 2000). O índio, acostumado a uma vida livre e com valores culturais voltados pra sua coletividade, com o passar dos tempos sofreu uma diminuição abrupta de sua representação no território, isolando-se em regiões afastadas.

Com o avanço das técnicas da Biologia Molecular, oferecem suporte as teorias de historiadores e antropólogos a cerca da formação do povo brasileiro e da identidade étnica deste povo visando a resguardar valores culturais.

Explicam os estudiosos que, o que mais contribuiu para a predominância de uma alta taxa de miscigenação verificada no Rio de Janeiro desde o século XVIII (Martis & Abreu, 2001), foi o fato de a cidade ter sido um dos principais portos de entrada para o tráfico de africanos, e o fato de ter sido por vinte anos a capital do império.

Neste trabalho, observou-se que as três populações não apresentaram diferença significativa populacional, todas as populações estudadas apresentaram grande miscigenação. Uma parte substancial dos colonizadores portugueses cruzarem-se com índios e africanos, em um processo muito importante para a formação do povo brasileiro.

5.10.1 Distanciamento genético entre populações

As análises de comparações de distanciamento entre as populações: Amazonas, Rio de Janeiro e Espírito Santo, o distanciamento geografico mostrado na FIGURA 26, verificou-

se que a distância geográfica não afeta significativamente o fluxo gênico. Por outro lado, os fatores étnico e cultural pareceram ser mais significativos na oscilação do fluxo gênico.

A comparação da população do Amazonas com as populações do Rio de Janeiro e Espírito Santo, como mostra na FIGURA 26 (mapa de localização das 3 cidades brasileiras), foram realizadas por meio do software *Arlequin*, programa (AMOVA), para estimar a heterozigosidade média para cada população, As medidas do efeito da subdivisão populacional, que representa a redução da heterozigosidade em uma subpopulação e avalia a divergência genética total entre subpopulações foram estimadas através dos coeficientes de diferenciação genética F_{ST} (Wright, 1965) e R_{ST} (Slatkin, 1995) bem como estimativas de diversidade genética que medem a distância genética entre essas populações.



Figura 27: Mapa do Brasil e a localização geográfica das três populações brasileiras (Amazonas, Rio de Janeiro e Espírito Santo).

Observou-se a hipótese de isolamento por distância por via teste de Mantel onde foram estimados os valores de $\Phi_{ST}/(1-\Phi_{ST})$ (Fetzner *et al.*, 2003). Os resultados obtidos pelo teste de Mantel foram ($r = 0.5929$; $p = 0.1630$). Esse resultado vem confirmar a distância geográfica que pode estar interferindo no fluxo gênico. Neste contexto, podem ser observadas algumas diferenças regionais em relação a alguns fatores como: étnicos, lingüísticos e até mesmo as frequências alélicas.

Na análise da estruturação genética, foi realizado cálculo do índice R_{ST} (*stepwise mutation model*), comparando objetivamente o efeito total da estrutura da população, levando-se em consideração as variações das frequências alélicas e a diversidade gênica. Nas tabelas 20 e 21, mostram os resultados dessas análises, onde as populações foram comparadas entre si, utilizando os 5 *loci* incluídos no multiplex PentaPlex-ISED. Observou-se uma estruturação nas populações do Amazonas, Rio de Janeiro e Espírito Santo.

Em relação à variância molecular (AMOVA), as populações estudadas foram agrupadas e comparadas par-a-par, realizados 10.000 passos da Cadeia de Markov. Essa análise mostrou que a variação genética é maior dentro das populações (98,89 %) e os valores de F_{ST} e $p = 0.03030 \pm 0.00562$. Este resultado evidencia que não esteja havendo uma subestrutura populacional, o que pode estar ocorrendo é uma estruturação populacional em decorrência de fluxo gênico.

Podemos também observar nos resultados apresentados na tabela 21, que a população do estado do Espírito Santo, geneticamente, está mais distante da população do estado do Amazonas onde ($p=0.0791 \pm 0.0078$). Essas diferenças podem ser decorrentes da mistura dos três grupos étnicos (ameríndios, negros e europeus) ocorrido com a expansão e domínio territorial pelos europeus. Como por exemplo, o *locus* D5S198498 apresentou 8 alelos, dentre estes, três só foram encontrados na população do Amazonas. Observou-se que no *locus*

D8S82535, os alelos 5.1 e 12 só foram tipados nas amostras da população do Espírito Santo.

No locus D11S118590, o alelo 5 apresenta frequência na amostra da população do Amazonas.

Tabela 20: Valores de F_{ST} o método de distância nos diferentes alelos, nas amostras das três populações brasileira (AM, RJ e ES) para 5 *loci* STR.

| | AM | RJ | ES |
|----|---------|---------|----|
| AM | 0 | | |
| RJ | 0.01769 | 0 | |
| ES | 0.00587 | 0.00688 | 0 |

O método AMOVA é baseado na frequência alélica e leva em consideração o número de unidades de repetições por *locus*, estimando a diferenciação populacional diretamente de dados moleculares (EXCOFFIER *et. al.*,1992).

Tabela 21: Estimativas de F_{ST} e correspondentes valores de p dos resultados da Análise de Variância Molecular (AMOVA) de amostras populacionais (AM, RJ e ES).

| Número de Permutação: 1023 | AM | RJ | ES |
|----------------------------|----------------|----------------|----|
| AM | * | | |
| RJ | 0.00391±0.0019 | * | |
| ES | 0.0791±0.0078 | 0.07129±0.0075 | * |

A diversidade inter-população média para cada população apresentadas na TABELA 21, pode-se observar a existência da diversidade genética entre as três populações. Embora a diferença seja significativa para a população do Amazonas onde ($p=0,003$; $p \geq 0,05$). O percentual da variação obtido entre as populações de cada estado brasileiro,

quando foram analisadas entre grupos, foi $\leq 1,03\%$, enquanto que acima de 98% se deve a diferença inter-populacional. Esse resultado vem corroborar para uma diversidade gênica significativa no contexto da biologia.

Para corroborar neste estudo, a variabilidade genética da população indígena é organizada de forma complexa sendo que dois fatores que mais exercem influência sobre a mesma é a variação lingüística e geográfica (Fagundes *et al.*, 2002). Os efeitos exercidos pela geografia são daqueles exercidos pela língua na variabilidade genética. O fluxo gênico entre as populações estudadas estaria corroborando a variabilidade genética.

6 CONCLUSÕES

Os resultados alcançados neste trabalho permitiram estabelecer diferentes conclusões:

- Foi desenvolvido um novo sistema multiplex denominado PentaPlex-ISED.
- O sistema PentaPlex-ISED permitiu a amplificação simultânea de cinco marcadores STR numa única reação de PCR.
- Foi construído um Banco de Dados das frequências alélicas para cada *loci*, permitindo a realização das análises genéticas populacionais.
- Mostrou-se que alguns *loci* não estão em Equilíbrio de Hardy-Weinberg nas populações estudadas e que as três populações são geneticamente diferenciadas, justificando a identificação de marcadores moleculares para a população do estado do Amazonas.
- O novo sistema PentaPlex-ISED para identificação humana, mostrou-se eficiente na tipagem de diferentes amostras biológicas.
- O sistema multiplex PentaPlex-ISED apresentou o poder de exclusão acumulado ($PE_{(A)}$) de 93,435 %, e o poder de discriminação acumulado ($PD_{(A)}$) em 99,995 %, índices aceitáveis para perícia forense que permite distinguir um indivíduo do outro.

Os resultados alcançados neste trabalho permitiram estabelecer diferentes conclusões:

- O sistema PentaPlex-ISED permitiu a amplificação simultânea de 5 marcadores STR numa única PCR. A utilização desse sistema facilitou a genotipagem de número expressivo de indivíduos e possibilitou a caracterização genética de três populações brasileiras (Amazonas, Rio de Janeiro e Espírito Santo), de forma rápida e utilizando pouca quantidade de DNA.
- O sistema multiplex PentaPlex-ISED mostrou-se eficiente na tipagem de diferentes amostras biológicas, sensível, reprodutível e com baixa frequência de alelos microvariantes.

7 CONSIDERAÇÕES FINAIS

O sistema PentaPlex-ISED, analisado neste trabalho, composto por 5 *loci* STR e mais a Amelogenina apresentou-se todos os requisitos necessários para a sua aplicação na identificação genética, permitindo uma amplificação simultânea utilizando quantidade mínima de DNA, os produtos de PCR foram separados em eletroforese capilar permitindo análise rápida e tipagem confiáveis.

Este novo sistema PentaPlex-ISED pode ser desenvolvido em laboratórios para análises de identificação humana, pode ser um excelente complemento para os testes de paternidade na exclusão. Para a investigação de paternidade pode ser aumentando o número de *loci* nos casos de inclusão, aumentando os valores do poder de discriminação (PD) em até 99,99999999%, gerando uma expressiva economia para laboratórios de identificação humana.

REFERÊNCIAS

AMERICAN ASSOCIATION OF BLOOD BANKS (AABB). **Paternity Testing. Seminar presented by Committee on Technical Workshops of AABB**, New Orleans, H. Silver (ed), Washington, D.C. 1989.

_____. **Parentage testing accreditation requirements manual**. Glenbrook Road, Bethesda, 3 ed. (1998).

_____. **Guidance for Standards for Parentage Testing Laboratories**. 4 th ed. Bethesda, Maryland, USA. 2000.

ALLENDORF, F. W & LUIKART, G. **Conservation and the genetics of populations**. Blackwell Publishing, 2007.

