

UNIVERSIDADE ESTADUAL PAULISTA
“JÚLIO DE MESQUITA FILHO”

INSTITUTO DE QUÍMICA – CÂMPUS DE ARARAQUARA
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOTECNOLOGIA

Danilo Faustino Braganholi

**Análise de polimorfismos INDELS na Identificação
Humana**

Araraquara - SP
2016

DANILO FAUSTINO BRAGANHOLI

Análise de polimorfismos INDELS na Identificação Humana

Tese apresentada ao Instituto de Química de Araraquara, Universidade Estadual Paulista “Julio de Mesquita Filho”, como parte dos requisitos para obtenção do título de Doutor em Biotecnologia.

Área de concentração: Biologia Molecular

Orientadora: Profa. Dra. Regina Maria Barretto Cicarelli

Araraquara - SP
2016

DANILO FAUSTINO BRAGANHOLI

Tese apresentada ao Instituto de Química,
Universidade Estadual Paulista, como parte
dos requisitos para obtenção do título de
Doutor em Biotecnologia.

Araraquara, 17 de fevereiro de 2016

BANCA EXAMINADORA



Profª Drª Regina Maria Barretto Cicarelli
Faculdade de Ciências Farmacêuticas – UNESP, Araraquara



Profª Drª Raquel Mantuaneli Scarel Caminaga
Faculdade de Odontologia – UNESP, Araraquara



Prof. Dr. Celso Teixeira Mendes Júnior
Faculdade de Ciências Farmacêuticas – USP, Ribeirão Preto



Prof. Dr. Rogério Nogueira de Oliveira
Faculdade de Odontologia – USP, São Paulo



Profª Drª Maria Leonor Rodrigues de Souza Botelho Gusmão
Universidade do Estado do Rio de Janeiro – UERJ, Rio de Janeiro

DEDICATÓRIA

Aos meus maiores exemplos de motivação,
Meus pais Antonio e Vera.

AGRADECIMENTOS

À minha família, pelo apoio em todos os sentidos imagináveis, em especial minha esposa Lis, quem multiplica diariamente minha alegria.

À Professora Regina Cicarelli, pela confiança, atenção e por proporcionar esta fase de grande aprendizado.

Ao Dr. Rui Pereira, pelo auxílio neste trabalho e ótima experiência em Portugal, e à Dra. Leonor Gusmão, pela disponibilidade e atenção.

Às amigas Isabela e Fernanda, pelo auxílio neste trabalho, pela amizade e por fazerem do laboratório um ótimo ambiente de trabalho.

À Dra. Joyce Martins, pelo auxílio e esclarecimento de dúvidas.

À todos os amigos dos laboratórios de Paternidade e Imunologia pela amizade e companhia diária.

Ao CNPq pela bolsa do Doutorado e a CAPES pela bolsa de Portugal.

“Cada um de nós compõe a sua história, cada ser em si carrega o dom de ser capaz e ser feliz”.
Almir Sater e Renato Teixeira

LISTA DE ABREVIATURAS E SIGLAS

AIM-INDELS	INDELS informativos de ancestralidade
DNA	Ácido desoxirribonucleico
DNA mt	DNA mitocondrial
EC	eletroforese capilar
GHEP-ISFG	Grupo de Línguas Espanhola e Portuguesa da Sociedade Internacional de Genética Forense
He	Heterozigosidade esperada
HID	<i>Human Identification</i> , identificação humana
HID-INDELS	INDELS para identificação humana
Ho	Heterozigosidade observada
IBGE	Instituto Brasileiro de Geografia e Estatística
INDELS	Polimorfismos de inserção/deleção
Kb	kilo pares de base (1000 pb)
LD	<i>Linkage Disequilibrium</i> , desequilíbrio de ligação
MCMC	Markov Chain Monte Carlo
MID	<i>Marshfield identification</i>
NA	não se aplica
NCBI	National Center for Biotechnology Information
NGS	<i>Next Generation Sequencing</i> , sequenciamento de nova geração
pb	pares de base
q.s.p.	quantidade suficiente para
rpm	rotações por minuto
STR	<i>Short tandem repeat</i>
SNP	<i>Single nucleotide polymorphism</i>

RESUMO

Os marcadores STR são os mais utilizados na rotina de identificação humana e genética forense, entretanto, os marcadores INDELS vem chamando a atenção dos pesquisadores desta área, pois sua análise pode ser uma ferramenta interessante por serem analisados com um fragmento menor que os STR e apresentarem baixa taxa de mutação, podendo ser utilizados na identificação de indivíduos e na avaliação de ancestralidade. Neste trabalho, caracterizamos as populações brasileiras dos estados de São Paulo e Espírito Santo pela análise de marcadores INDEL através de dois sistemas: 38 HID-INDELS, verificando a eficiência forense nas duas populações e comparando os dados com os de STRs rotineiramente utilizados na população de São Paulo; e 46 AIM-INDELS, avaliando as proporções de ancestralidade nas duas populações, e comparando os dados com os de marcadores uniparentais na população do Espírito Santo. Ambos os métodos foram eficientes para suas respectivas finalidades, sendo que o sistema 38 HID-INDELS apresentou alto poder de discriminação, para Espírito Santo (PD = 0,9999999999999990) e para São Paulo (PD = 0,9999999999999994); e o sistema 46 AIM-INDELS confirmou a miscigenação das populações estudadas, e neste caso, com maior ancestralidade genética de europeus, em comparação a africanos e nativo-americanos. Além disso, inserimos o marcador amelogenina no sistema multiplex 38 HID-INDELS como uma ferramenta complementar para identificação de sexo de amostras degradadas.

Palavras-chave: INDELS. Identificação humana. Ancestralidade

ABSTRACT

The STR markers are the most used in routine of human identification and forensic genetics, however, INDEL markers has attracted the attention of those researchers in this area, because their analysis can be an interesting tool to be analyzed with a smaller fragment than STR and to present low mutation rate, may be used to identify individuals and evaluating ancestry. In this work, we characterized the Brazilian populations of the states of São Paulo and Espírito Santo by INDEL markers analysis through two systems: 38 HID-INDELS, checking the forensic efficiency in this two populations and comparing the data with STRs routinely used in São Paulo population; and 46 AIM-INDELS, assessing the proportions of ancestry in this two populations, and comparing the data with uniparental markers in Espírito Santo population. Both methods were effective for their respective purposes, the 38 HID- INDELS system showed high discrimination power to Espírito Santo ($PD = 0.9999999999999990$) and to São Paulo ($PD = 0.9999999999999994$); and the 46 AIM- INDELS system confirmed the mixing of the populations studied, and in this case , with greater genetic ancestry of europeans, compared to africans and native americans. In addition, we insert the amelogenin marker in the 38 HID-INDELS multiplex system as a complementary tool to identificate the sex of degraded samples.

Keywords: INDELS. Human identification. Ancestry

SUMÁRIO

1	INTRODUÇÃO	14
1.1	Identificação humana	14
1.2	Análise do genoma humano	14
1.3	<i>Short Tandem Repeats</i> (STRs)	16
1.4	<i>Single Nucleotide Polymorphisms</i> (SNPs)	17
1.5	Polimorfismos de inserção/deleção (INDELs)	18
1.6	Marcadores genéticos de uso forense	20
1.7	Marcadores INDELs para identificação humana (HID-INDELs)	22
1.8	Marcadores INDELs informativos de ancestralidade (AIM-INDELs)	23
2	JUSTIFICATIVA	26
3	OBJETIVO GERAL	29
3.1	Objetivos específicos	29
4	MATERIAIS E MÉTODOS	31
4.1	Aspectos Éticos	31
4.2	Casuística	31
4.3	Extração de DNA	31
4.4	Análise de 38 HID-INDELs (human identification)	32
4.4.1	Implantação da metodologia	32
4.4.2	Reação em cadeia da polimerase (PCR)	34
4.4.3	Purificação	35
4.4.4	Eletroforese capilar	36
4.4.5	Determinação dos perfis alélicos e análise estatística	36
4.5	Inserção do marcador para amelogenina na reação de 38 HID-	37

INDELS	
4.5.1 Seleção dos primers	37
4.5.2 Reação de PCR	38
4.5.3 Purificação e Eletroforese Capilar	38
4.5.4 Análise	38
4.6 Análise de 20 STRs autossômicos	38
4.6.1 Reação em cadeia da polimerase (PCR)	38
4.6.2 Eletroforese Capilar	40
4.6.3 Determinação dos perfis alélicos e análise estatística	40
4.7 Análise de 46 AIM-INDELS (marcadores informativos de ancestralidade)	40
4.7.1 Implantação da metodologia	40
4.7.2 Período sanduíche	43
4.7.3 Reação em cadeia da polimerase (PCR)	43
4.7.4 Purificação	44
4.7.5 Eletroforese capilar	44
4.7.6 Determinação dos perfis alélicos e análise estatística	45
4.7.7 Comparação de 46 AIM-INDELS com cromossomo Y e DNA mt	45
4.8 Fluxograma de trabalho	45
5 RESULTADOS E DISCUSSÃO	48
5.1 Análise de 38 HID-INDELS	48
5.1.1 Implantação e avaliação da técnica	48
5.1.2 Variação genética das populações	51
5.1.3 Avaliação da eficiência forense	53
5.2 Inserção do marcador da amelogenina na reação de 38 HID-INDELS	61
5.3 Comparação entre STRs e 38 HID-INDELS em amostras do	64

estado de São Paulo	
5.4 Análise de 46 AIM-INDELS	67
5.4.1 Implantação e avaliação da técnica	68
5.4.2 Diversidade gênica e distância genética entre populações	71
5.4.3 Proporções de ancestralidade	80
5.4.4 Identificação do alelo 3 nas amostras analisadas	85
5.5 Comparação de dados do cromossomo Y, DNA mt e AIM-INDELS na população do estado de Espírito Santo	86
6 CONCLUSÕES	89
REFERÊNCIAS	90
ANEXO	96

1 INTRODUÇÃO

A identificação humana (*human identification* – HID) é o processo científico pelo qual se determina a identidade de uma pessoa, por meio do estabelecimento de um conjunto de caracteres que a individualize, fazendo-a igual apenas a si mesma (FRANÇA, 2011). Dentre os processos de identificação humana, tem-se o campo da ciência que envolve a análise de marcadores biológicos protéicos e moleculares.

1.1 Identificação humana

O uso de sistemas protéicos apresenta alguns inconvenientes para a aquisição de altos níveis de diferenciação entre os indivíduos, tais como: baixa estabilidade em amostras expostas ao ambiente; baixo poder de discriminação e a ausência dos mesmos marcadores protéicos em todos os tecidos de um mesmo indivíduo (WEEDN; SWARNEN, 1998). A tipagem de polimorfismos genéticos pela análise do DNA contornou essas limitações e revolucionou a ciência voltada à identificação humana.

A variação genética existente entre os indivíduos, uma das principais características das populações humanas, faz com que todos os indivíduos (com exceção de gêmeos monozigóticos) tenham um genoma único.

O estudo das variações genéticas, utilizando os polimorfismos do DNA que são encontrados com ampla distribuição em todo o genoma, permite uma melhor compreensão da história e da diversidade das populações humanas, além de proporcionar um sistema para a identificação genética de indivíduos (BUTLER, 2005).

1.2 Análise do genoma humano

O genoma humano é o conjunto de toda a informação genética existente nas células do organismo; está contido no núcleo das células somáticas em 46 cromossomos (22 pares de cromossomos homólogos e 1 par de cromossomos sexuais X/Y), denominado de DNA nuclear, e em duplas fitas circulares presentes no

interior da organela celular mitocôndria, denominado DNA mitocondrial (DNA mt) (BUTLER, 2005).

O DNA nuclear é biparental, ou seja, metade dos cromossomos é herdada da mãe e a outra metade do pai e possui apenas uma cópia por célula (Figura 1). Nos gametas (óvulo e espermatozóide), a diferença é o número de cromossomos, sendo 23 cromossomos de origem materna, no caso do óvulo, ou de origem paterna, no caso do espermatozóide, assim, em uma análise de investigação de paternidade é possível identificar o verdadeiro pai, pois metade do DNA do filho foi herdado de seu pai.

A palavra “polimorfismo” tem origem grega e significa “muitas formas”, assim, regiões no genoma que apresentam variabilidade superior a 1% são denominados polimorfismos genéticos.

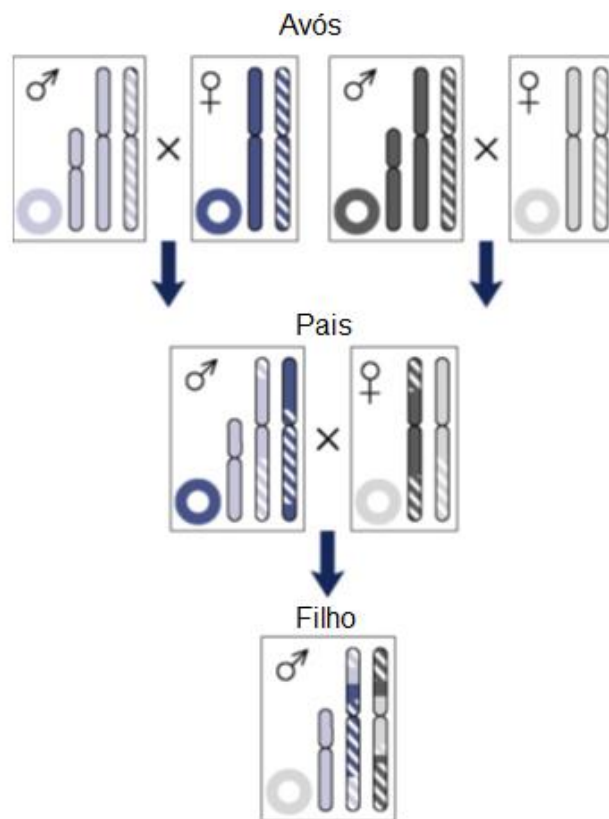
A aplicação do DNA na identificação humana (HID) decorre do seu alto poder de discriminação gerado por seus polimorfismos, que são classificados em dois tipos:

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

Dentre os polimorfismos de comprimento destacam-se os *short tandem repeats* (STRs) e os polimorfismos de inserção/deleção (INDELs), e, dentre os polimorfismos de seqüência os mais frequentes são os *single nucleotide polymorphisms* (SNPs).