ALLEN, C. R.; GRAVES, G.; BUDOWLW, B. **Polymerase Chain Reaction Amplification Products Separated on Rehydratable Polycrylamide Gels and Stained with Silver**. Bio Techniques. 7(7): 736-744. 1989.

ANDERSON, S.; BANKIER, A.; BARRELL, B. **Sequence and organization of the human mitochondrial genome**. Nature. 290(56): 457-65. 1981.

ANDREWS, R.M.; KUBACK, I.; CHINNERY, P.F.; LIGHTOWLERS, R.N.; TURNBULL, D.M.; HOWELL, N. **Reanalysis and revision of the Cambridge Reference Sequence for human mitochondrial DNA**. Nature Genet 23 (1999) 147.

APPLIED BIOSYSTEMS, **AmpFISTR Identifiler PCR**. Amplification Kit User`s Manual, 2001 Foster City, CA, USA.

_____. **Available from**. <http://www.appliedbiosystems.com>. 2006.

AREZI, B.; XING, W.; SORGE, J.A.; HOGREFE, H. **Amplification efficiency of thermostable DNA polymerases**. Analytical Biochemistry 2003; 321:226-235.

BALAZS, I.; BAIRD, M.; Mc ELFRESH, K. **Hardy-Weinberg equilibrium analysis of DNA polymorphic loci that exhibit a continuous distribution of fragment sizes**. In H. Polesky and W. Mayr. Eds. *Advances in Forensic Haemogenetics 3*. New York: Springer-Verlag. P: 71-74.1990.

BAR, W. **International Journal of Legal Medicina**. Nº 107, p 159 –160 (1994).

_____. **International Journal of Legal Medicina**. V. 110, p 175-176 (1997).

BARRETO, C. *O povoamento e População: política populacional brasileira*. 2a. edição Rio de Janeiro. J. Olympo, 1959

BENCHIMOL. S. *Amazônia: Um pouco – Antes e Além – Depois*. Editora Manaus: Umberto Calderaro, 247-251.1977

BEIGUELMAN, B. **Dinâmica dos genes nas famílias e nas populações. Ribeirão Preto**, Ed. Sociedade Brasileira de Genética, 1994.

BLAKE, E.; MIHALOVICH, A.; HIGUCHI, R. **Polymerase chain reaction (PCR) amplification and human leukocyte antigen (HLA)**. Do a oligonucleotide typing on biological evidence samples: Casework Experience. *J Forensic Sci.* 37:700-726. 1992.

BODY POLITIC. **Alphonse Bertillon: his measuring system**. <http://www.cimm.jcu.edu.au/hist/stats/bert/illust.htm>. 2003.

BOTSTEIN, D.; WHITE, R.L.; SKOLNICK, M.; DAVIS, R.W. **Construction of a genetic linkage map in man using restriction fragments length polymorphisms**. *Am.J.hum.Genet.* 1980; 32: 314-331.

BRETTELL, T.A.; BUTLER, J.M.; SAFRERSTEIN, R. **Forensic Science**. *Anal. Chem.*, nº 77, p. 3839-3860. 2005.

BODY POLITIC. **Alphonse Bertillon: his measuring system**. <http://www.cimm.jcu.edu.au/hist/stats/bert/illust.htm>. 2003.

BUDOWLE, B.; CHAKRABORTY, R.; CARMODY, G. **Source attribution of a Forensic DNA Profile**. *Forensic Science Communications*. 2(3):1-6. 2000.

_____; MORETTI, T.R.; KEYS, K.M. **Validation studies of the CTT STR multiplex system**. *J. Forensic Sci.* 42: 701-707. 1997.

_____; MORETTI, BROW, B. L. **CODIS and PCR-Based short tandem repeat loci: law enforcement tool**. Laboratory Divisin of the Federal Bureau of Investigation. Virginia. Publication number 98-06. 1998.

BUTLER, J.M. **Forensic DNA typing. Biology and technology behind STR markers**. Academic Press, San Diego. 2001.

_____. **Methods in molecular biology: Forensic DNA typing protocols** Humana Press: Totowa. New Jersey; 2003.

_____. **Forensic DNA typing. Biology and technology behind STR markers**. Burlington, London Elsevier Academic Press. 2005a.

_____. **Constructing STR multiplex assays. Methods in Molecular Biology**. 2005b; 297: 53-66.

_____. **Genetics and Genomics of core Short Tandem Repeat Loci Used in Human Identity Testing**. *J Forensic Sci.* 2006; 51: 253-265.

CHAKRABORTY, R. **Quantitative traits in relation to population structure: why and how are they used and what do they imply?** *Hum Biol.* 62(1):147-62. 1990.

COBLE, M.D.; BUTLER, J.M. **Characterização of New MiniSTR Loci to Aid (Analysis of Degraded DNA)**. J Forensic Sci, v.50, n. 01, jan. 2005.

COLLINS, P.J. ; HENNESSY, L.K. ; LEIBELT, C.S.; ROBY, R.K.; REEDER, D.J.; FOXALL, P.A. **Developmental validation of a Single-Tube amplification of the 13 CODIS STR loci, D2S1338, D19S433, and Amelogenin : The AmpF1STR Identifier PCR Amplification Kit**. J.Forensic Sci, 2004, n. 49(6):1265-1277.

CONNOR, A.; STONEKING, M. **Assessing Ethnicity from Human Mitochondrial DNA Types Determined by Hybridization with Sequence-Specific Oligonucleotides**. J. Forensic Sci. 39(6):1360-1371. 1994.

COMEY, J.E. **DNA extraction**. J. Forensic Sci. p: 1261-1269. 1995.

CORACH, D.; SALA, A.; PENACINO, G.; SOTELO, A. **Mass disasters: rapid molecular screening of human remains by means of short tandem repeats typing**. Electrophoresis. 1995. Sep; 16(9): 1617-23.

CRESTE, S. ; TULMANN NETO, A. ; FIGUEIRA, A. **Detection of single sequence repeat polymorphisms in denaturing polyacrylamide sequencing gels by silver staining**. Plant Molecular Biology Reporter, v. 19, p. 299-306, 2001.

CROUSE, C. A., ROGERS, S., AMIOTT, E. **Analysis and Interpretation of Short Tandem Repeat Microvariants and Three-Banded Allele Patterns Using Multiple Allele Detection Systems**. J. Forensic Sci. 44(1):87-94. 1999.

CHUNG, D.T. ; DRÁBEK, J. ; OPEL, K.L.; BUTLER, J.M.; McCORD, B.R. **A study on the effects of degradation and template concentration on the amplification efficiency of the STR Miniplex Primer Sets**. J. Forensic Sci., v. 34, n 4, jul. 2004.

CHAKRABORTY, R.; SHAW, M.; SCHULL, S. **Exclusion of Paternity: The current state of the art**. Am.J.Hum. Genet. 26:477-488. 1974.

_____; STIVERS DN.; SU B.; ZOHONG B. **The utility of short tandem repeat loci beyond human identification: Implications for development of new DNA typing systems**. Electrophoresis, 1999; 20: 1682-1696.

CHAMBERLAIN, J.S; GIBBS, R.A; RANIER, J.E ; NQUVEN, P.N; CASKEY, C.T. 1988. **Deletion screening of the Duchenne muscular dystrophy locus via multiplex DNA amplification**. Institute for Molecular Genetics, Baylor College of Medicina, Houston, TX 77030.

DANIELS, D.L.; HALL, A.M.; BALLANTYNE, J. **SWGDAM Developmental Validation of a 19-Locus Y-STR System for Forensic Casework**. J.Forensic Sci. 2004; 49(4):668-83.

DECORTE, R.; CUPPENS, H.; MAYNEM, P.; CASSIMAN, J.J. 1990. **Rapid Detection of Hypervariable Regions by the Polymerase Chain Reaction Techique**. DNA and cell Biology. 9(6): 461-469.

EXCOFFIER, L.; SMOUSE, P.E.; QUATTRO, J.M. **Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data.** *Genetics*. 131: 479-91.1992

_____. **Analýsis of Population Subdivision.** In: Balding d, Bishop M, Cannings C, editors. *Handbook of Statistical Genetics*, 2nd Edition. New York: John Wiley & Sons, Ltd. Pp. 713-750. 2003.

_____; SCHNEIDER, S. Arleqin ver. 3.0: Na integrated software package for population genetics data analysis. **Evolutionary Bioinformatics Online**, v.1, p.47-50, 2005.

_____; ESTOUP, A. ; CORNUET, J-M. (2005). **Bayesian analysis of an admixture model with mutation and arbitrarily linked markers.** *Genetics* 169: 1727-1738.

EVETT I,W. & GILL, P. **A discussion of the robustness of methods for assessing the evidential value of DNA single locus profile in crime investigations.** *Electrophoresis*. 12:226-230. 1991.

EDWARDS, A.; CHAKRABORTY, R.; CASKEY, C. T. **DNA typing with trimeric and tetrameric tandem repeats: polymorphic loci, detection systems and population genetics.** In: *Proceedings from the Second International Symposium on Human Identification*. Madison, WI: Promega Corp. p:31-52. 1991.