Estas formas de variação genética ocorrem em todo o genoma nuclear e algumas também no genoma mitocondrial. Devido ao modo diferente de herança dos cromossomos autossomos, sexuais e das mitocôndrias (Figura 1), a localização de um marcador genético no genoma tem um impacto importante sobre suas características e, conseqüentemente, sobre sua aplicabilidade.

Figura 1 - Ilustração da hereditariedade de segmentos recombinantes e não recombinantes do genoma em três gerações.



Fonte: Pereira, R. (2011).

Cromossomo Y do filho (simbolizado como pequeno cromossomo) que descende do avô paterno e DNA mt (simbolizado como círculo) que descende da avó materna. Em contraste, os cromossomos autossomos (simbolizados como grandes cromossomos) descendem de todos os seus avós, e se recombinam por gerações (JOBILING et al., 2004).

1.3 *Short Tandem Repeats (STRs)*

Os STRs são unidades de repetição do DNA constituídas de 2 a 7 pb que se repetem geralmente de 5 a 30 vezes (LITT et al., 1989), são encontrados em genomas eucariotos e estão amplamente distribuídos no genoma humano (LANDER et al., 2011). As sequências de repetição dos STRs são nomeadas pelo comprimento da unidade de repetição, por exemplo, di-, tri- e tetranucleotídeos. A Sociedade Internacional de Genética Forense (ISFG) recomenda a utilização de tetranucleotídeos para identificação humana, pois estes possuem um grau de

polimorfismo razoável e baixa taxa de mutação em comparação com os outros tamanhos de repetição (MONTEIRO, 2007).

Os STRs tornaram-se marcadores de DNA muito utilizados na identificação humana porque são facilmente amplificados pela reação em cadeia da polimerase (PCR), o número de repetições pode ser altamente variável entre os indivíduos (BUTLER, 2011) e, permitem a utilização de reações em multiplex, que é a amplificação simultânea de múltiplas regiões do DNA, adicionando mais de um par de iniciadores na reação de PCR.

Além disso, a detecção pode ser realizada em sistemas fluorescentes que permitem a automação dos sistemas de eletroforese e a interpretação dos perfis de DNA (BUTLER, 2005).

O FBI (*Federal Bureau of Investigation*) estabeleceu treze marcadores STRs como o conjunto mínimo a ser utilizado em genética forense nos Estados Unidos, sendo este denominado CODIS (*Combined DNA Index System*) que é composto pelos seguintes marcadores: TPOX, D3S1358, FGA, D5S818, CSF1PO, D7S820, D8S1179, TH01, vWA, D13S317, D16S539, D18S51 e D21S11 (BUDOWLE et al., 1998).

1.4 Single Nucleotide Polymorphisms (SNPs)

O SNP (*single nucleotide polymorphism*), que é a substituição de uma única base, é a forma mais simples e mais comum de variação genética (DAWSON et al., 2001).

Um SNP caracteriza-se, por exemplo, pela substituição de uma base “A” (adenina) por “T” (timina) num determinado ponto no genoma.

Pela sua natureza, este tipo de polimorfismo com dois estados alélicos (ancestral e derivado) também pode ser nomeado de polimorfismo binário ou bialélico, onde eventos mutacionais podem ocorrer ao longo de gerações, porém, com uma frequência muito baixa, apresentando alta especificidade geográfica (NACHMAN; CROWELL, 2000).

Apesar da baixa taxa de mutação, os SNPs são distribuídos abundantemente no genoma ocorrendo em média entre 1-2 Kb (CLIFFORD et al., 2000)

Uma simples busca no banco de dados NCBI (*National Center for Biotechnology Information*)-<http://www.ncbi.nlm.nih.gov/SNP>, consulta realizada em 22 de fevereiro de 2016 - resulta em mais de 710 milhões de SNPs no genoma humano.

SNP é o mais estudado tipo de polimorfismo do DNA podendo ter grande impacto sobre a forma como os seres humanos respondem às doenças, fatores ambientais, drogas e outras terapias. Isto gerou um avanço nas tecnologias de genotipagem e aumento de estudos de associação do genoma a diversas características (SOBRINO; BRION; CARRACEDO, 2005; HIRSCHHORN; GAJDOS, 2011).

Uma vez que os SNPs são mais estáveis que os STRs, eles são úteis para o estudo da evolução história das populações humanas em escalas de tempo mais profundas (JAKOBSSON et al., 2008). Na área forense, apesar de apresentar uma diversidade genética menor que os STRs, os SNPs têm várias características desejáveis como marcadores genéticos: podem ser analisados em fragmentos muito curtos, o que é importante para melhorar o sucesso de amplificação em DNA altamente degradado e apresentam baixa taxa de mutação, o que é interessante em testes de parentesco (NACHMAN; CROWELL 2000).

Os SNPs são utilizados na identificação de haplótipos do DNA mitocondrial (DNA mt) e cromossomo Y e, mais recentemente, na análise de amostras altamente degradadas como na identificação de vítimas de desastres, por exemplo, bem como para inferir ascendência biogeográfica de indivíduos e desvendar características fenotípicas (JOBILING; GILL, 2004; BUDOWLE; van DAAL, 2008; KAYSER; de KNIJFF, 2011).

1.5 Polimorfismos de inserção/deleção (INDELs)

Os polimorfismos de inserção/deleção (INDELs) são polimorfismos de comprimento, caracterizados pela inserção ou deleção de um ou mais nucleotídeos em uma determinada região do genoma (PEREIRA, R. et al., 2009); pode-se utilizar como exemplo o polimorfismo rs16363, que apresenta a deleção da sequência "TGTTT" na localização cromossômica 22q13.1.

Weber et al. (2002) foram pioneiros na identificação e caracterização de INDELs e, devido à abundância no genoma humano e a facilidade de análise, destacaram a sua utilidade para estudos genéticos.

Além disso, este trabalho representou o início de um banco de dados online de polimorfismos INDELs Marshfield (<http://www.marshfieldclinic.org/mgs/>), incluindo várias informações sobre os polimorfismos: alelos identificados, diferenças de comprimento, *primers* para amplificação, posição no genoma, bem como as frequências alélicas nos principais grupos populacionais do mundo (africanos, europeus, asiáticos e nativo-americanos). Os INDELs também foram incluídos no banco de dados de variação genética dbSNP (<http://www.ncbi.nlm.nih.gov/projects/SNP/>).

Desde então, alguns estudos foram publicados utilizando INDELs para uma variedade de fins como, por exemplo, abordando a estrutura genética de populações humanas (ROSENBERG et al., 2003; BASTOS-RODRIGUES; PIMENTA; PENA, 2006; TISHKOFF et al. 2009), inferindo proporções de ancestralidade de indivíduos e de populações (YANG et al., 2005) e na identificação de espécies (PEREIRA, F. et al., 2010).

Em 2006, foram identificados novos INDELs, e relatado um mapa inicial contendo mais de 415.000 polimorfismos únicos no genoma humano (MILLS et al., 2006).

Neste mapa, os INDELs representam aproximadamente 16 a 20% de todos os polimorfismos do DNA humano e sendo amplamente difundidos ao longo de todo o genoma, com uma densidade média de um INDEL por 7,2 kb. Cerca de um terço dos INDELs relatados foram identificados em genes conhecidos, a partir dos quais cerca de 3,7% localizados em exons e regiões promotoras, em alguns casos, em sítios funcionalmente relevantes, sendo possível que os INDELs influenciem em traços humanos e doenças (MILLS et al., 2006).

Atualmente, com a utilização da tecnologia de sequenciamento de nova geração (NGS), estima-se que o genoma humano contenha aproximadamente 1 milhão de polimorfismos INDELs (JIANG; TURINSKY; BRUDNO, 2015)

Estes polimorfismos possuem características interessantes para uso na identificação humana: baixa taxa de mutação em comparação com os STRs; são amplamente distribuídos no genoma, como já mencionado; podem ser detectados

por diferentes técnicas (MULLANEY, 2010); o tamanho do fragmento a ser amplificado na PCR é pequeno (50 a 150 pb) o que facilita a análise em amostras com DNA degradado (ZIDKOVA et al., 2011); as frequências alélicas apresentam diferenças significativas entre grupos de populações geograficamente diferentes, sendo potenciais marcadores para estudo de ancestralidade (YANG et al., 2005).

1.6 Marcadores genéticos de uso forense

Atualmente, muitos kits comerciais para análise de STRs estão disponíveis no mercado e são utilizados na rotina de identificação humana, porém, em amostras com alto nível de degradação e contaminação, a análise de STRs pode apresentar dificuldades devido ao tamanho relativamente elevado do fragmento a ser amplificado (100 a 400 pb) obtendo-se perfis genéticos parciais ou a ausência de perfil (ZIDKOVA et al., 2011) (Figura 2 A).

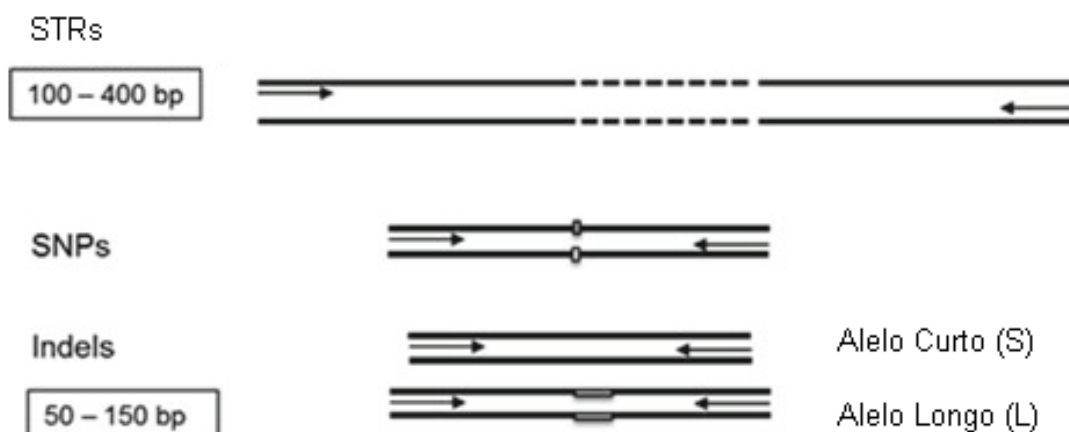
Alguns autores (PHILLIPS et al., 2008) sugerem a análise de SNPs nestes casos pois o fragmento a ser amplificado na PCR costuma ser menor que para STRs, e apresentam baixa taxa de mutação.

Porém, estes polimorfismos apresentam algumas limitações como: o número de SNPs necessários para que se consiga realizar uma análise de parentesco é cerca de quatro vezes maior do que os STRs; a técnica para detecção de SNPs é denominada de mini-sequenciamento, ou reação de SNaPshot, que é o nome do kit comercial utilizado da empresa (Thermo Fisher Scientific), e é diferente das utilizadas na rotina de um laboratório de identificação humana, que analisam rotineiramente STRs, sendo estas últimas mais simples e de menor custo (Figura 2 B).

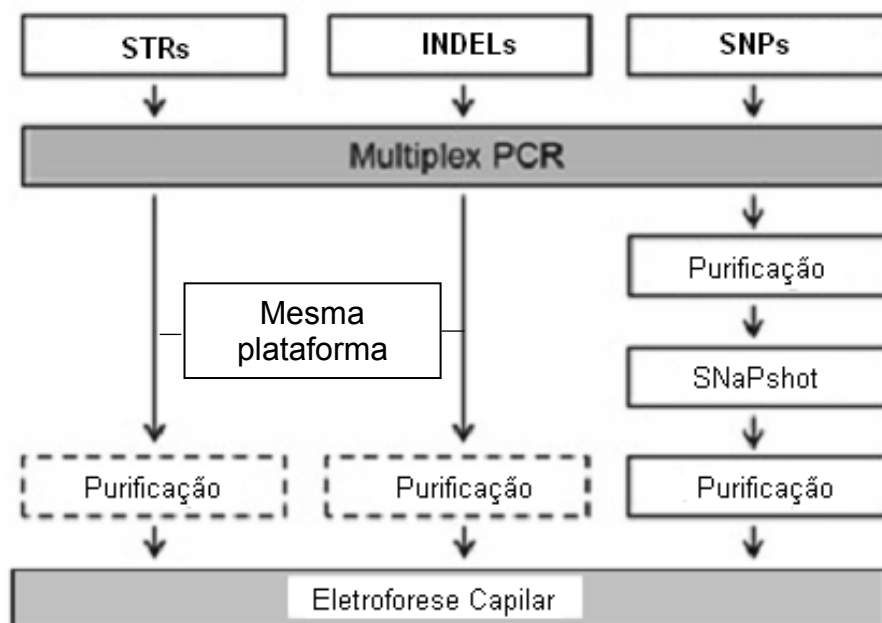
Devido a estas dificuldades, os polimorfismos de inserção-deleção (*insertion-deletion polymorphisms* - INDELs) atraem atualmente a atenção dos cientistas da área de identificação humana como uma nova ferramenta a ser utilizada para acrescentar robustez na análise deste tipo de amostra ou em casos com resultado inconclusivo, devidos às características supramencionadas, especialmente um menor fragmento a ser amplificado com metodologia de análise similar a dos STRs.

Figura 2 - A - Esquema ilustrando o tamanho médio de fragmentos amplificados na PCR para análise de STRs, SNPs e INDELS, sendo que os últimos apresentam tamanho similar. B - Fluxograma para análise de STRs, INDELS e SNPs, sendo que os primeiros são analisados na mesma plataforma no analisador genético ABI 3500 (Thermo Fisher Scientific). Na análise de STRs e INDELS a purificação é opcional.

A



B



Fonte: Pereira e Gusmão (2012).

Neste contexto, algumas metodologias foram padronizadas para análise de marcadores INDELS visando aplicação em identificação humana ou na identificação

de ancestralidade (BASTOS-RODRIGUES; PIMENTA; PENA, 2006; RIBEIRO-RODRIGUES et al., 2009; SANTOS, et al., 2010; PIMENTA; PENA, 2010).

Além destes métodos, atualmente é disponível no mercado o kit comercial DIPplex (Qiagen), para análise de 30 marcadores INDELS autossômicos para aplicação na identificação humana.

Em seguida, serão melhor discutidas duas metodologias de análise de INDELS pois, são as que foram utilizadas no presente trabalho.

1.7 Marcadores INDELS para identificação humana (HID-INDELS)

Pereira, R. et al. (2009) reportaram uma metodologia para análise em *multiplex* de 38 polimorfismos INDELS, distribuídos entre os 22 cromossomos autossômicos e sabidamente polimórficos para populações da África, Europa e Ásia, o que permite sua utilização na identificação humana (HID).

Os autores utilizaram os seguintes critérios para seleção dos marcadores: INDELS bialélicos não localizados em zonas codificadoras; frequência alélica mínima $\geq 0,25$ em europeus, africanos e asiáticos; heterozigosidade média $\geq 0,40$ e, comprimento de variação de alelos de 2-5 pb.

Os autores utilizaram uma reação de PCR *multiplex* sensível aos 38 polimorfismos, seguida de detecção por eletroforese capilar onde todos os fragmentos amplificados são menores que 160 pb utilizando os fluorocromos 6-FAM (azul), VIC (verde), NED (amarelo), e PET (vermelho) (Thermo Fisher Scientific).

A validação da metodologia foi realizada pela análise em 306 indivíduos provenientes da África, Europa e Ásia, não sendo identificado desvio no equilíbrio de Hardy-Weinberg ou associação significativa estatisticamente entre os pares de marcadores, confirmando sua possibilidade de aplicação na identificação de indivíduos. Além disso, obtiveram perfis completos utilizando quantidades de DNA de até 0,3 ng e, ainda, em amostras de DNA degradado, onde a amplificação de STRs falhou parcialmente.

Esta metodologia foi utilizada na análise de indivíduos de populações brasileiras do estado do Rio de Janeiro (incluindo uma amostragem de indivíduos auto-declarados afrodescendentes) e da tribo nativo-americana Terena do Mato Grosso do Sul.

Os dados indicaram que este sistema pode ser utilizado na identificação de indivíduos nativo-americanos e/ou miscigenados, pois, os autores obtiveram alto poder de discriminação para estas amostragens (MANTA et al., 2012a).

1.8 Marcadores INDELS informativos de ancestralidade (AIM-INDELS)

Pereira, R. et al., (2012) reportaram uma metodologia para análise em *multiplex* de 46 marcadores INDELS informativos de ancestralidade (AIM-INDELS), para estimar de forma eficiente as proporções de quatro origens populacionais diferentes (africana, europeia, asiática e nativo-americana).

Os marcadores selecionados apresentam diferentes frequências alélicas entre estes quatro grupos parentais, o que possibilita a identificação de ancestralidade, sendo que, cada marcador apresenta um diferencial de frequência alélica $\geq 0,40$, entre pelo menos dois destes grupos.

Todos os marcadores são analisados em fragmentos menores que 230 pb através de PCR seguida de eletroforese capilar utilizando os fluorocromos 6-FAM (azul), VIC (verde), NED (amarelo), e PET (vermelho) (Thermo Fisher Scientific).

Os autores utilizaram a metodologia na análise de 1002 amostras dos 4 grupos populacionais (europeus, africanos, asiáticos e nativo-americanos), além da Oceania e de uma população miscigenada (Belém-Brasil), e conseguiram classificar as amostras por sua origem continental, e no caso da população miscigenada, avaliar a proporção de ancestralidade de cada grupo parental.

Manta et al., (2012b) utilizaram esta metodologia no estudo de populações das cinco principais regiões brasileiras, norte, nordeste, centro-oeste, sudeste e sul, verificando as proporções de ancestralidade.

Segundo os autores, a metodologia foi eficiente para obtenção das proporções de ancestralidade, entretanto, os dados obtidos para a população estudada do estado do Espírito Santo apresentou uma proporção de ancestralidade inesperada com um nível de ancestralidade europeia muito elevado, similar a de populações do sul do país.

Isto possivelmente poderia ter sido causado, segundo os autores, por um efeito amostral gerado pelo método de seleção dos indivíduos participantes no

estudo, onde as amostras foram coletadas de professores e estudantes de uma instituição de ensino.

JUSTIFICATIVA

2 JUSTIFICATIVA

Em identificação humana e no estudo de populações, marcadores SNP e principalmente STR são mais comumente utilizados. Os INDELS passaram a ser utilizados recentemente apesar de apresentarem características interessantes, já mencionadas. Assim, novos estudos com marcadores INDELS são de grande interesse, visando aumentar o conhecimento das populações, possibilitando sua utilização em genética forense.

Neste trabalho, caracterizamos as populações dos estados de São Paulo e Espírito Santo pela análise de marcadores INDELS, para isto, foram feitas análises de INDELS em amostras provenientes de indivíduos não aparentados dos dois estados por meio das metodologias citadas anteriormente (38 HID-INDELS e 46 AIM-INDELS), sendo as duas populações, caracterizadas com ambas as metodologias.

O uso combinado de marcadores STRs e INDELS pode ser interessante, principalmente quando se deseja aumentar o poder discriminatório, mas para isto, é importante conhecer o desequilíbrio de ligação entre os marcadores na população estudada. Assim, 20 STRs autossômicos foram analisados em amostras do estado de São Paulo, e os dados foram analisados com os obtidos através da análise dos 38 HID-INDELS, avaliando assim o desequilíbrio de ligação entre os STRs e HID-INDELS analisados nesta população. Além disso, comparamos a eficiência forense destes STRs em relação ao sistema 38 HID-INDELS.

Polimorfismos do DNA mt e cromossomo Y possibilitam a classificação em haplogrupos conforme a origem ancestral, materna e paterna, respectivamente. Em estudos anteriores realizados por nosso grupo, amostras de indivíduos do estado de Espírito Santo foram classificadas em haplogrupos evolutivos pela análise de STRs e SNPs do cromossomo Y, e do DNA mt. Estes dados de ancestralidade da população do estado de Espírito Santo, obtidos da análise de marcadores uniparentais, foram comparados com os dados obtidos com os 46 AIM-INDELS, no presente estudo, verificando a relação entre os dados obtidos com diferentes marcadores, uniparentais e autossômicos.

Os INDELS podem ser interessantes para análise de amostras com DNA degradado, pelos motivos já mencionados, e visando uma alternativa de identificação do sexo na análise destas amostras, inserimos o marcador

amelogenina no sistema multiplex 38 HID-INDELS, uma vez que este sistema não apresenta originalmente o marcador.

OBJETIVOS

3 OBJETIVO GERAL

Caracterizar as populações dos estados de São Paulo e Espírito Santo pela análise de polimorfismos INDELS, avaliando a eficiência forense de 38 HID-INDELS em comparação com 20 STRs autossômicos, e comparando os dados de ancestralidade obtidos com 46 AIM-INDELS com os de cromossomo Y e DNA mt.

3.1 Objetivos específicos

- Padronizar a análise de 38 HID-INDELS e 46 AIM-INDELS no analisador genético ABI 3500 (Thermo Fisher Scientific), seguindo os protocolos descritos por Pereira, R. et al. (2009) e Pereira, R. et al. (2012), respectivamente;

- Caracterizar as populações dos estados de São Paulo e Espírito Santo pela análise de 38 HID-INDELS e 46 AIM-INDELS verificando o equilíbrio de Hardy-Weinberg, desequilíbrio de ligação entre os pares de marcadores, as frequências genotípicas esperadas e alélicas observadas, além da distância genética (F_{st});

- Avaliar a eficiência forense do sistema 38 HID-INDELS em relação a 20 STRs autossômicos na população do estado de São Paulo e verificar possível desequilíbrio de ligação entre os pares de marcadores;

- Inserir o marcador para amelogenina no sistema multiplex de análise de 38 HID-INDELS;

- Comparar os dados de ancestralidade obtidos com o sistema 46 AIM-INDELS na população do estado de Espírito Santo com os dados do cromossomo Y e região hipervariável do DNA mt.

MATERIAIS E MÉTODOS

4 MATERIAIS E MÉTODOS

Nesta parte, é apresentada a metodologia usada no presente estudo.

4.1 Aspectos Éticos

Este trabalho foi aprovado pelo Comitê de Ética em Pesquisa da Faculdade de Ciências Farmacêuticas da UNESP de Araraquara com o código: CAAE 05647912.6.0000.5426.

4.2 Casuística

Foram coletadas amostras de sangue de 100 indivíduos (50 homens e 50 mulheres) do estado de São Paulo e 106 (todos homens) do estado de Espírito Santo em papel de filtro FTA Classic (Whatman).

Os indivíduos não são aparentados e são naturais de diversas cidades destes estados. As amostras de Espírito Santo foram coletadas no Centro de Hemoterapia e Hematologia do Espírito Santo – HEMOES. As amostras de São Paulo foram cedidas de casos de investigação de paternidade do Laboratório de Investigação de Paternidade da UNESP de Araraquara.

Os indivíduos coletados foram selecionados com aleatoriedade, visando à representação dos estados a partir destas amostragens.

4.3 Extração de DNA

A extração de DNA das amostras foi realizada com o auxílio da resina Chelex 100 (Biorad) (SINGER-SAM; TANGUAY; RIGGS, 1989) conforme segue:

1 disco de papel FTA Classic (Whatman) de 1,2 mm com sangue foi cortado com o auxílio do *Harris Micro Punche e Cutting Mat* (Whatman) e depositado em microtubo de 1,5 mL, no qual adicionou-se 50 µL de H₂O MILLI-Q, vortexou-se rapidamente e descartou-se a H₂O. Adicionou-se novamente 50 µL de H₂O MILLI-Q e incubou-se por 30 minutos à temperatura ambiente (TA). Posteriormente, centrifugou-se a

13.000 rpm por 1 minuto e descartou-se a H₂O. Adicionou-se 100 µL de Chelex 5%, incubando-se a 56 °C por 1 hora. O tubo foi vortexado, incubou-se por 100 °C por 8 minutos, centrifugou-se a 13.000 rpm por 3 minutos e transferiu-se o sobrenadante para um novo microtubo, o qual foi armazenado a 4 °C.

4.4 Análise de 38 HID-INDELS (*human identification*)

A análise de 38 HID-INDELS autossômicos foi feita em todas as amostras dos estados de São Paulo e Espírito Santo pela metodologia descrita por Pereira, R. et al. (2009).

4.4.1 Implantação da metodologia

A metodologia descrita por Pereira, R. et al. (2009) para a análise de 38 HID-INDELS autossômicos (Tabela 1) utiliza uma reação de PCR seguida por análise em eletroforese capilar (EC). Os autores utilizaram o analisador genético ABI 3130 (Thermo Fisher Scientific) com polímero POP 7 (Thermo Fisher Scientific) para padronização da metodologia. Como em nosso laboratório utiliza-se outro modelo de analisador genético - ABI 3500 (Thermo Fisher Scientific) com polímero POP 4 (Thermo Fisher Scientific) - houve a necessidade de ajustar o painel de *bins* (região em que os alelos são identificados no eletroferograma) para a correta identificação dos alelos no software GeneMapper ID-X v. 1.2 (Thermo Fisher Scientific). Para isto, foram genotipadas as amostras controle 9947A (Thermo Fisher Scientific) e 9948 (Promega) e as amostras M1, M2, M3 e M4 do exercício de intercomparação do ano de 2012 do GHEP-ISFG (Grupo de Línguas Espanhola e Portuguesa da Sociedade Internacional de Genética Forense). Estes dados foram enviados para o grupo de trabalho do GHEP para avaliação e após a padronização, a técnica foi utilizada para a genotipagem dos 38 HID-INDELS nas amostras dos estados de São Paulo e Espírito Santo.