EDWARDS, A.1991. **DNA typing and Genetic Mapping With Trimetric and Tetrametric Tandem Repeats.** *Am. J. Hum. Genet.*, n. 49, p. 746-756,.

EXCOFFIER, L. & SLATKIN, M.. 1995. **Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population.** *Mol. Biol. Evol.* 12:921-927.

FERREIRA, M.E. & GRATTAPAGLIA, D. **Introdução ao uso de marcadores moleculares em análise genética.** 3ª ed. Brasília: EMBRAPA-CENARGEN, 1998.

FBI “Federal Bureau of Investigation – EUA) em 1988. **Combined DNA Index system (CODIS)** Homepage; <http://www.fbi.gov/hq/lob/codis/index1.htm>.

FISHER, R. **Standard calculations for evaluating a blood group system.** *Heredity*. 1951; 55: 95-102.

FISHER, R. **Standard calculation for evaluating a blood group system.** *Heredity*. 55: 95-102. 1951.

FOURNEY, R. *Genome Architecture – Repetitive DNA. The Application of DNA STR Analysis in Forensic Science.* Jun. 1998.

FRÉGEAU, C.J. **The polymerase chain reaction.** Part I: Its various components. Carleton University. 1998; June: 22-26.

_____; BOWEN, K.L. & FOURNEY, R.M. **Validation of highly polymorphic fluorescent multiplex short tandem repeat systems using two generations of DNA sequencers.** *J Forensic Sci.* 44(1):133-66. 1999.

FREYRE, G. & PUTNAM, S. **The masters and the slaves: a study in the development of brazilian civilization.** Am Soc Review. 12(5): 607-9.1947.

GINTHER, C.; ISSEL-TARVER, L. & KING, M.C. Identifying **individuals by sequencing mitochondrial DNA from teeth.** Nature genetics (1992) 2:135-138.

GÓES, A.C.S. **Tipagem Humana por DNA: otimização de condições para análises post mortem.** Tese. 2003.

GRADZICKER, T.; WILLIAMS, J., SHARP, P. & SAMBROOK, J. 1974. **Physical mapping of temperature sensitive mutations of adenoviruses.** Cold Spring Harbor Symp. Quant. Biol. 1974 39:439-446.

GRIFFITHS, A.J.F. **Na Introduction to Genetic analysis.** 7 ed. W.H.Freeman, 2000.

GRODZICKER, T. 1974. **Physical mapping of temperature-sensitive mutation of adenoviruses.** Cold Spring Harbor Symp. Quant. Biol. 39:439-446.

GUO, S. & THOMPSON, E. 1992 Performing **the exact test of Hardy-Weinberg proportion for multiple alleles.** Biometrics 48:361-372.

HAMM, W. A., SCHUELER, S. R. A. **Chemiluminescent technique to detect and quantitate human DNA in contaminated samples prior to restriction fragment length polymorphism (RFPL) analysis.** Kansas Bureau of Investigation, Topeka, Kansas. 1995.

HARDY, G.H. **Mendelian proportions in a mixed population.** Science.28:49-50. 1908.
HARTL, D.L. & CLARCK, A.G. 1997. Principles of population genetics, 3 rd edn. Sinauer, Sunderland, MA.

HOLLAND, M.M.; FISHER, D.L.; MITCHELL, L.G.; RODRIGUEZ, W.C.; CANIK, J.J.; MERRIL, C.R.; WEEDN, V.W. **Mitochondrial DNA sequence analysis of human skeletal remains: Identification of remains from the Vietnam War,** Journal of Forensic Sciences 38: 542-553. 1993.

HEDRICK, P.W. 1999. **Perspective: Highly variable genetic loci and their interpretation in evolution and conservation.** Evolutions 53:313-318.

JEFFREYS, A.J.; WILSON, V. & THEIN, S.L. 1985a. **Hypervariable "minisatellite regions in human DNA.** Nature London **314**:67-73.

_____; WILSON V., THEIN S. L. 1985b. **Individual Specific Fingerprint of Human DNA.** Nature **316**:76-79.

_____; WILSON, V. & NEUMANN R. **Amplification of human minisatellites by the polymerase chain reaction: towards DNA fingerprinting of. Single cells.** Nucleic Acids Research. 16(23):10953-10970. 1988.

KIMPTON, C. P.; GILL, P.; WALTON, A. **PCR methods and applications.** 3:13-22. 1993.

JAYLE, M.F. & GUILLARD, P. **Précipitation saline d'une protéine définie dans un mélange hétérogène.** Bull. Soc. Chim. Biol. 29:149. 1947.

LANDSTEINE, K., WINER, A.S. **An agglutinable factor in human blood recognized by immune sera for rhesus blood.** Proc. Soc. Exp. Biol. Med. 43:223. 1940.

LANDER, E.S.; LINTON, L.M.; BIRREN, B.; NUSBAUM, C.; ZODY, M.C.; BALDWIN, J.; DEVON, K.; DEWAR, K.; DOYLE, ..; FITZHUGH, W.; FUNKE, R.; GAGE, D.; HARRIS K.; HEAFORD, A.; HOWLAND, J.; KANN, L.; LEHOCZKY, J.; LEVINE, R.; MCEWAN, P.; MCKERNAN, K.; MELDRIM, J.; MESIROV, JP.; MIRANDA, C.; MORRIS, W.; NAYLOR, J.; RAYMOND, C.; ROSETTI, M.; SANTOS, R.; SHERIDAN, A.; SOUGNEZ, C.; STANGE-THOMANN, N.; STOJANOVIC, N.; SUBRAMANIAN, A.; WYMAN, D.; ROGERS, J.; SULSTON, J.; AINSCOUGH, R.; BECK, S.; BENTLEY, D.; BURTON, J.; CLEE, C.; CARTER N; COULSON A; DEADMAN R; DELOUKAS P; DUNHAM, A.; DUNHAM, I.; DURBIN, R.; FRENCH, L.; GRAFHAM, D.; GREGORY, S.; HUBBARD, T.; HUMPHRAY, S.; HUNT, A.; JONES, M.; LLOYD, C.; MCMURRAY, A.; MATTHEWS, L.; MERCER, S.; MILNE, S.; MULLIKIN, JC.; MUNGALL, A.; PLUMB ,R.; ROSS, M.; SHOWNKEEN, R.; SIMS, S.; WATERSTON, R.H.; WILSON, R.K.; HILLIER, L.W.; MCPHERSON, JD.; MARRA, M.A.; MARDIS, E.R.; FULTON, L.A.; CHINWALLA, A.T.; PEPIN, K.H.; GISH, W.R.; CHISSOE, S.L.; WENDL, M.C.; DELEHAUNTY, K.D.; MINER, T.L.; DELEHAUNTY, A.; KRAMER, J.B.; COOK, L.L.; FULTON, R.S.; JOHNSON, D.L.; MINX, P.J.; CLIFTON, S.W.; HAWKINS, T.; BRANSCOMB, E.; PREDKI, P.; RICHARDSON, P.; WENNING, S.; SLEZAK ,T.; DOGGETT, N.; CHENG, J.F.; OLSEN, A.; LUCAS, S.; ELKIN, C.; UBERBACHER, E.; FRAZIER, M.; GIBBS, R.A.; MUZNY, D.M.; SCHERER, S.E.; BOUCK, J.B.; SODERGREN, E.J.; WORLEY, K.C.; RIVES, C.M.; GORRELL, J.H.; METZKER, M.L.; NAYLOR, S.L.; KUCHERLAPATI, R.S.; NELSON, D.L.; WEINSTOCK, G.M.; SAKAKI, Y.; FUJIYAMA, A.; HATTORI, M.; YADA, T.; TOYODA, A.; ITOH, T.; KAWAGOE, C.; WATANABE, H.; TOTOKI, Y.; TAYLOR, T.; WEISSENBACH, J.; HEILIG, R.; SAURIN, W.; ARTIGUENAVE, F.; BROTTIER, P.; BRULS, T.; PELLETIER, E.; ROBERT, C.; WINCKER, P.; ROSENTHAL, A.; PLATZER, M.; NYAKATURA, G.; TAUDIEN, S.; RUMP, A.; YANG, H.M.; YU, J.; WANG, J.; HUANG, G.Y.; GU, J.; HOOD, L.; ROWEN, L.; MADAN, A.; QIN, S.Z.; DAVIS, R.W.; FEDERSPIEL, N.A.; ABOLA, A.P.; PROCTOR, M.J.; MYERS, R.M.; SCHMUTZ, J.; DICKSON, M.; GRIMWOOD, J.; COX, D.R.; OLSON, M.V.; KAUL, R.; RAYMOND, C.; SHIMIZU, N.; KAWASAKI, K.; MINOSHIMA, S.; EVANS, G.A.; ATHANASIOU, M.; SCHULTZ ,R.; ROE, B.A.; CHEN, F.; PAN, H.Q.; RAMSER, J.; LEHRACH, H.; REINHARDT, R.; MCCOMBIE, W.R.; DE LA BASTIDE, M.; DEDHIA, N.; BLOCKER, H.; HORNISCHER, K.; NORDSIEK, G.; AGARWALA, R.; ARAVIND, L.; BAILEY, J.A.; BATEMAN, A.; BATZOGLOU, S.; BIRNEY, E.; BORK, P.; BROWN, D.G.; BURGE, C.B.; CERUTTI, L.; CHEN, H.C.; CHURCH, D.; CLAMP, M.; COPLEY, R.R.; DOERKS, T.; EDDY, S.R.; EICHLER, E.E.; FUREY, T.S.; GALAGAN, J.; GILBERT, J.G.R.; HARMON, C.; HAYASHIZAKI, Y.; HAUSSLER, D.; HERMJAKOB, H.; HOKAMP, K.; JANG, W.H.; JOHNSON, L.S.; JONES, T.A.; KASIF, S.; KASPRYZK, A.; KENNEDY, S.; KENT, W.J.; KITTS, P.; KOONIN, E.V.; KORF, I.; KULP, D.; LANCET, D.; LOWE, T.M.; MCLYSAGHT, A.; MIKKELSEN, T.; MORAN, J.V.; MULDER, N.; POLLARA, V.J.; PONTING, C.P.; SCHULER, G.; SCHULTZ, JR; SLATER, G.; SMIT, A.F.A.; STUPKA, E.; SZUSTAKOWKI, J.; THIERRY-MIEG D.; THIERRY-MIEG, J.; WAGNER, L.; WALLIS ,J.; WHEELER, R.; WILLIAMS, A.; WOLF, Y.I.; WOLFE, K.H.; YANG, S.P.; YEH, R.F.; COLLINS, F.; GUYER, M.S.; PETERSON, J.; FELSENFELD, A.;