Nos resultados, serão mostradas as diferenças entre as posições de leitura dos alelos no eletroferograma para o ABI 3130 e o ABI 3500 (Thermo Fisher Scientific).

Tabela 1 - INDELS analisados mostrando a localização de cada marcador no genoma e os alelos possíveis.

Marcador	Cromossomo	Posição (pb)	Alelos
rs3047269	1	161077452	-/CTGA
rs2307579	1	245878706	-/ATG
rs16624	2	234681130	-/GT
rs2308242	3	8591709	-/CT
rs2308026	4	119404855	-/CA
rs2307526	5	5178112	-/ACAC
rs1160956	5	65414216	-/AGA
rs1610871	5	171020572	-/TAGG
rs2307710	6	47929222	-/AGGA
rs2307839	6	117200251	-/GA
rs2308137	6	149655891	-/GA
rs2307978	7	83121850	-/GA
rs35769550	8	76681235	-/TGAC
rs5895447	8	138489776	-/CA
rs16402	9	38396788	-/TTAT
rs2067294	9	70504241	-/CTT
rs2307580	9	104626014	-/AATT
rs140809	10	6027167	-/CAA
rs1160886	10	54112392	-/ACT
rs10688868	11	258180	-/CT
rs34811743	11	30134266	-/TG
rs33972805	11	125794082	-/CT

rs1610919	12	14801263	-/AT
rs2067238	12	113772931	-/GCT
rs2308171	13	43778155	-/TCTG
rs2308189	14	28106508	-/AACTA
rs2308020	15	51268809	-/TT
rs2067208	16	83139788	-/GCCAG
rs3051300	17	10076666	-/GTAT
rs3080855	18	21507205	-/AATT
rs34511541	18	34677042	-/CTCTT
rs36040336	19	1353662	-/AT
rs2307689	19	48896180	-/TTC
rs33917182	20	11643625	-/CA
rs34541393	20	30165066	-/AACT
rs35605984	21	14556736	-/TAAAG
rs10629077	21	30294208	-/AT
rs2307700	22	25120901	-/TCAC

Dados de acordo com o *build* 129 do dbSNP.

Fonte: Pereira, R. et al. (2009).

4.4.2 Reação em cadeia da polimerase (PCR)

As amostras de DNA foram submetidas à reação de amplificação por PCR dos 38 HID-INDELS (PEREIRA, R. et al. 2009), conforme segue:

- 5 µL de Qiagen Multiplex PCR master *mix* (2x)
- 1 µL de Primer *mix* (10x)
- 1 µL de DNA (0.3–5 ng/ µL)
- H₂O ultrapura (MILLI-Q) autoclavada q.s.p. 10 µL

Obs. O mix de primers contém todos os primers na concentração de 1 μM exceto para os marcadores rs2308137 (2 μM) e rs3047269 (3 μM).

A ciclagem foi realizada em termociclador Veriti (Thermo Fisher Scientific) nas seguintes condições:

95°C - 15 min		
94°C - 30 s	}	10 ciclos
60°C - 90 s		
72°C - 60 s		
94°C - 30 s	}	18 ciclos
58°C - 90 s		
72°C - 60 s		
72°C - 80 min		
4°C - até a retirada do termociclador.		

Obs. Esta ciclagem apresenta duas modificações em comparação com aquela publicada pelos autores. O segundo passo de ciclagens foi diminuído de 20 para 18 ciclos para reduzir o sinal obtido no eletroferograma e o tempo de extensão final foi aumentado em 20 minutos para melhorar a adenilação dos fragmentos.

4.4.3 Purificação

Os autores citam que a purificação do produto de PCR previamente à EC é opcional, mas neste trabalho foi realizada pela técnica de precipitação por álcool conforme segue:

- Adicionou-se 1 μL do produto da PCR em um tubo de 1,5 mL contendo 80 μL de isopropanol 75%;
- A mistura foi deixada à temperatura ambiente (22°C) por 15 minutos no escuro, centrifugada por 15 minutos a 14.000 G (Eppendorf Centrifugal 5417R) e 22°C;
- O sobrenadante foi então cuidadosamente descartado, invertendo-se o tubo sobre papel absorvente;
- Adicionou-se 1 mL de etanol 70%;

- Procedeu-se nova centrifugação por 5 minutos, 14.000 rpm à 22°C, e o sobrenadante foi vertido cuidadosamente sobre papel absorvente;
- O precipitado foi seco a vácuo por 15 minutos em aparelho DNA Speed Vac (Savant) e guardado em freezer a -20°C até a EC.

4.4.4 Eletroforese capilar

As amostras foram preparadas da seguinte forma: adicionou-se 9,7 µL de formamida HI-DI (Thermo Fisher Scientific) e 0,3 µL de LIZ 600 (Thermo Fisher Scientific) ao tubo de 1,5 mL contendo o produto de PCR purificado e seco. O volume foi agitado para ressuspender o produto de PCR e em seguida transferido para placa de 96 poços.

Em seguida, as amostras foram desnaturadas a 95 °C por 3 minutos e rapidamente inseridas no gelo onde permaneceram também por 3 minutos.

A eletroforese capilar foi realizada no analisador genético ABI 3500 (Thermo Fisher Scientific), utilizando capilar de 36 cm e polímero POP 4.

4.4.5 Determinação dos perfis alélicos e análise estatística

Os resultados foram analisados no software GeneMapper ID-X v. 1.2 (Thermo Fisher Scientific).

Com o auxílio do software Arlequin v. 3.11 (EXCOFFIER; LAVAL; SCHNEIDER; 2005), foram estimadas as frequências alélicas para as populações, o teste exato para avaliar o desequilíbrio de ligação (LD) entre os pares de marcadores, a divergência do equilíbrio de Hardy-Weinberg, a heterozigosidade observada (H_o) e a heterozigosidade esperada (H_e), o cálculo de distância genética (F_{st}) e os valores de p correspondentes para as populações estudadas e as de Portugal, África, Ásia, tribo Terena e do estado de Rio de Janeiro. Os valores de F_{st} foram utilizados para representação gráfica em escala multidimensional (MDS plot) por meio do software STATISTICA v. 13 (Statsoft). Com o auxílio da tabela PowerStats (Promega) foram calculados os seguintes parâmetros forense: poder de discriminação (PD) e poder de exclusão (PE).

4.5 Inserção do marcador para amelogenina na reação de 38 HID-INDELS

No sistema 38 HID-INDELS estabelecido por Pereira, R. et al. (2009) há um espaço entre 103 e 120 pb para fluorocromo NED (amarelo) no eletroferograma, onde o marcador da amelogenina poderia ser inserido.

4.5.1 Seleção dos primers

Após busca na literatura, selecionou-se o par de primers descritos em Krenke et al. (2002), pois os fragmentos amplificados têm 108 e 112 pb para os alelos X e Y, respectivamente. A ferramenta *online* PCR *in silico* (<http://genome.ucsc.edu/>) foi utilizada para confirmação dos tamanhos dos fragmentos. *Observação: Para as populações de Espírito Santo e São Paulo estudadas neste trabalho este sistema multiplex com o marcador da amelogenina não foi utilizado.

4.5.2 Reação de PCR

Inicialmente, o par de primers foi testado individualmente na PCR para verificar sua eficiência e, posteriormente, utilizado em conjunto com o *mix* de primers para os 38 HID-INDELS conforme descrito no item 4.4.2. O *mix* de primers da amelogenina foi utilizado na concentração de 2 μ M.

4.5.3 Purificação e Eletroforese Capilar

A purificação do produto de PCR e a EC foram realizadas conforme descrito nos itens 4.4.3. e 4.4.4., respectivamente.

4.5.4 Análise

Os resultados foram analisados no software GeneMapper ID-X v. 1.2 (Thermo Fisher Scientific).

4.6 Análise de 20 STRs autossômicos

O kit PowerPlex 21 (Promega), que analisa 20 STRs autossômicos mais o marcador da amelogenina (Tabela 2), foi utilizado na análise de 100 amostras de indivíduos não aparentados do estado de São Paulo.

4.6.1 Reação em cadeia da polimerase (PCR)

As PCRs foram feitas com volume total de 12,5 µL (metade do volume indicado no manual do fabricante), utilizando-se 1 disco de papel FTA Classic (Whatman) com a amostra de sangue adsorvida, sem a necessidade de extração de DNA, conforme segue:

- 2,5 µL de Master *mix* (5X)
- 2,5 µL de Primer *mix* (5X)
- 1 disco (1,2 mm) do cartão FTA Classic (Whatman) com sangue
- H₂O ultrapura (MILLI-Q) autoclavada q.s.p. 12,5 µL

A ciclagem foi realizada conforme as seguintes condições em termociclador Veriti (Thermo Fisher Scientific):

96°C - 1 min	
94°C - 10 s	} 25 ciclos
59°C - 1 min	
72°C - 30 s	
60°C - 20 min	
4°C - até a retirada do termociclador	

Tabela 2 - STRs analiados com o kit PowerPlex 21 (Promega), localização cromossômica e sequência repetidora do polimorfismo. NA: não se aplica.

STR	Cromossomo	Sequência repetidora 5' → 3'
Amelogenina	X e Y	NA
D3S1358	3	TCTA
D1S1656	1	TAGA
D6S1043	6	AGAT
D13S317	13	TATC
Penta E	15	AAAGA
D16S539	16	GATA
D18S51	18	AGAA
D2S1338	2	TGCC/TTCC
CSF1PO	5	AGAT
Penta D	21	AAAGA
TH01	11	AATG
Vwa	12	TCTA
D21S11	21	TCTA
D7S820	7	GATA
D5S818	5	AGAT
TPOX	2	AATG
D8S1179	8	TCTA
D12S391	12	AGAT/AGAC
D19S433	19	AAGG
FGA	4	TTTC

Fonte: PowerPlex 21 (Promega) manual.

4.6.2 Eletroforese Capilar

As amostras foram preparadas em placas de 96 poços da seguinte forma:

- 1 μ L do produto de PCR
- 10 μ L de formamida HI-DI (Thermo Fisher Scientific)
- 1 μ L de CC5-ILS 500 (Promega)

Em seguida, as amostras foram desnaturadas a 95 °C por 3 minutos e rapidamente inseridas no gelo onde permaneceram também por 3 minutos.

A eletroforese capilar foi realizada no analisador genético ABI 3500 (Thermo Fisher Scientific) utilizando capilar de 36 cm e polímero POP 4.

4.6.3 Determinação dos perfis alélicos e análise estatística

Os resultados foram analisados no software GeneMapper ID-X v. 1.2 (Thermo Fisher Scientific).

Com o auxílio do software Arlequin v. 3.11 (EXCOFFIER; LAVAL; SCHNEIDER; 2005), foi realizado o teste exato para avaliar o desequilíbrio de ligação entre os pares de marcadores. A tabela PowerStats (Promega) foi usada para calcular o poder de discriminação (PD) e o poder de exclusão (PE).

4.7 Análise de 46 AIM-INDELs (marcadores informativos de ancestralidade)

A metodologia descrita por Pereira, R. et al. (2012) visa a análise de 46 AIM-INDELs (marcadores informativos de ancestralidade) autossômicos (Tabela 3) por meio de reação de PCR seguida por análise em eletroforese capilar (EC).

4.7.1 Implantação da metodologia

O ajuste desta metodologia em nosso laboratório foi necessário pelos mesmos motivos descritos no item 4.4.1. Para isto, foram genotipadas as amostras controle 9947A (Thermo Fisher Scientific) e 9948 (Promega) e as amostras M1, M2,

e M3 do exercício de intercomparação do ano de 2014 do GHEP-ISFG. Estes dados foram enviados para o grupo de trabalho do GHEP para avaliação.

Nos resultados, serão mostradas as diferenças entre as posições de leitura dos alelos no eletroferograma para o ABI 3130 e o ABI 3500 (Thermo Fisher Scientific).

Tabela 3 - INDELs analisados com a nomenclatura segundo o banco de dados Marshfield (MID) e dbSNP, sua localização no genoma e os alelos possíveis.