- WETTERSTRAND, K.A.; PATRINOS, A.; MORGAN, M.J. **Internacional Human Genome Sequencing Consortium**. Initial Sequencing and Analysis of the Human Genome. *Nature*. 2001; 409: 860-921.
- LEITE, F.P.N.; MENEGASSI, F.J.; SCHWENGBER, S.P.; RAIMANN. **STR data for 09 autosomal STR markers from Rio Grande do Sul (southern Brazil)**. *Forensic Sci Int*; 132: 223-224. 2003.
- _____; TOCCHETTO, D.; ESPINDULA, A. **Exames de DNA Humano**. *Criminalística: procedimentos e metodologias*, p. 242-259, 2005.
- LEVEDAKOU, E.N., FREEMAN, D.A., BUDZYNSKI, M.J., EARLY, B.E, MCELRESH, M.J, SCHUMM, J.W. **Allele Frequencies for Fourteen STR Loci of the PowerPlex™ 1.1 and 2.1 Multiplex Systems and Penta D Locus In Caucasians, African-Americans, Hispanics, and Other Populations of the United States of America and Brazil [published erratum appears in J Forensic Sci 2001; 46(6): 1533. J Forensic Sci 2001; 46(3): 736-61.**
- LEWIN, B. 1990. **Genes IV**. Oxford University Press, Oxford.
- LITT, M. & LUTY, J.A. 1989. **A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene**. *Am. J. Hum. Genet.* **44**:398-401.
- LORENTE J.A.A. **The GITAD Project: DNA typing in Latin America**. *Profiles in DNA*. 3: 8-9. 1999.
- MACAULAY, V.; RICHARDS, M.; HICHEY, E. **The Emerging Tree of West Eurasian mtDNAs: A Synthesis of Control-Region Sequences and RFLPs**. *Am. J. Hum. Genet.* 64 (p): 232-249. 1999.
- MATHER, K. & JINKS, J. L. **Biometrical genetics**. 3 ed. Chapman Hall. London. 1994.
- PROMEGA CORPORATION. **GenePrint™ PowerPlex™ 2.1 System. Technical Manual**, 1998.
- MANTEL, N. 1967. **The detection of disease clustering and a generalized regression approach**. *Cancer Res* 27:209-220.
- MANTA, F.S. 2008. **Desenvolvimento e validação de sistema multiplex de tipagem para identificação humana por DNA**. Rio de Janeiro, 132p. Dissertação de mestrado.
- MICKA, K.; AMIOTT, E.; HOCHENBERRY, T.L. TWGDAN. **Validation of a nine-locus and a four-locus fluorescent STR multiplex system**. *J. Forensic Sci.* 44(6): 1-15, 1995.
- _____; AMIOTT, E.A.; HOCHENBERRY, T. TWGDAM. **Validation of a Nine-Locus and a Four-Locus Fluorescent STR Multiplex System**. *J. Forensic Sci.* 44(6): 1-15. 1999.
- MORRAL, N.; ESTIVILL, X. **Multiplex PCR amplification of three microsatellites within the CFTR gene**. *Genomics*. 13:1362-1364. 1992.
- MORTARA, G. **The development and structure of Brazil's population**. *Population studies*. 8(2):121-139. 1954.

MORTON, N.E. **Genetic structures of forensic populations**. Proc. Natl. Acas. Sci. 89: 2556-2560. 1992.

MULLIS, K.B. & FALOONA, F.A. **Specific synthesis of DNA in vitro via a polymerase catalyzed chain reaction**. Methods in Enzymology. **155**: 335-350. 1987.

NADIR, E.; MARGALIT, H. ; GALLILIY, T. **Proceedings of the National Academy of Sciences USA**. 93 6470-6475. 1996.

NAKAMURA, Y. ; LEPPERT, M. ; O'CONNELL, P. **Variable number tandem repeat (VNTR) markers for human gene mapping**. Science. 235:1616-1622. 1987.

NEI, M. & ROYCHOUDHURY, A.K. **Sampling variances of heterozygosity and genetic distance**. Genetics. 76. 379-390. 1973.

_____; CHACRAVARTY, A. **Drift variance of Fst and Gst statistics obtained from a finite number of isolated populations**. Theoretical Population Biology. 11(3):307-315. 1977.

_____. **Molecular evolutionary genetics**. New York: Columbia University Press. 512pp. 1987.

_____. **F- statistics and analysis of gene diversity in subdivided populations**. Annals of Human Genetics 41:225-233. 1977.

NODARI, R.O.; TSAI, S.M.; GILBERTSON, R.L. & GEPTS, P. 1993. **Towards an integrated linkage map of common bean**. II. Development of an RFLP-based linkage map. Theor. Appl. Genet. 85:513-520.

PARAN, I.; KESSELI, R. & MICHELMORE ,R. **Identification of restriction fragment length polymorphism and random amplified polymorphic DNA markers linked to downy mildew resistance genes in lettuce, using near-isogenic lines**. Genome 34:1021-1027. 1991.

PEARSON, W. R. **Identifying distantly related protein sequences**. Comput. Appl. Biosc. **31**:325-32. 1997.

PENA, S.D.J. Razões para banir o conceito de raça da medicina brasileira. História, Ciências, Saúde – manguinhos. **(12)1**: 321-46. 2005.

PONTES, I.M. 2003. **Desenvolvimento de novos marcadores microssatélites para análise genética em humanos**. Dissertação, S. Carlos, SP.

PRIMACK, R. **Essentials of conservation biology**. 4 th Edition, Sinauer Associates, 2006.

PROVINE, W.B. 2001. **The origins of theoretical population genetics with a new afterword.** University Chicago Press, Chicago.

WEBER, B.; RIESS, O.; WOLFF, G. **Delineation of a 50 kilobase segment containing the recombination site in a sporadic case of Huntington's disease.** *Nature Genetics.* 2:216-222. 1992.

PUERS, C.; HAMMOND, H.; JIN, L.; CASKEY, T. **Identification of repeat sequence Heterogeneity at the polymorphic short Tandem Repeat Locus HUMTH01 [AATG]_n and Reassignment of alleles in Population Analysis by Using Locus Specific Allelic Ladder.** *Am J Hum Genet.* 53:953-958. 1993.

RAMOS, A. 1946. **Las poblaciones del Brazil.** *Am Soc Review.* 11(3):362-63.

REYNOLDS, R.; SENSABAUGH, G.; BLAKE, E. **Analysis of Genetic Markers in Forensic DNA Samles Using the Polymerase Chain Reaction.** *Analytical Chemistry.* 63(1): 2-14. 1991.

RIBEIRO, D. **Os índios e a civilização: A integração das populações indígenas no Brasil** Moderno Companhia das Letras, São Paulo, 1996

RIBEIRO, D. **Os índios e a civilização: A integração das populações indígenas no Brasil** Moderno Companhia das Letras, São Paulo, 1996

SALZANO, FM; BLACK, FL; CALLEGARI-JACQUES, SM; SANTOS, SE; WEIMER, TA; MESTRINER, MA; PANDEY, Jp; HUTZ, MH; RIEGER, TT. **Genetic variation within a linguistic group: Apalai-Wayana and other Carib tribes.** *Am J Phys Antropol* 1988 Mar; 75(3):347-56, 1988

SAMBROOK, J.; FRITSCH, E.F.; MANIATIS, T. **Molecular cloning: a laboratory manual,** 1989.

SANGER, F.; NICKLEN, S.; COULSON, A.R. **DNA Sequencing with chain – terminating inhibitors.** *Proc. Natl. Acad. Sci. EUA.* 74(12): 5463-5467. 1977.

SAJANTILA, A.; LUKKA, M.; SYVANEN, A.C. **European Journal of Human Genetics.** 7-263-266. 1999.

SCHAFFER, A.J.; HAWKINS, J.R. **DNA variation and the FUTURE OF Human Genetics.** *Nature Biotechnology,* n.16, p. 33-39, 1998.

SCHYMA, C.; HUCKENBECK, W; BONETE, W. **DNA-PCR Analysis of bloodstains sampled the polyvinyl method,** *J. Forensic Sci.* 44(1):95-99. 1999.