MID	dbSNP	Cromossomo	Posição (pb)	Alelos
MID-1470	rs2307666	11	64729920	-/GTTAC
MID-777	rs1610863	16	6551830	-/GAA
MID-196	rs16635	6	99789775	-/CAT
MID-881	rs1610965	5	79746093	-/ACTT
MID-3122	rs35451359	18	45110983	-/ATCT
MID-548	rs140837	6	3708909	-/CT
MID-659	rs1160893	2	224794577	-/CT
MID-2011	rs2308203	2	109401291	-/CTAGA
MID-2929	rs33974167	8	87813725	-/TA
MID-593	rs1160852	6	137345857	-/TT
MID-798	rs1610884	5	56122323	-/GGGAAA
MID-1193	rs2067280	5	89818959	-/AT
MID-1871	rs2308067	7	127291541	-/TT
MID-17	rs4183	3	3192524	-/TAAC
MID-2538	rs3054057	15	86010538	-/AACA
MID-1644	rs2307840	1	36099090	-/GT
MID-3854	rs60612424	6	84017514	-/TCTA

MID-2275	rs3033053	14	42554496	-/TCAGCAG
MID-94 r	rs16384	22	42045009	-/AAC
MID-3072	rs34611875	18	67623917	-/GCCCCCA
MID-772	rs1610859	5	128317275	-/TAG
MID-2313	rs3045215	1	234740917	-/ATTATAACT
MID-397	rs25621	6	139858158	-/TTCT
MID-1636	rs2307832	1	55590789	-/AA
MID-51 r	rs16343	4	17635560	-/TTTAT
MID-2431	rs3031979	8	73501951	-/ATTG
MID-2264	rs34122827	13	63778778	-/AAGT
MID-2256	rs133052	22	41042364	-/CAT
MID-128	rs6490	12	108127168	-/ATT
MID-15	rs4181	2	42577803	-/AAATACACAC
MID-2241	rs3030826	6	67176774	-/GTCCAATA
MID-419	rs140708	6	170720016	-/AATGGCA
MID-943	rs1611026	5	82545545	-/TGAT
MID-159	rs16438	20	25278470	-/CCCCA
MID-2005	rs2308161	10	69800909	-/AACAAT
MID-250	rs16687	7	83887882	-/CA
MID-1802	rs2307998	5	7814345	-/GGA
MID-1607	rs2307803	3	108981031	-/TG
MID-1734	rs2307930	6	84476378	-/CCAT
MID-406	rs25630	6	14734341	-/AG

MID-1386	rs2307582	1	247768775	-/AAACTATTCATTTTTTCACCCT
MID-1726	rs2307922	1	39896964	-/CAAGAACTATAAT/CACTATCTATTAT
MID-3626	rs11267926	15	45526069	-/AATATAATTTCTCCA
MID-360	rs25584	12	112145217	-/AA
MID-1603	rs2307799	5	70828427	-/TTGT
MID-2719	rs34541393	20	30701405	-/AACT

Dados de acordo com o *build* 132 do dbSNP.

Fonte: Pereira, R. et al. (2012).

4.7.2 Período do doutorado sanduíche

Durante os meses de junho a novembro de 2015, realizou-se um estágio no Instituto de Patologia e Imunologia Molecular da Universidade do Porto (IPATIMUP) – Porto, Portugal, sob a supervisão do Dr. Rui Pereira, onde foi feita a análise dos 46 AIM-INDELS para as duas populações estudadas, conforme descrito nos itens seguintes (4.7.3. a 4.7.7.).

4.7.3 Reação em cadeia da polimerase (PCR)

As PCRs para amplificação dos 46 AIM-INDELS foram realizadas conforme descrito por Pereira, R. et al. (2012):

- 5 µL de Qiagen Multiplex PCR master mix (2x)
- 1 µL de Primer *mix* (10x)
- 1 µL de Primer *mix* (reforço)
- 1 µL de DNA (0.3–5 ng/ µL)
- H₂O ultrapura (MILLI-Q) autoclavada q.s.p. 10 µL

Obs. O *mix* de primers contém todos os primers na concentração de 1 µM.

O Primer *mix* (reforço) contém os primers a 1 μ M para os marcadores MIDs-196, 881, 3854, 772, 3626 e 2719.

A ciclagem foi realizada em termociclador Veriti (Thermo Fisher Scientific) conforme as seguintes condições:

95°C - 15 min	}	30 ciclos
94°C - 30 s		
60°C - 90 s		
72°C - 45 s		
72°C - 60 min		
4°C - até a retirada do termociclador.		

4.7.4 Purificação

Conforme os autores, o passo de purificação é opcional e não foi utilizado, pois não foi necessário.

4.7.5 Eletroforese capilar

As amostras foram preparadas em placas de 96 poços da seguinte forma:

- 1 μ L do produto de PCR
- 9,7 μ L de formamida HI-DI (Thermo Fisher Scientific)
- 0,3 μ L de LIZ 600 (Thermo Fisher Scientific)

Em seguida, as amostras foram desnaturadas a 95 °C por 3 minutos e rapidamente inseridas no gelo onde permaneceram também por 3 minutos.

A eletroforese capilar foi realizada no analisador genético ABI 3130 (Thermo Fisher Scientific) utilizando capilar de 36 cm e polímero POP 7.

4.7.6 Determinação dos perfis alélicos e análise estatística

Os resultados foram analisados no software GeneMapper v. 4.0 (Thermo Fisher Scientific).

Com o auxílio do software Arlequin v. 3.11 (EXCOFFIER; LAVAL; SCHNEIDER; 2005), foram estimadas as frequências alélicas para as populações, o teste exato para avaliar o desequilíbrio de ligação (LD) entre os pares de marcadores, a divergência do equilíbrio de Hardy-Weinberg, a heterozigosidade observada (H_o) e a heterozigosidade esperada (H_e), o cálculo de distância genética (F_{st}) e os valores de p correspondentes para as populações estudadas e da Europa, África, nativo-americanos (PEREIRA, R. et al., 2012) e algumas populações brasileiras (MANTA et al., 2012b). Os valores de F_{st} foram utilizados para representação gráfica em escala multidimensional (MDS plot) através do software STATISTICA v. 13 (Statsoft). Com o auxílio do software STRUCTURE v. 2.3.4 (FALUSH; STEPHENS; PRITCHARD, 2003) foi realizada a comparação para proporção de ancestralidade das populações de Espírito Santo e São Paulo em comparação com africanos, europeus e nativo-americanos, além da plotagem destes dados.

4.7.7 Comparação de 46 AIM-INDELS com cromossomo Y e DNA mt

As amostras provenientes de indivíduos do estado de Espírito Santo em outros estudos realizados no nosso laboratório (FIGUEIREDO et al., 2015; SANCHES et al., 2014) foram classificadas em haplogrupos evolutivos por meio da análise de Y-SNPs, Y-STRs, e região hipervariável do DNA mt. As proporções de origem ancestral obtidas com as análises de regiões uniparentais foram comparadas com aquela obtida com os 46 AIM-INDELS.

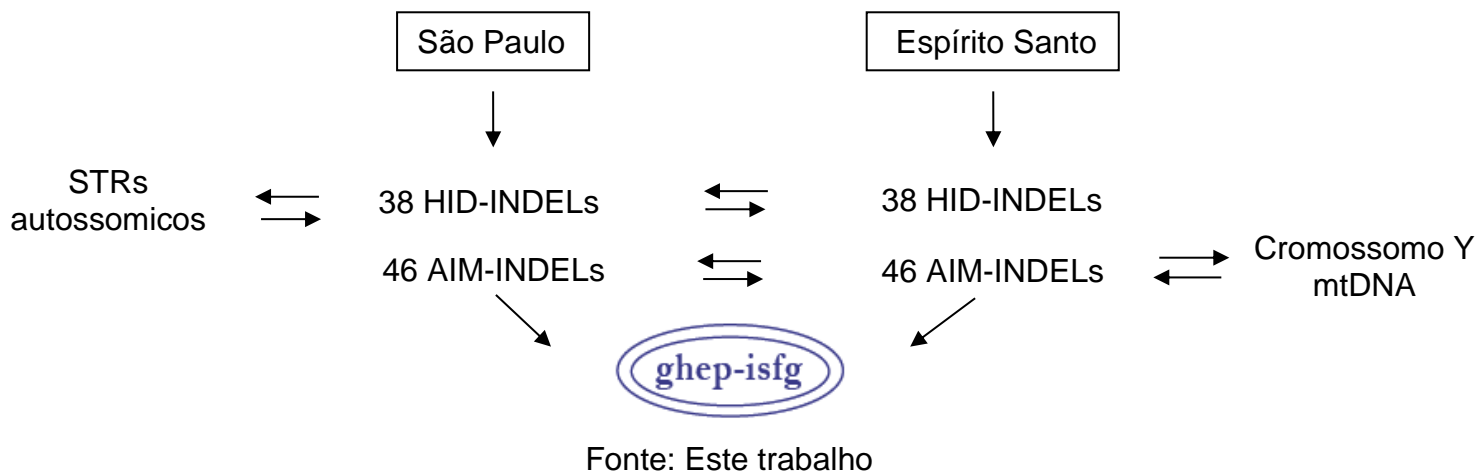
4.8 Fluxograma de trabalho

Os dados obtidos para os 38 HID-INDELS e 46 AIM-INDELS foram comparados entre as populações dos estados de São Paulo e Espírito Santo, sendo que estes dados foram submetidos a duas comissões de trabalho interlaboratorial do

Grupo de Línguas Espanhola e Portuguesa da Sociedade Internacional de Genética Forense (GHEP-ISFG). Além disso, para a população de São Paulo os dados dos 38 HID-INDELs foram comparados com STRs autossômicos e para a população do Espírito Santo os dados dos AIM-INDELs foram comparados com dados de polimorfismos do cromossomo Y e região hipervariável do DNA mt (Figura 3).

Os dados de 38 HID-INDELs e 46 AIM-INDELs serão publicados em conjunto com dados de outras populações mundiais enviadas pelos laboratórios participantes do exercício colaborativo.

Figura 3 - Esquema ilustrando o fluxograma de trabalho com os dados das análises que foram utilizadas nas duas populações.



RESULTADOS E DISCUSSÃO

5 RESULTADOS E DISCUSSÃO

Os resultados obtidos no presente trabalho são apresentados e comparados com dados existentes na literatura, á seguir.

5.1 Análise de 38 HID-INDELS

Nesta parte, são discutidos os dados obtidos da análise de 38 HID-INDELS nas populações dos estados de São Paulo e Espírito Santo.

5.1.1 Implantação e avaliação da técnica

Houve a necessidade de ajustar o protocolo descrito por Pereira, R. et al. (2009) devido à diferença de mobilidade em pares de base (pb) dos fragmentos amplificados de cada marcador quando analisados no ABI 3130 (Thermo Fisher Scientific), o qual foi o utilizado pelos autores, e no ABI 3500 (Thermo Fisher Scientific), utilizado neste trabalho.

Esta diferença de mobilidade ocorre, pois, no equipamento ABI 3130 (Thermo Fisher Scientific) os autores utilizaram polímero POP-7, enquanto no presente trabalho, foi utilizado o ABI 3500 (Thermo Fisher Scientific) com polímero POP-4, uma vez que este equipamento não permite a utilização do polímero POP-7 para a análise de fragmentos.

As diferentes posições de leitura em pb são mostradas a seguir (Tabela 4):

Tabela 4 - Comparação das posições em pares de base (pb) no eletroferograma dos alelos de cada um dos 38 marcadores com os respectivos fluorocromos no ABI 3130 e no ABI 3500 (Thermo Fisher Scientific). 1 (alelo curto) e 2 (alelo longo).

Marcador	Fluorocromo	ABI 3130 (polímero POP-7)		ABI 3500 (polímero POP-4)	
		Pereira, R. et al, (2009)		Este trabalho	
		1 (pb)	2 (pb)	1 (pb)	2 (pb)
B01	Azul	55,43	58,91	47,17	51,28
B02	Azul	61,19	63,06	55,13	57,26
B03	Azul	69,9	72,93	65,01	68,36
B04	Azul	86,83	91,15	82,96	87,21
B05	Azul	99,38	103,03	96,05	99,67
B06	Azul	112,17	115,5	109,21	112,45
B07	Azul	124,58	128,77	122,27	126,36
B08	Azul	133,68	135,78	130,41	132,46
B09	Azul	141,49	143,61	139,18	141,14
B10	Azul	149,09	153,27	145,13	149,17
G01	Verde	59,14	63,11	52,9	57,72
G02	Verde	70,13	73,3	65,04	68,79
G03	Verde	77,08	80,07	73,95	77,11
G04	Verde	91,15	95,17	87,05	91,21
G05	Verde	105,66	107,69	101,79	103,82
G06	Verde	120,13	124,19	116,33	120,22
G07	Verde	128,21	131,12	124,41	127,96
G08	Verde	141,92	147,55	137,58	142,71
G09	Verde	157,67	159,62	151,73	153,99

Y01	Amarelo	60,69	64,31	53,19	58,31
Y02	Amarelo	73,96	76,11	68,24	70,77
Y03	Amarelo	82,77	84,93	78,17	80,29
Y04	Amarelo	92,39	97,66	88,92	93,91
Y05	Amarelo	102,24	105,14	98,05	101,25
Y06	Amarelo	124,98	127,02	121,25	123,2
Y07	Amarelo	132,21	136,36	128,08	132,16
Y08	Amarelo	141,76	144,05	137,09	139,1
Y09	Amarelo	151,2	153,38	145,99	148,08
R01	Vermelho	63,15	64,966	57,24	59,41
R02	Vermelho	67,66	69,48	62,81	64,93
R03	Vermelho	78,9	81,89	74,49	77,63
R04	Vermelho	87,23	89,14	83,58	85,64
R05	Vermelho	95,5	100,08	91,78	96,38
R06	Vermelho	109,28	111,34	106,04	108,21
R07	Vermelho	119,42	124,12	117,74	122,53
R08	Vermelho	129,38	131,5	126,53	128,59
R09	Vermelho	135,78	140,11	133,32	137,52
R10	Vermelho	154,79	160,01	150,17	155,16

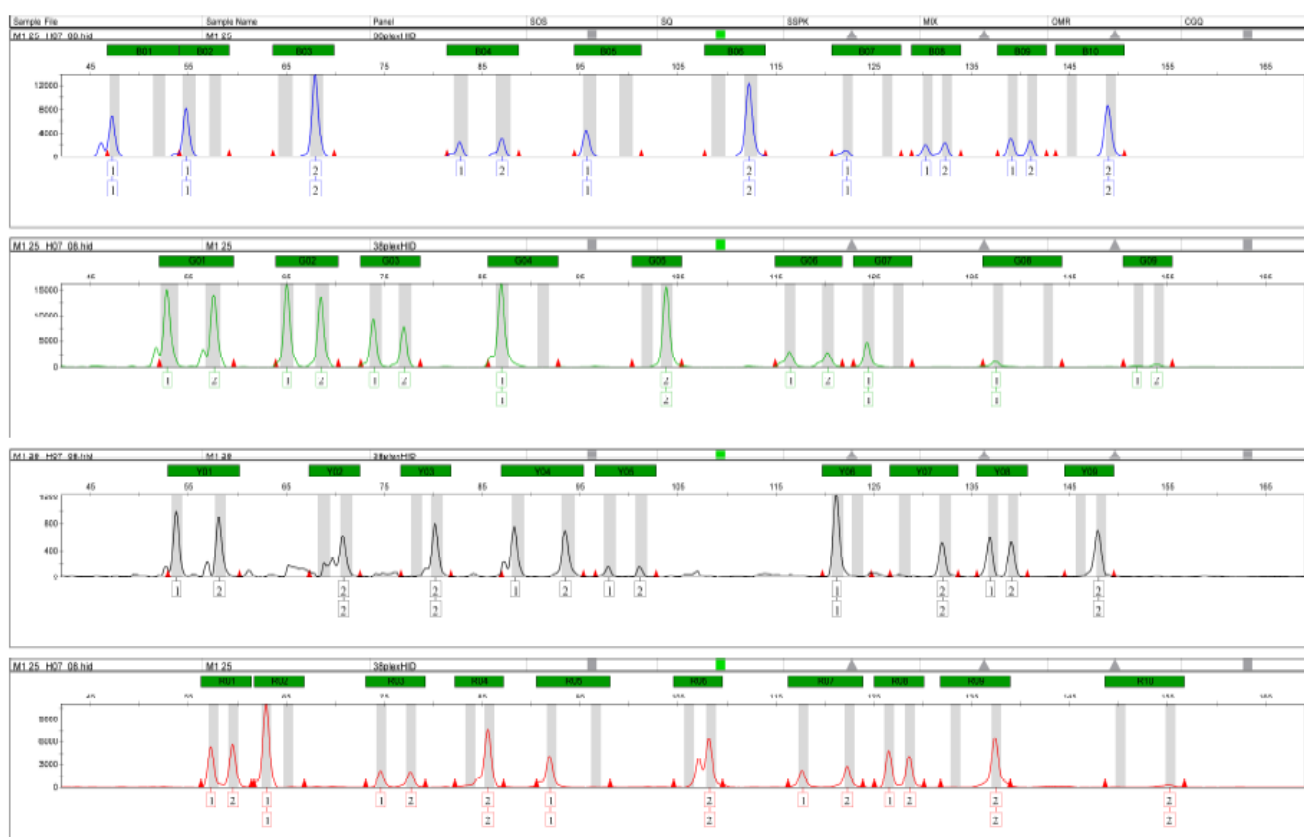
Fonte: Este trabalho.

O diferencial de mobilidade observado na tabela 4 foi causado não só por usarmos um equipamento diferente, mas principalmente, pela utilização de um tipo diferente de polímero, uma vez que o ABI 3500 (Thermo Fisher Scientific) não permite o uso do polímero POP-7 para análise de fragmentos.

Após o ajuste de leitura para a correta identificação dos alelos, verificamos que a reação de amplificação foi eficiente para análise dos 38 INDELS.

Os perfis dos indivíduos de cada população foram avaliados no software GeneMapper ID-X v. 1.2 (Thermo Fisher Scientific) (Figura 4).

Figura 4 - Eletroferograma de uma reação multiplex dos 38 HID-NDELS analisada no ABI 3500 (Thermo Fisher Scientific). O eixo Y indica a intensidade de fluorescência em RFU (*relative fluorescence units*) e o eixo X, mostra a posição do fragmento em pares de base.



Fonte: Este trabalho

5.1.2 Variação genética das populações

Na análise de estimativa de distância genética (F_{st}) entre as populações estudadas de São Paulo e Espírito Santo, o valor obtido foi $F_{st} = 0,00060$ ($p = 0,35135$), indicando que as populações são bastante próximas geneticamente.

Além disso, os dados do Espírito Santo (ES) e de São Paulo (SP) foram comparados com dados dos 38 HID-INDELS das populações de Portugal, África

(AFR), Ásia (EAS), tribo Terena do Mato Grosso do Sul (nativo-americanos) e Rio de Janeiro (RJ) (MANTA et al., 2012a) para obtenção das distâncias genéticas (Fst) que são mostradas na tabela 5.

Os dados indicam uma proximidade genética entre as populações dos estados de São Paulo, Espírito Santo e Rio de Janeiro (MANTA et al., 2012a). O maior valor Fst obtido (indicando maior distância genética) foi entre as populações de África e Terena (0,14128) e o menor entre Espírito Santo e Rio de Janeiro (0,00059).

Tabela 5 - Distâncias genéticas entre as populações (abaixo da diagonal) e valores de p correspondentes (acima da diagonal) obtidos com os dados de 38 HID-INDELS.

	Portugal	AFR	EAS	Terena	RJ	ES	SP
Portugal	*	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
AFR	0,10758	*	0,00000	0,00000	0,00000	0,00000	0,00000
EAS	0,09066	0,12176	*	0,00000	0,00000	0,00000	0,00000
Terena	0,09066	0,14128	0,09603	*	0,00000	0,00000	0,00000
RJ	0,01086	0,05737	0,06901	0,06287	*	0,27928	0,08108
ES	0,01208	0,05441	0,07156	0,06691	0,00059	*	0,35135
SP	0,00464	0,07490	0,07034	0,06728	0,00136	0,00060	*

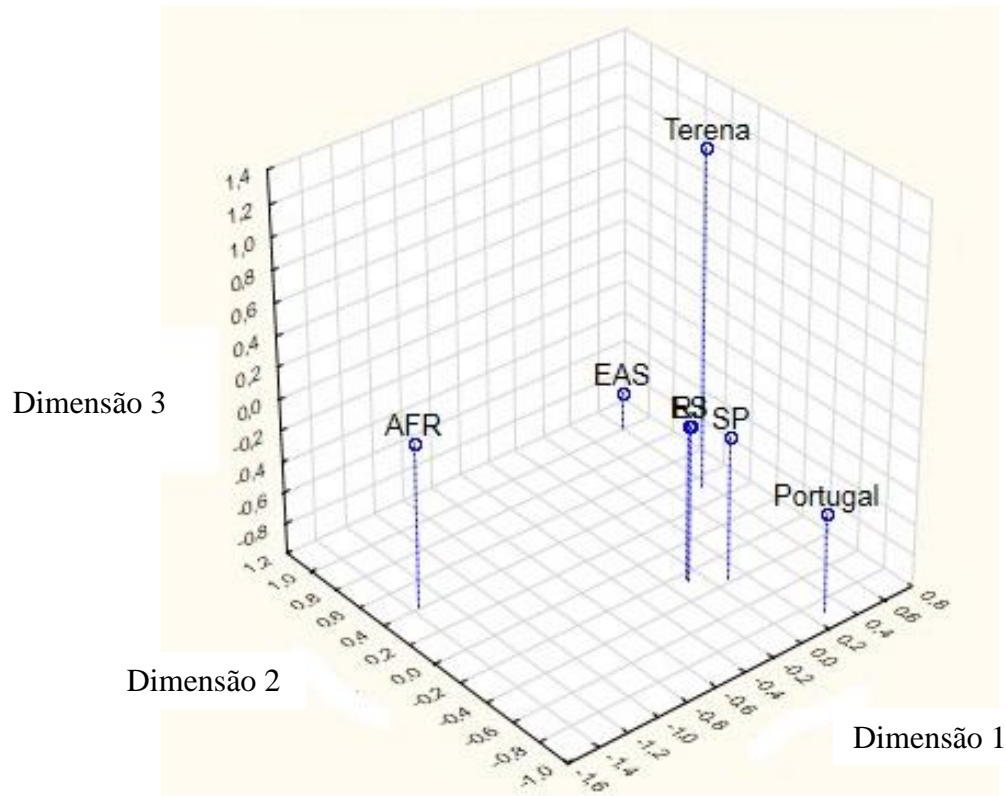
Fonte: Este trabalho.

Um detalhe interessante a ser observado na tabela 5, é que a amostragem populacional de Terena apresentou uma menor distância em relação à amostragem de Portugal, do que em relação as amostragens da África e Ásia.

Apesar de Terena ser uma tribo indígena localizada no estado de Mato Grosso do Sul, este dado pode ser um indicativo da presença de uma componente genética de origem europeia, inserida nesta população indígena na época em que o Brasil foi colônia de Portugal.

Com os valores de distância genética foi realizado um MDS plot tridimensional com o auxílio do software STATISTICA v. 13 (Figura 5).

Figura 5 - MDS plot tridimensional a partir das distâncias genéticas de Portugal, África (AFR), Ásia (EAS), Portugal, Terena, Rio de Janeiro (RJ), São Paulo (SP) e Espírito Santo (ES), obtidas com os dados de 38 HID-INDELS. As populações ES e RJ aparecem sobrepostas.



Fonte: Este trabalho

As populações de África, Portugal, Ásia e Terena (nativo americanos) se apresentam totalmente distintas na plotagem, e as três populações da região Sudeste do Brasil, bastante próximas entre si, e mais próximas da população de Portugal do que dos outros grupos populacionais.

Este é um exemplo que demonstra a forte influência genética europeia nas populações brasileiras, gerada no período de colonização do Brasil por Portugal.

5.1.3 Avaliação da eficiência forense

No desenvolvimento da metodologia, os autores buscaram selecionar marcadores com características adequadas para uso em identificação humana,

dentre as quais, frequência alélica mínima de 0,25 em populações africanas, europeias e asiáticas, e heterozigosidade média de no mínimo 0,40 (PEREIRA, R. et al., 2009). Os 38 HID-INDELS foram analisados em 45 indivíduos de uma população nativo-americana (Terena-MS) tendo sido identificado um valor médio de diversidade menor quando comparado com o esperado para os outros três grupos populacionais (MANTA et al., 2012a).

Os dados de Espírito Santo e São Paulo, obtidos nos cálculos de diversidade gênica e equilíbrio de Hardy-Weinberg são mostrados na tabela 6, como heterozigosidade observada (H_o), esperada (H_e) e equilíbrio de Hardy-Weinberg ($P-HWE$) e comparados com dados da população do estado do Rio de Janeiro (MANTA et al., 2012a).

Os marcadores que apresentaram maior H_e para Espírito Santo e São Paulo foram, respectivamente, G2 (0,50237) e R10 (0,50251); ao contrário, Y9 (0,34141) e G5 e Y2 (0,32759) apresentaram menor H_e . Em comparação, no estado do Rio de Janeiro foi identificada maior H_e nos marcadores B7, G2 e R10, ambos com 0,499, e B10 com a menor H_o que foi 0,372 (MANTA et al., 2012a).

Destaca-se os menores e maiores valores obtidos para H_e , pois, é a H_e é representativa da população total, e não só da amostragem estudada.

Como mencionado anteriormente, os autores selecionaram marcadores com heterozigosidade mínima $\geq 0,40$ para os grupos parentais de europeus, africanos e asiáticos.

As populacionais brasileiras são formadas principalmente por europeus, africanos e nativo-americanos, sendo que estes 38 marcadores não foram selecionados visando à aplicação na última população parental mencionada.

Ainda assim, a grande maioria dos marcadores apresentou H_e maior que 0,40 para as duas populações aqui estudadas.

Tabela 6 - Heterozigosidade observada (Ho), heterozigosidade esperada (He) e teste exato para o equilíbrio de Hardy-Weinberg (P-HWE) calculados para os 38 HID-INDELS nas populações de Espírito Santo e São Paulo. Em comparação a He para os mesmos INDELS no estado do Rio de Janeiro (MANTA et al., 2012a).