SCHNEIDER, S.; ROESSLI, D.; EXCOFFIER, L. **ARLEQUIN v 2000. A software for population genetic data analysis,** University of Geneva, 2000.

_____; EXCOFIER, L. LAVAL, G. 2006. **Arlequin version 3.1 : An Integrated Software package for population genetics data analysis**. Computational and Molecular Population Genetics Lab (CMPG), **Institute of Zoology, University of Berne**.
<http://cmpg.unibe.ch/software/arlequin3>.

SILVA, M.T.D.; SILVA, D.A.; GÓES, A.C.C.; CARVALHO, E.F. **Análise da conservação do DNA de amostras de sangue submetidas à variação de temperatura e tempo de estocagem**. 48 Congresso da Sociedade Brasileira de genética. Águas de Lindóia. 24-27. 2002.

SILVA, D.A.; CROUSE, C.A.; CHAKRABORTY, R.; GÓES, A.C.S.; CARVALHO, E.F. **Statistical analyses of 14 short tandem repeat loci in Brazilian populations from Rio de Janeiro and Mato Grosso do Sul states for forensic and identity testing purposes**, *Forensic Sci Int*. 2004a; 139: 173-176.

_____; GÓES, A.C.S.; CARVALHO, E.F. **Eight short tandem repeat (STR) frequencies for descendants from Terena indigenous Brazilian group**. *J. Forensic Sci*. 2004b; 49(2).

_____; GÓES, A.C.S.; CARVALHO, J.J.; CARVALHO, E.F. **The usefulness of DNA extracted from vaginal smear slides in forensic casework - sperm DNA typing in rape cases**. *São Paulo Medical Journal, São Paulo*. 2004c; 122: 34-38.

_____. **Metodologia para Análise Forense por DNA: Estudo de Validação para 14 Loci STR e Construção de Banco de Dados**. Tese de Doutorado. Universidade do Estado do Rio de Janeiro. 2004.

SLATKIN, M. 1995. **A measure of population subdivision based on microsatellite allele frequencies**. *Genetics* 139: 457-462.

SPRECHER, E. J.; PUERS, C.; LINS, A.M. **General approach to analysis of polymorphic short tandem repeat loci**. *Biotechniques*. 20:266-276. 1996.

STRACHAN, T. & READ, P.A. **Genética Molecular Humana**. Artmed Editora, Porto Alegre, 2002.

TORRONI, A.; NEEL, J.V.; BARRANTES, R. **Mitochondrial DNA (clock) for the Amerinds and its implications for timing their antry into North America**. 91:1158-1162.1994.

VAN HOOFFSTAT, D.E.; DEFORCE, D.L.; MILLECAMP, S R.E., **Population genetic study of four short tandem repeat loci in the Belgian population, using capillary electrophoresis**. *Electrophoresis*. 19(5):719-22.1998.

VIGILANT, L.; STONEKING, M.; HARPENDING, H.; HAWKES, K.; WILSON, A.C. **African population and the evolution of mitochondrial DNA**. *Science*. 253: 1503-1507. 1991.

WATSON, S.; ALLSOP, R.; FOREMAN, L.; KELSEY, Z.; GILL, P. **Sequenced allelic ladders and population genetics of a new STR multiplex system**. *Forensic Sci. Int*. 2001; 115; 207-217.

- WALSH, P.S.; FILDES, N.J.; REYNOLDS, R. **Sequence analysis and characterization of stutter products at the tetranucleotide repeat locus vWA.** Nucleic Acids Res. 24(14):2807-12. 1996.
- WEBER, J. L.; WONG, C. **Human Molecular genetics.** 2:1123-1128. 1993.
- WEIR, B. S. and COCKERHAM, C.C. 1984. **Estimating F-statistics for the analysis of population structure.** Evolution 38:1358-1370.
- WELSH, J. & McCLELLAND, M. 1990. **Fingerprinting genomes using PCR with arbitrary primers.** Nucleic Acids Res. 18:7213-7218.
- WHITAKER, J.P. **Short Tandem Repeat Typing of Bodies from a Mass Disaster: High success rate and characteristic amplification patterns in highly degraded samples.** Bio Techniques, v. 4, n. 18, p. 670-677, 1995.
- WILLIAMS, J.G.; KUBELIK, A.R.; LIVAK, K.J; RAFALSK, L.A & TINGEY, S.V. **DNA polymorphism amplified by arbitrary primers are useful as genetic markers.** Nucleic Acids Res. 18: 6531-6535. 1990.
- WILLOTT, G. M.; CROSSE, M. A. **The Detections of Spermatozoa in the Mouth.** Journal of the Forensic Science Society. 26:125-128. 1986.
- WRIGHT, S. 1969. **Evolution and the Genetics of Populations.** Vol. 2. The Theory of Gene Frequencies. University of Chicago Press, Chicago.
- WYMAN, A. R.; WHITE, R. A highly polymorphic *locus* in human DNA. Proc. Natl. Acad. Sci. EUA. 77:6754-6758. 1980.
- YAFFE, M.P. **The Machinery of Mitochondrial inheritance and behavior.** Science. 283;1493-1497.1999.

APÊNDICE

Tabela de soluções e tampões utilizadas na fase experimental deste trabalho são mostrados a seguir.

| Soluções | Especificação |
|-----------------------------------|---|
| Para reação da PCR | <ul style="list-style-type: none"> - Tampão 10X para PCR <ul style="list-style-type: none"> Tris-HCl 200 mM pH= 8,4 KCl 500 mM - Solução dNTPs <ul style="list-style-type: none"> Mistura com 2,5 mM cada dATP 2,5 mM dGTP 2,5 mM dTTP 2,5 mM dCTP 2,5 mM - <i>Taq.</i> Polimerase 5U/μL (CENBIOT/UFRGS) - Iniciadores 1 pmol/μL de cada (F / R) - MgCl₂ 20 mM |
| Para Eletroforese | <ul style="list-style-type: none"> - Tampão TBE 10X (estoque) <ul style="list-style-type: none"> Tris-base 108 g Ácido bórico 55 g EDTA (0,5M, pH 8,0) 40 mL Água destilada completar para 1000 mL, pH = 8,0 / 8,4 - Tampão de Amostra TBE 5X <ul style="list-style-type: none"> TEB 10X 5 mL Glicerol – 30 % 3 mL Azul de bromofenol 5 mg Água Milli-Q completar volume para 10 mL - Brometo de etídio 10 mg/mL <ul style="list-style-type: none"> Brometo de etídio 10 mg Água Milli-Q completar volume para 10 mL |
| Para gel de agarose 0,8 % | <ul style="list-style-type: none"> - Agarose 0,8 g TEB 10X 10 mL Água Destilada completar para 100 mL |
| Para gel de poliacrilamida | <ul style="list-style-type: none"> - Acrilamida/bis-Acrilamida 40 % (solução estoque) <ul style="list-style-type: none"> Acrilamida 380 g Bis-acrilamida 20 g Água Milli-Q completar para 1 L - Gel de poliacrilamida 4 % |

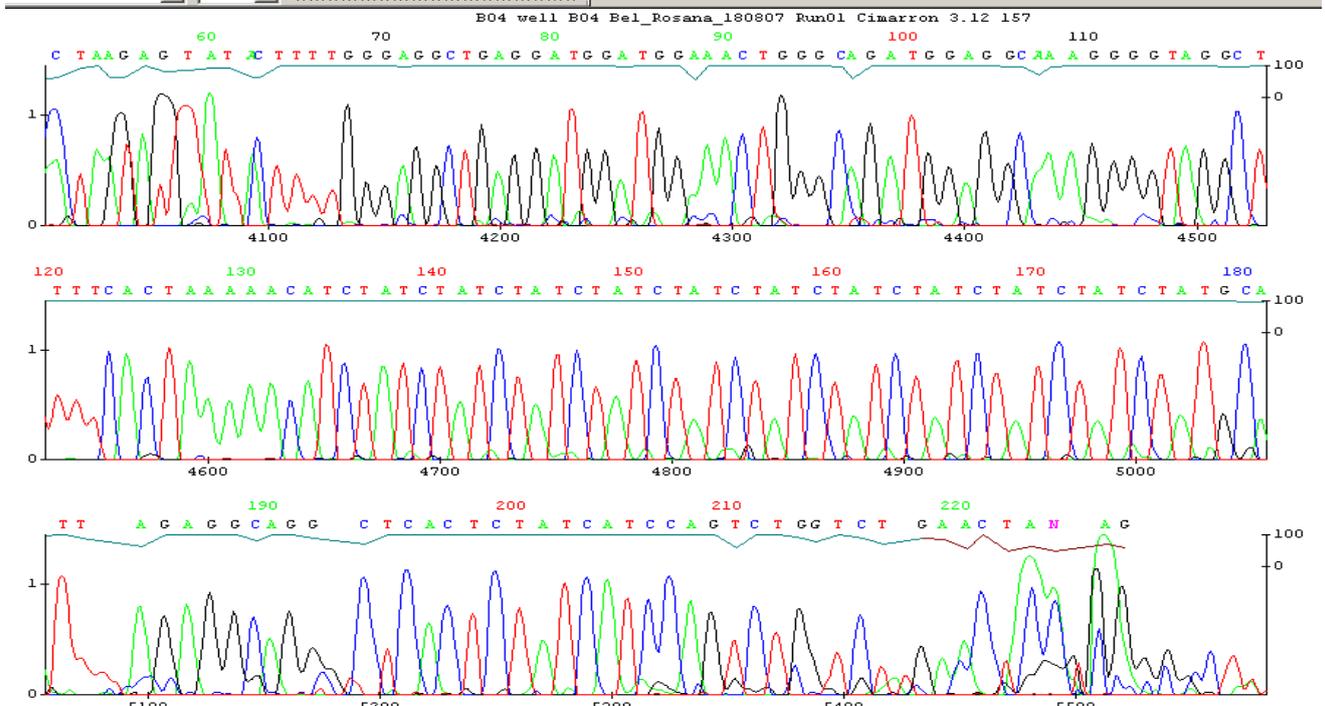
| | | |
|---|--|-----------------------------|
| | Uréia | 420 g |
| | TBE 10X | 50 mL |
| | Acrilamida/bis-acrilamida 40 % | 100 mL |
| | Água Milli-Q | completar o volume para 1 L |
| Para coloração em prata (Creste et al., 2001). | - Solução de Fixação (Fix/Stop) | |
| | Etanol 10 % | 100 mL |
| | Ac. Acético (1 %) | 10 mL |
| | H ₂ O Milli-Q | - completar para 1 L |
| | - Solução de Oxidação | |
| | Ácido Nítrico (absoluto) | 15 mL |
| | H ₂ O Milli-Q | 985 mL |
| | - Solução de Impregnação-Nitrato de Prata 0,2 % (Staining Solution): | |
| | Nitrato de Prata | 2 g |
| | H ₂ O Milli-Q | 1000 mL |
| | - Solução de Revelação – Carbonato de Sódio 3 % (Developer) | |
| | Carbonato de sódio | 30 g |
| | H ₂ O Milli-Q | - completar para 1 L |
| | OBS: Na hora do uso adicionar | 540 µL de formaldeído 37 %. |
| | - Solução de Bloqueio (STOP) | |
| Ácido acético 5 % | 50 mL | |
| H ₂ O Milli-Q | 950 mL | |
| - Tampão de amostra para gel de poliacrilamida | | |
| Formamida 98 % | 19,6 mL | |
| EDTA 0,5M pH=8.0 | 400 µL | |
| Azul de Bromofenol | 5 mg | |
| Para preparo das placas: | | |
| - (Solução adesiva) BIND | | |
| Álcool absoluto | 994 µL | |
| Ácido Acético Glacial | 5 µL | |
| Metacriloxipropiltrimetoxisilane | 1 µL | |
| - REPEL (Solução repelente) | | |
| Solução de silicone especial em heptano | | |
| (sigmacote) | | |
| - Ladder | | |
| Extração de DNA | Solução para 12µL (0,5µg/µL) | |
| | Água milli-Q | 12 µL |
| - SEB (pH 8) | | |
| EDTA (pH 8) | 10 mM | |
| NaCl | 100 mM | |
| Tris-Cl (pH 8) | 10 mM | |
| SDS | 2 % | |
| Tampão Platinum Plus 4x (para 100 µL) | - 10x PCR buffer para 40 µL | |
| | MgCl ₂ 50 mM | 08 µL |
| | dNTP Mix 40 mM – Promega. | 12 µL |

H2O Milli Q

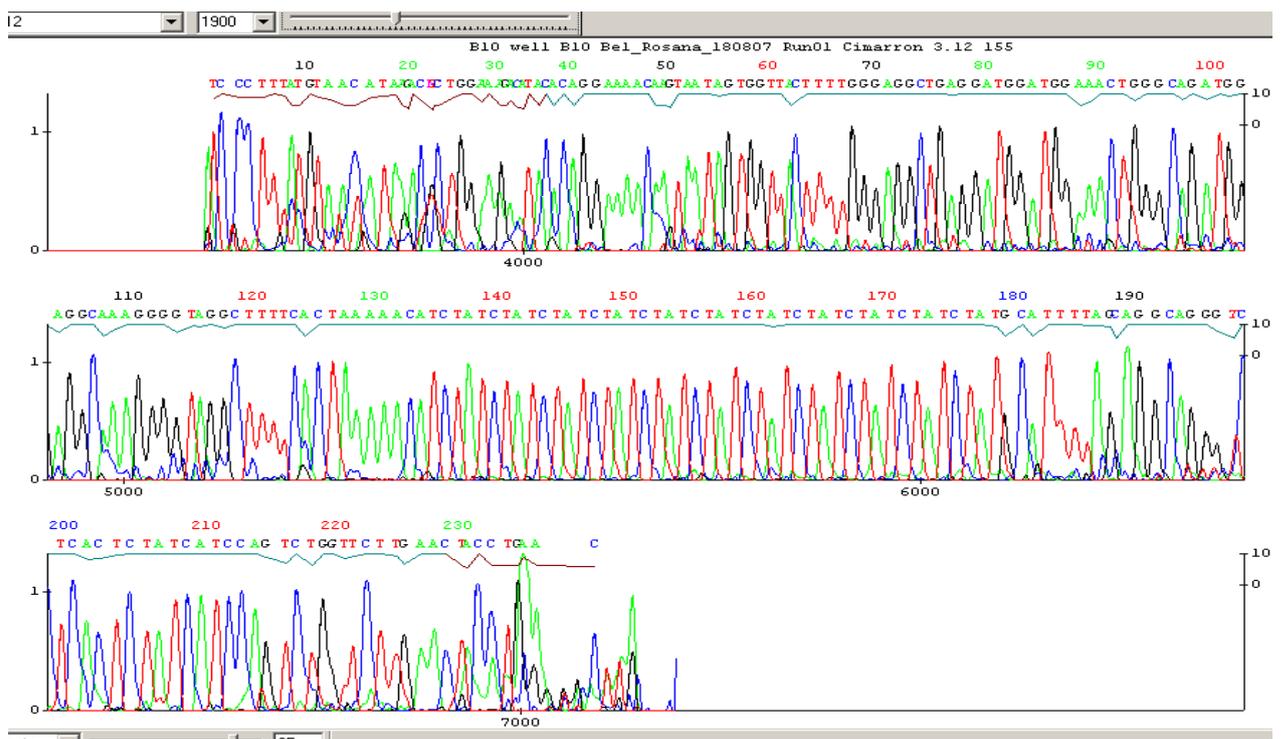
40 µL

Resultado de Sequenciamento

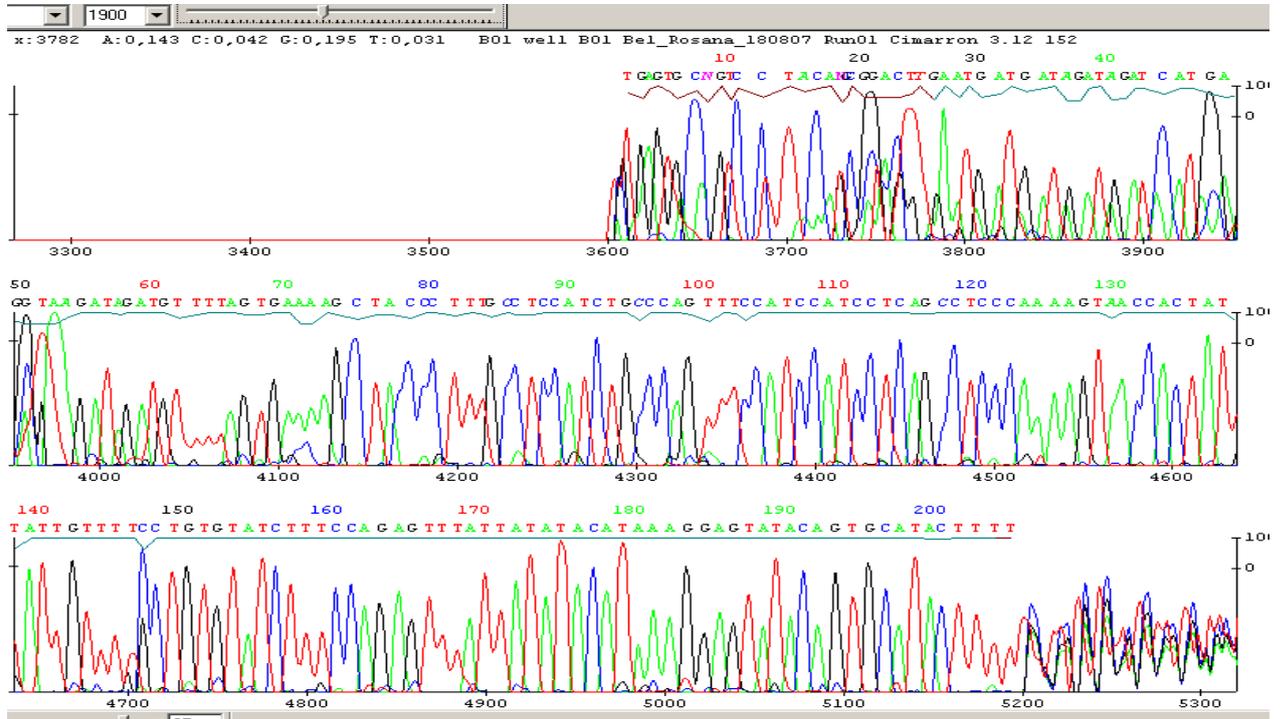
Primer ISSO-3G(AGAT)10
Amostras 101-(Reverse-TCTA) (1uL)