Marcador	Espírito Santo			São Paulo			Rio de Janeiro (MANTA et al., 2012a)
	Ho	He	P-HWE	Ho	He	P-HWE	He
B1	0,49057	0,47662	0,83873	0,55102	0,49922	0,31788	0.494
B2	0,53774	0,50197	0,5609	0,49	0,47111	0,83117	0.475
B3	0,50943	0,48073	0,55048	0,49	0,49121	1,00000	0.472
B4	0,41509	0,41581	1,00000	0,55	0,48035	0,20742	0.461
B5	0,45714	0,47829	0,68457	0,48	0,46312	0,82874	0.481
B6	0,45283	0,47662	0,68129	0,44	0,42211	0,81139	0.461
B7	0,41509	0,49951	0,11529	0,50505	0,50208	1,00000	<u>0.499</u>
B8	0,4434	0,50197	0,24733	0,65	0,50206	0,00488	0.498
B9	0,4434	0,48265	0,4205	0,54	0,47357	0,20282	0.457
B10	0,42453	0,4197	1,00000	0,48	0,42995	0,34699	<u>0.372</u>
G1	0,40952	0,50221	0,07811	0,48	0,50231	0,6929	0.498
G2	0,49057	<u>0,50237</u>	0,8473	0,45	0,49643	0,41806	<u>0.499</u>
G3	0,37736	0,39936	0,62435	0,50505	0,45942	0,38269	0.390
G4	0,56604	0,48073	0,07014	0,5	0,45106	0,37081	0.454
G5	0,40566	0,39502	1,00000	0,39	<u>0,32759</u>	0,06559	0.419
G6	0,45283	0,49092	0,43761	0,57	0,49402	0,15297	0.495
G7	0,5	0,46477	0,52977	0,41	0,38186	0,59889	0.443
G8	0,51887	0,49481	0,69559	0,50505	0,49633	1,00000	0.492
G9	0,35849	0,39059	0,45456	0,33333	0,33005	1,00000	0.431

Y1	0,48113	0,46477	0,83322	0,59	0,48799	0,041	0.480
Y2	0,31132	0,3571	0,27064	0,35	<u>0,32759</u>	0,75701	0.397
Y3	0,34286	0,42579	0,06374	0,45918	0,43846	0,81681	0.413
Y4	0,4717	0,4235	0,2583	0,5	0,45106	0,37097	0.345
Y5	0,53774	0,49231	0,42408	0,55	0,50005	0,41977	0.487
Y6	0,43396	0,45068	0,82837	0,56	0,46312	0,04932	0.407
Y7	0,34906	0,42721	0,06926	0,43	0,39151	0,44061	0.416
Y8	0,43396	0,48449	0,3177	0,47	0,49121	0,68745	0.487
Y9	0,35849	<u>0,34141</u>	0,77655	0,4	0,36663	0,41959	0.333
R1	0,49057	0,46732	0,67935	0,41	0,44779	0,50048	0.488
R2	0,40566	0,41183	1,00000	0,49495	0,43609	0,24458	0.472
R3	0,48113	0,44116	0,38307	0,51	0,46025	0,3811	0.443
R4	0,33962	0,38147	0,30545	0,37	0,43372	0,16692	0.430
R5	0,43396	0,45068	0,82865	0,51	0,47593	0,52785	0.452
R6	0,42453	0,45368	0,52423	0,48	0,46312	0,82805	0.463
R7	0,45283	0,50165	0,3344	0,51	0,50246	1,00000	0.497
R8	0,42453	0,4594	0,52358	0,4	0,40523	1,00000	0.416
R9	0,38095	0,43308	0,25715	0,36364	0,39871	0,44424	0.416
R10	0,38679	0,48623	0,04584	0,56	<u>0,50251</u>	0,31848	<u>0.499</u>

Fonte: Este trabalho.

*Os maiores e menores valores de He em cada população estão em negrito e sublinhado.

Não foi identificado valor de P-HWE significativo após correção por Bonferroni ($p < 0,0013$).

As frequências alélicas obtidas para cada um dos 38 marcadores nas duas populações são mostradas a seguir (Tabela 7):

Tabela 7 - Frequências alélicas calculadas para 38 HID-INDELS nas populações de Espírito Santo e São Paulo; nome do marcador e respectivo código usado no eletroferograma; 1 (alelo curto) e 2 (alelo longo).

Marcador	Código	Espírito Santo		São Paulo	
		1	2	1	2
rs34541393	B1	0,386792	0,613208	0,452128	0,547872
rs16624	B2	0,514151	0,485849	0,632979	0,367021
rs2307689	B3	0,396226	0,603774	0,425532	0,574468
rs35769550	B4	0,292453	0,707547	0,388298	0,611702
rs2307700	B5	0,386792	0,603774	0,356383	0,643617
rs140809	B6	0,386792	0,613208	0,31383	0,68617
rs3047269	B7	0,537736	0,462264	0,5	0,5
rs33972805	B8	0,485849	0,514151	0,510638	0,489362
rs33917182	B9	0,599057	0,400943	0,606383	0,393617
rs16402	B10	0,297170	0,702830	0,31383	0,68617
rs1610871	G1	0,485849	0,504717	0,494681	0,505319
rs2067238	G2	0,500000	0,500000	0,542553	0,457447
rs2067294	G3	0,273585	0,726415	0,356383	0,643617
rs2307710	G4	0,396226	0,603774	0,335106	0,664894
rs2308242	G5	0,268868	0,731132	0,207447	0,792553
rs2307580	G6	0,424528	0,575472	0,43617	0,56383
rs1160956	G7	0,636792	0,363208	0,744681	0,255319
rs34511541	G8	0,438679	0,561321	0,446809	0,553191
rs2307978	G9	0,264151	0,735849	0,218085	0,781915
rs3051300	Y1	0,363208	0,636792	0,409574	0,590426

rs10629077	Y2	0,231132	0,768868	0,212766	0,787234
rs10688868	Y3	0,301887	0,688679	0,319149	0,680851
rs2067208	Y4	0,301887	0,698113	0,345745	0,654255
rs2307579	Y5	0,429245	0,570755	0,468085	0,531915
rs2308020	Y6	0,660377	0,339623	0,62766	0,37234
rs3080855	Y7	0,306604	0,693396	0,276596	0,723404
rs1610919	Y8	0,594340	0,405660	0,558511	0,441489
rs2307839	Y9	0,216981	0,783019	0,239362	0,760638
rs2308137	R1	0,367925	0,632075	0,345745	0,654255
rs36040336	R2	0,712264	0,287736	0,68617	0,31383
rs1160886	R3	0,325472	0,674528	0,351064	0,648936
rs2308026	R4	0,254717	0,745283	0,31383	0,68617
rs2307526	R5	0,339623	0,660377	0,393617	0,606383
rs34811743	R6	0,655660	0,344340	0,62766	0,37234
rs2308189	R7	0,481132	0,518868	0,484043	0,515957
rs5895447	R8	0,353774	0,646226	0,271277	0,728723
rs2308171	R9	0,311321	0,679245	0,281915	0,718085
rs35605984	R10	0,589623	0,410377	0,521277	0,478723

Fonte: Este trabalho.

Como observado na tabela 7, todos os marcadores, com exceção do rs10629077, apresentaram frequência alélica superior a 0,25, que foi o limite mínimo utilizado pelos autores da metodologia como critério de seleção.

Um dos alelos deste marcador pode ter sido observado com frequência ligeiramente menor a 0,25, devido a não consideração de nativo-americanos na seleção deste conjunto de marcadores na padronização da metodologia, sendo este um grupo parental importante na formação genética da população do Brasil.

Não foi identificado valor de p significativo ($p < 0,0000711$ após correção por Bonferroni) na avaliação do desequilíbrio de ligação (LD) entre os pares de marcadores.

Tabela 8 Poder de discriminação (PD) e poder de exclusão (PE) dos 38 marcadores nas populações de Espírito Santo e São Paulo.

Marcadores	Espírito Santo		São Paulo	
	PD	PE	PD	PE
B01	0,607867569	0,163914069	0,592	0,236
B02	0,603595586	0,222706414	0,5986	0,17889899
B03	0,608045568	0,179377935	0,6186	0,17889899
B04	0,568707725	0,117398893	0,5742	0,235118125
B05	0,626213152	0,138831534	0,5952	0,170591846
B06	0,610359559	0,149454466	0,5696	0,140172595
B07	0,657777778	0,11392	0,621977349	0,191957581
B08	0,648451406	0,142587218	0,5158	0,355216875
B09	0,62744749	0,149454466	0,5738	0,224961235
B10	0,578141687	0,123365777	0,5622	0,170591846
G01	0,661768707	0,108233703	0,6342	0,170591846
G02	0,629583482	0,179377935	0,6402	0,147369375
G03	0,561231755	0,100734706	0,579532701	0,191957581
G04	0,559807761	0,252090952	0,5738	0,1875
G05	0,551975792	0,117398893	0,4878	0,1079548
G06	0,626557494	0,163914069	0,5742	0,256415629
G07	0,58757565	0,1875	0,5378	0,1201172
G08	0,603239587	0,213487638	0,616263647	0,191957581

G09	0,546279815	0,081176743	0,495051525	0,0781893
Y01	0,585261659	0,195889048	0,5534	0,27905158
Y02	0,532039872	0,076723152	0,4922	0,086270625
Y03	0,590294785	0,07380437	0,579133694	0,154214125
Y04	0,559095764	0,171517725	0,5738	0,1875
Y05	0,589355643	0,232210606	0,5938	0,235118125
Y06	0,593983624	0,142587218	0,5504	0,245601485
Y07	0,591491634	0,085798916	0,5422	0,133236351
Y08	0,633677465	0,135953959	0,6274	0,162572239
Y09	0,500711997	0,085798916	0,5248	0,11392
R01	0,594695621	0,179377935	0,6034	0,1201172
R02	0,558205767	0,111642114	0,561371289	0,18311894
R03	0,571911712	0,179377935	0,5778	0,19640099
R04	0,535955856	0,090595402	0,5962	0,096691649
R05	0,6000356	0,142587218	0,5934	0,19640099
R06	0,604307583	0,117398893	0,5952	0,170591846
R07	0,640619438	0,156561513	0,6198	0,19640099
R08	0,605731577	0,129548798	0,5632	0,11392
R09	0,589206349	0,09237536	0,561983471	0,09328722
R10	0,646315415	0,111642114	0,5896	0,245601485
Combinados	0,9999999999999990	0,997	0,9999999999999994	0,99945
Probabilidade				
de match	$1,1 \times 10^{-15}$		$6,4 \times 10^{-15}$	

Fonte: Este trabalho.

O poder de discriminação acumulado obtido com os 38 HID-INDELS nas populações de Espírito Santo e São Paulo são apresentados na tabela 8.

O poder de discriminação obtido com este mesmo conjunto de 38 HID-INDELS foi para africanos (0,999999999999994), europeus (0,999999999999995), asiáticos (0,999999999999997) (PEREIRA, R. et al., 2009), Rio de Janeiro (0,9999999999999990) e Terena (0,999999999999997) (MANTA et al., 2012a).

Nota-se que o poder de discriminação obtido para Espírito Santo, São Paulo e Rio de Janeiro (MANTA et al., 2012a) são semelhantes, e mais elevado quando comparado com o valor obtido para as populações de europeus, africanos, asiáticos e nativo-americanos, que são as populações parentais. Sendo que para Espírito Santo, o poder de discriminação obtido foi ligeiramente maior que para São Paulo.

Isto ocorre devido à alta miscigenação destas populações brasileiras, que tem sua formação baseada em europeus, africanos e nativo-americanos.

Um menor poder de discriminação para este conjunto de marcadores foi obtido para a amostragem nativo-americana de Terena. Este dado é gerado, pois esta população é mais sujeita a deriva genética e, além disso, os autores deste sistema 38 HID-INDELS utilizaram como critério de seleção, marcadores INDELS altamente polimórficos em europeus, africanos e asiáticos.

5.2 Inserção do marcador da amelogenina na reação de 38 HID-INDELS

Na resolução de casos de investigação de paternidade o sistema de análise de 38 HID-INDELS (PEREIRA, R., 2009) pode ser utilizado em associação com STRs, para aumentar o poder de discriminação. Os kits comerciais para análise de STRs autossômicos apresentam o marcador da amelogenina, porém, na análise de DNA degradado, pode-se não obter boa visualização da amelogenina no eletroferograma por fatores como: ausência de amplificação ou ocorrência de alelo nulo devido à baixa quantidade de DNA amplificável e, sobreposição de artefatos gerados nos marcadores STR aos alelos da amelogenina, formados pela baixa qualidade do DNA ou contaminação provenientes da longa exposição da amostra ao ambiente.

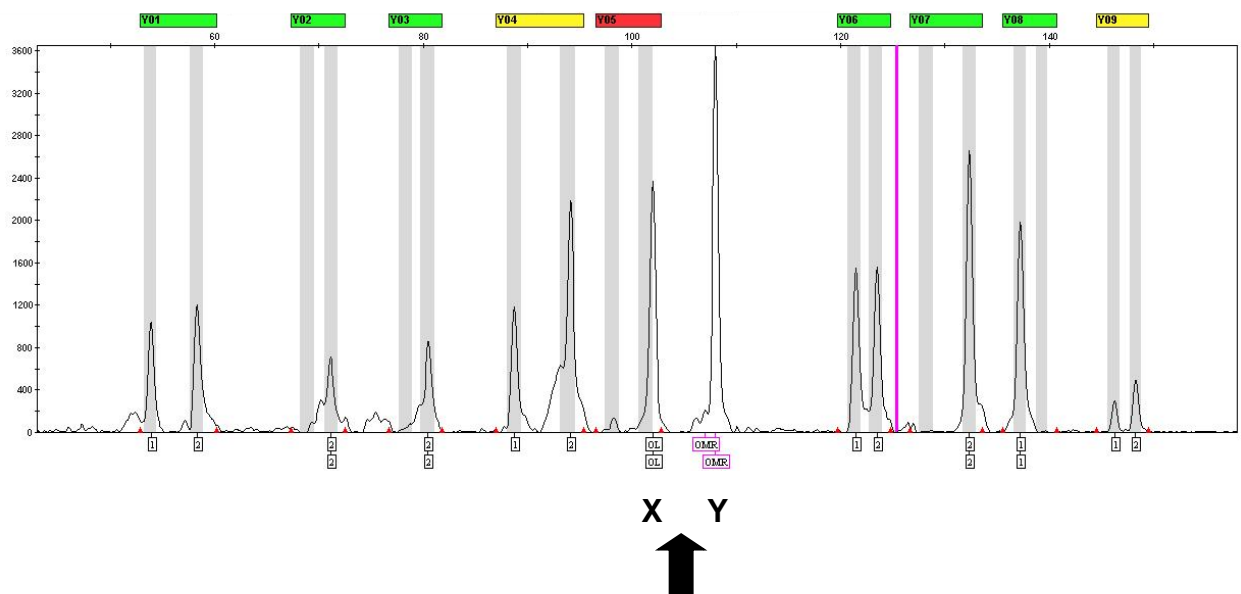
Nestes casos, é interessante uma opção complementar de amplificação da amelogenina para confirmação do sexo das amostras em questão.