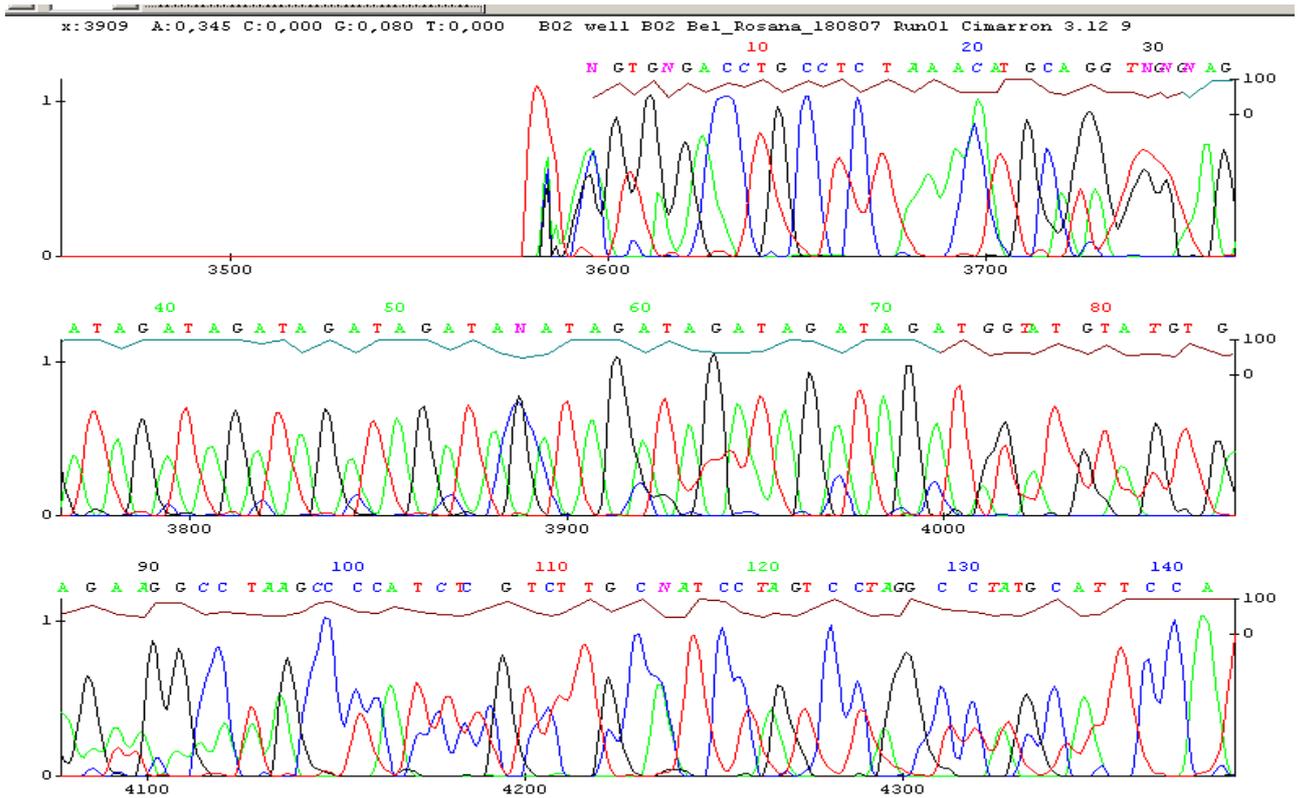
Primer ISSO-3G(AGAT)10
Amostras 101-(Reverse-TCTA) (2uL)



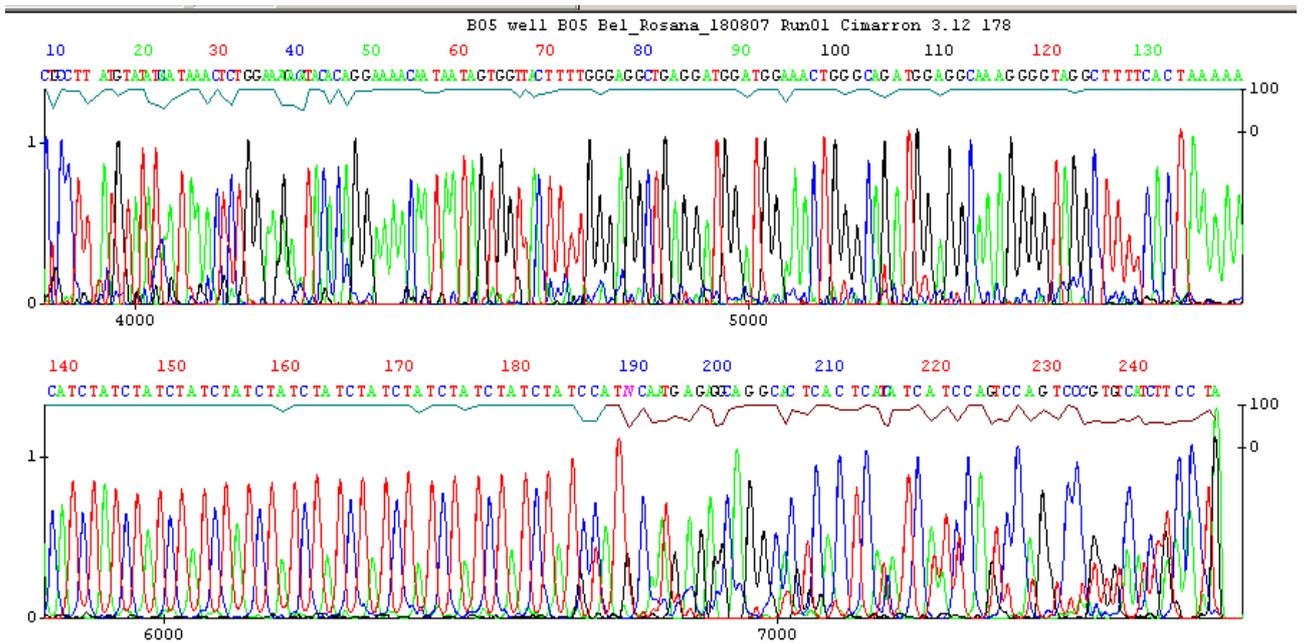
Primer ISSO-3G(AGAT)10
 Amostras 101-(Forward) (1uL)



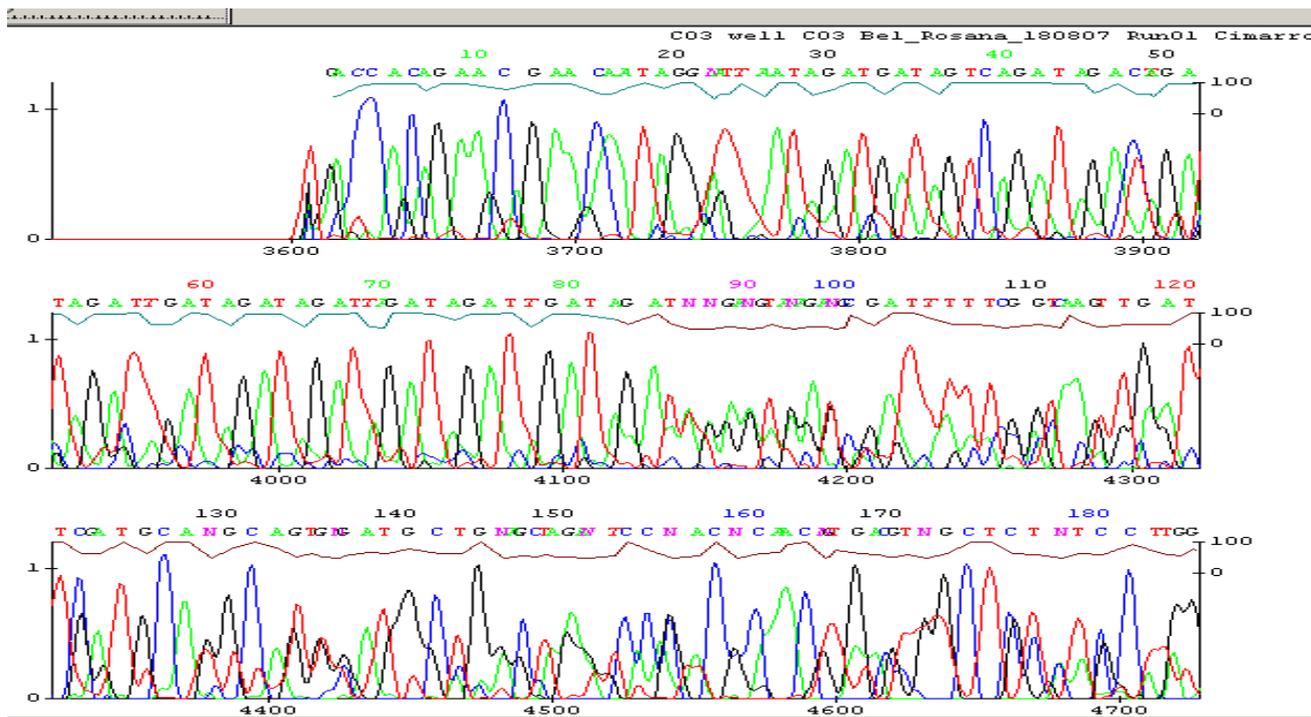
Primer ISSO-3G(AGAT)10
 Amostras 103-(Forward) (2uL)



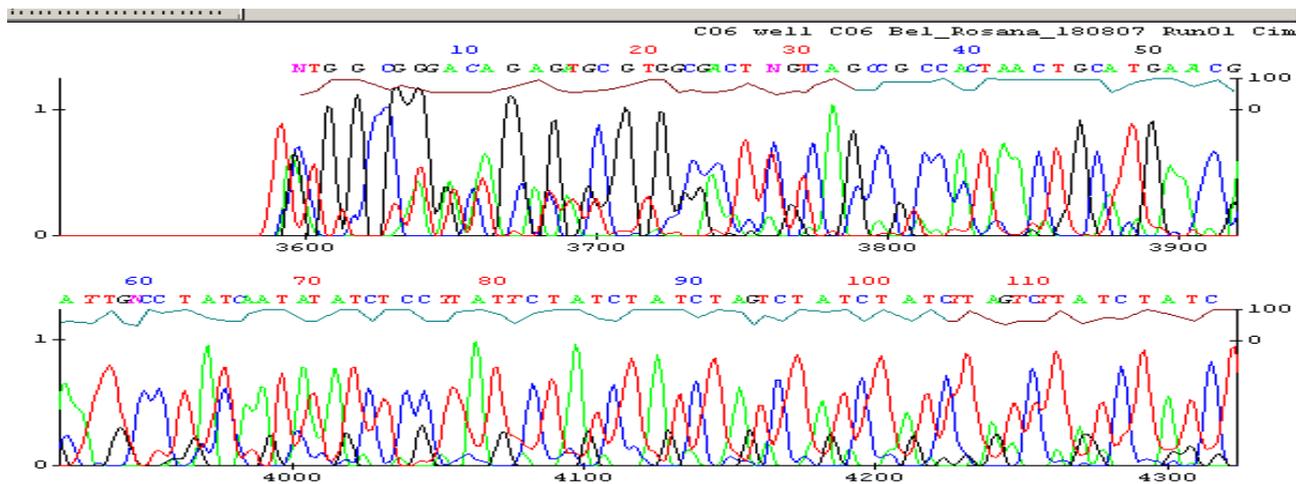
Primer ISSO-3G(AGAT)10
 Amostras 103-(Reverse) (1uL)



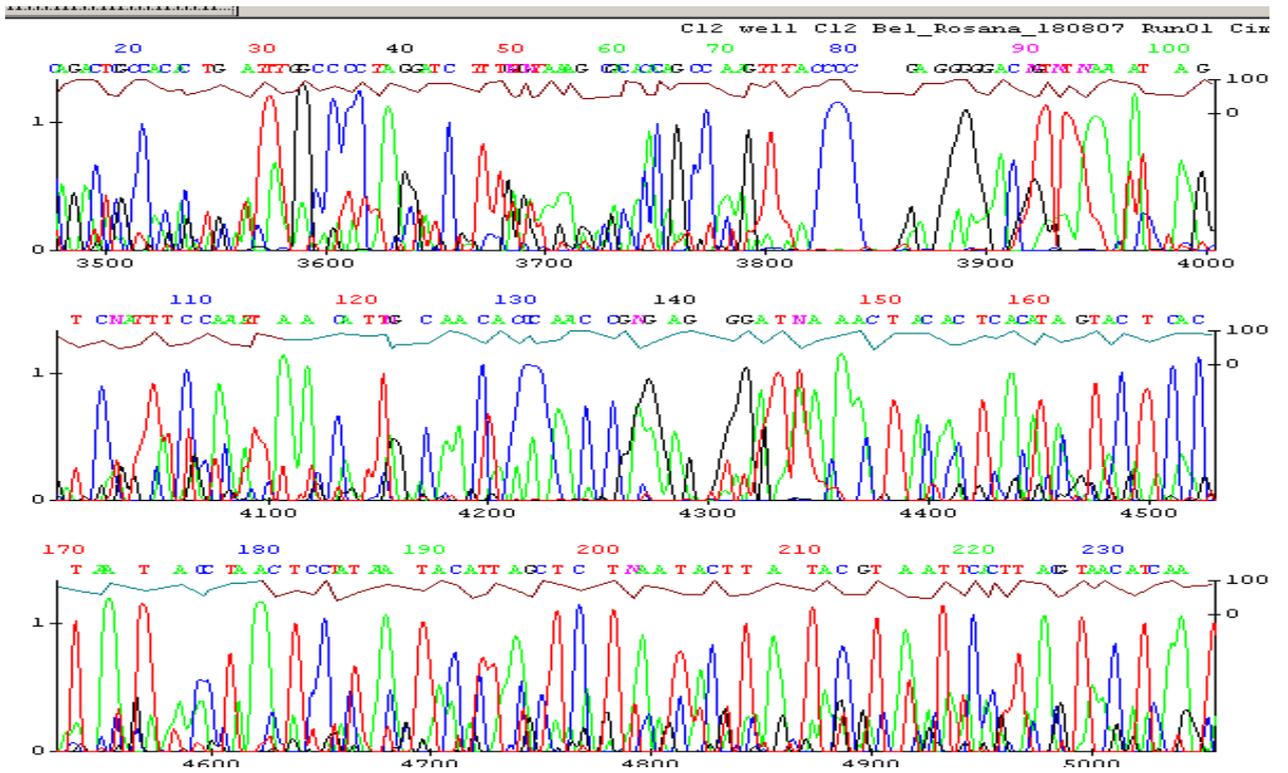
Primer ISSO-3G(AGAT)10
 Amostras 103-(Rverse) (2uL)



Primer ISSO-1M(CTAT)10
 Amostras 104-(Reverse) (1uL)



Primer ISSO-1M(CTAT)10 / Amostras 104-(Reverse) (1uL)



RELAÇÃO DOS ALELOS ENCONTRADOS NOS 5 LOCI

| MARCADOR | NOME DO ALELO | TAMANHO DO ALELO | MARCADOR | NOME DO ALELO | TAMANHO DO ALELO |
|--------------------|---------------|------------------|--------------------|---------------|------------------|
| ISO2M (FAM) | | | ISO3G (FAM) | | |
| | 8 | 187 | | 5 | 242 |
| | 9 | 191 | | 6 | 246 |
| (Gene Bank 200pb) | 9.3 | 194 | (Gene Bank 260pb) | 7 | 250 |
| Cromossomo 5 | 10 | 195 | Cromossomo 11 | 8 | 254 |
| (GGCT)11 | 10.3 | 198 | (AGAT)10 | 8.3 | 257 |
| | 11 | 199 | | 9 | 258 |
| (9 alelos) | 12 | 203 | (12 alelos) | 9.3 | 261 |
| | 13 | 207 | | 10 | 262 |
| | 14 | 211 | | 11 | 266 |
| | | | | 11.3 | 269 |
| | | | | 12 | 270 |
| | | | | 13 | 274 |
| ISO5P (HEX) | | | ISO2G (HEX) | | |
| | 4.2 | 141 | | 6 | 245 |
| | 5 | 143 | | 6.1 | 246 |
| | 6 | 147 | | 6.2 | 247 |
| | 7 | 151 | | 6.3 | 248 |
| (Gene Bank 170pb) | 8 | 155 | (Gene Bank 260pb) | 7 | 249 |
| Cromossomo 15 | 8.2 | 157 | Cromossomo 8 | 7.2 | 251 |
| (GATA)13 | 9 | 159 | (TCTG)10 | 8 | 253 |
| | 10 | 163 | | 8.2 | 255 |
| | 11 | 167 | | 9 | 257 |
| | 12 | 171 | | 9.1 | 258 |
| (17 alelos) | 13 | 175 | (20 alelos) | 9.2 | 259 |
| | 14 | 179 | | 9.3 | 260 |
| | 15 | 183 | | 10 | 261 |
| | 16 | 187 | | 10.1 | 262 |
| | 17 | 191 | | 10.2 | 263 |
| | 18 | 195 | | 11 | 265 |
| | 19 | 199 | | 11.2 | 267 |
| | | | | 12 | 269 |
| | | | | 13 | 272 |
| | | | | 14 | 276 |
| ISO1M (NED) | | | | | |
| | 10 | 204 | | | |
| | 11 | 208 | | | |
| (Gene Bank 220pb) | 12 | 212 | | | |
| Cromossomo 14 | 13 | 216 | | | |
| (ACAT)14 | 14 | 220 | | | |
| | 15 | 224 | | | |
| (7 alelos) | 16 | 228 | | | |

Negrito: tipagens do controle positivo K562

Tabela 21: A análise do Equilíbrio de Hardy-Weinberg nos 5 *loci* STR, em três populações do Brasil (Amazonas, Rio de Janeiro e Espírito Santo), utilizando o pool 308 indivíduos.

| <i>Locus</i> | Genótipos | Heterozigosidade Observada | Heterozigosidade Esperada | Valor de ($P < 0,05$) |
|--------------|-----------|----------------------------|---------------------------|-------------------------|
| D5S198498 | 308 | 0.58766 | 0.58343 | 0.00360 |
| D11S118590 | 308 | 0.72403 | 0.68533 | 0.06677 |
| GH15 | 308 | 0.72078 | 0.77056 | 0.09564 |
| D8S82535 | 308 | 0.75078 | 0.75056 | 0.11471 |
| D3S18773 | 308 | 0.67532 | 0.67575 | 0.01142 |