Além disso, vimos nos resultados apresentados anteriormente, que o sistema 38 HID-INDELS é eficiente na identificação de indivíduos, por isto, o interesse de se inserir este marcador na reação multiplex para análise de 38 HID-INDELS, uma vez que esta reação não apresenta originalmente o marcador.

Como apresentado anteriormente, selecionou-se o par de primers descrito em Krenke et al. (2002) e a análise *in silico* com o auxílio da ferramenta *online* In Silico PCR (KENT, 2001) mostrou que o tamanho dos fragmentos a serem amplificados ocupariam perfeitamente o espaço entre 103 e 120 pb no eletroferograma.

Inicialmente, o par de primers foi testado por PCR para verificação da eficiência, e posteriormente testado em conjunto com os primers para amplificação multiplex dos 38 HID-INDELS. Os fragmentos amplificados da amelogenina apareceram no eletroferograma algumas bases a menos do que o esperado (Figura 6) e o alelo X do marcador da amelogenina apareceu muito próximo ao marcador Y05 prejudicando sua identificação.

Figura 6 - Eletroferograma mostrando os marcadores INDELS na fluorescência NED. A seta indica os picos referentes aos alelos X e Y amplificados com o par de primers original (KRENKE et al., 2002). O eixo Y indica a intensidade de fluorescência em RFU (*relative fluorescence units*) e o eixo X, a posição do fragmento em pares de base.



Fonte: Este trabalho

Para solucionar este efeito, o primer reverso para a amelogenina foi deslocado 6 pb no sentido da extremidade 3' para possibilitar a verificação correta do fragmento no eletroferograma sem prejudicar a análise de nenhum INDEL, conforme esquema a seguir (Figura 7):

Figura 7 - Ilustração da sequência do fragmento amplificado para o marcador amelogenina no cromossomo X. 1) Localização do primer reverso original; 2) Alteração do primer reverso visando a obtenção de um fragmento amplificado com 6 bases a mais.

■ Primer forward ■ fragmento a ser amplificado; ■ primer reverso original; ■ primer reverso alterado.

1)

5'CAATGCCCTGGGCTCTGTAAAGAA TAGTGTGTTGATTCTTTATCCCAGATGTTT
CTCAAGTGGTCCTGATTTTACAGTTCCTACCAC CAGCTTCCCAGTTTAAGCTCTG
ATGGTTGGCCTCAAGCCT 3'

2)

5'CAATGCCCTGGGCTCTGTAAAGAA TAGTGTGTTGATTCTTTATCCCAGATGTTT
CTCAAGTGGTCCTGATTTTACAGTTCCTACCAC CAGCTT CCCAGTTTAAGCTCTG
ATGGTTGG CCTCAAGCCT 3'

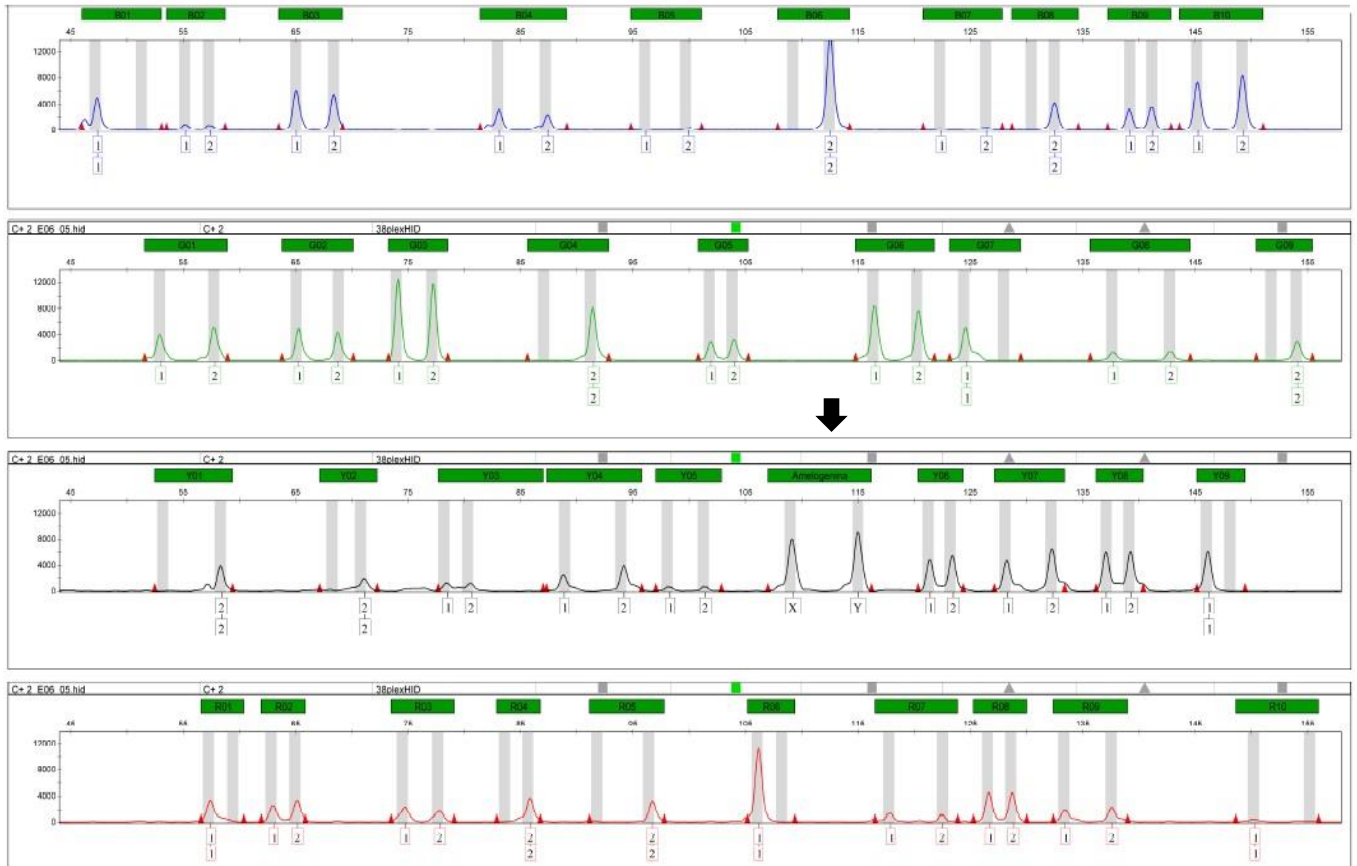


6 bases

Fonte: (BRAGANHOLI, 2015).

Assim, conforme apresentado na figura 8, foi possível obter a identificação correta da amelogenina e dos 38 HID-INDELS pela reação multiplex.

Figura 8 - Eletroferograma da reação multiplex de 38 HID-INDELs e o marcador da amelogenina (indicado pela seta). O eixo Y indica a intensidade de fluorescência em RFU (*relative fluorescence units*) e o eixo X, a posição do fragmento em pares de base.

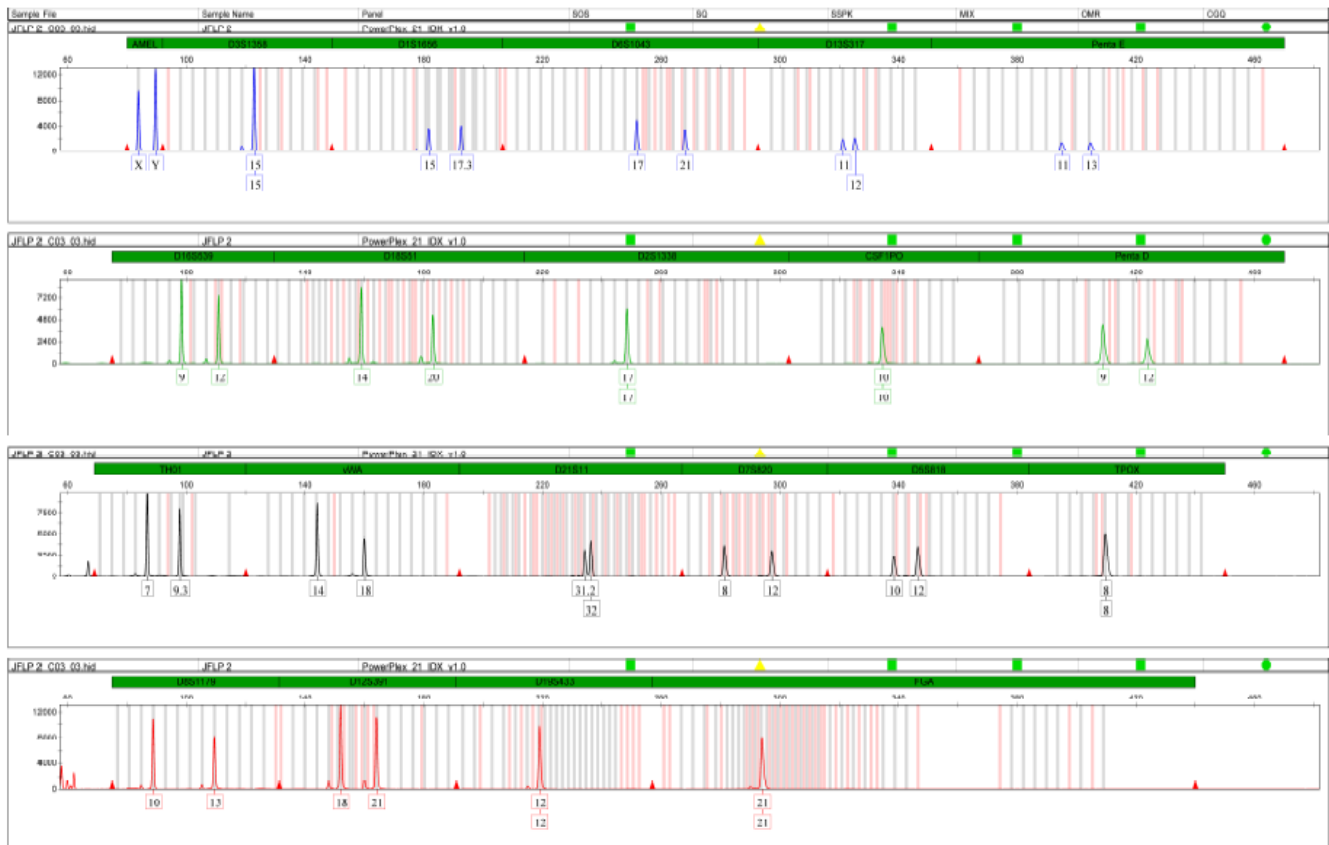


Fonte: (BRAGANHOLI, 2015).

5.3 Comparação entre STRs e 38 HID-INDELs em amostras do estado de São Paulo

Foram analisados 20 STRs autossômicos rotineiramente usados na identificação humana (Figura 9) nas amostras provenientes de indivíduos do estado de São Paulo para avaliação da eficiência forense e comparação com os dados obtidos dos 38 HID-INDELs.

Figura 9 - Eletroferograma de uma reação para análise de 20 STRs mais amelogenina com o kit PowerPlex 21 (Promega). O eixo Y indica a intensidade de fluorescência em RFU (*relative fluorescence units*) e o eixo X, a posição do fragmento em pares de base.



Fonte: Este trabalho

Os perfis individuais de cada amostra, obtidos pela análise do kit PowerPlex 21 (Promega), foram analisados individualmente para verificação da qualidade no software GeneMapper ID-X v. 1.2 (Thermo Fisher Scientific).

Foram calculados o poder de discriminação (PD) e o poder de exclusão (PE) para os 20 STRs (Tabela 9):

Tabela 9 - Poder de discriminação (PD) e poder de exclusão (PE) para os 20 STRs e comparados com os resultados acumulados dos 38 HID-INDELS.

Marcadores	PD	PE
CSF1PO	0,898375677	0,403458402
Penta D	0,947603358	0,74728545
TH01	0,85687904	0,45874277
vWA	0,9298	0,694864375
D21S11	0,896868716	0,502544038
D7S820	0,916437098	0,633209356
D5S818	0,8924	0,476181321
TPOX	0,831751862	0,423587748
D8S1179	0,9278	0,754773606
D12S391	0,965100833	0,687298172
D13S317	0,906	0,544598809
Penta E	0,9758	0,81589215
D16S539	0,8978	0,636671354
D18S51	0,961890879	0,591486616
D2S1338	0,97	0,81589215
D19S433	0,896468144	0,628666226
FGA	0,95867036	0,742042281
D3S1358	0,883	0,544598809
D1S1656	0,945153061	0,744689941
D6S1043	0,943122811	0,66584501
Combinados	>0,999999999999999	0,9999999991
38 HID-INDELS	0,999999999999994	0,99945

Fonte: Este trabalho.

Os valores do PD e PE obtidos com os 20 STRs são significativamente superiores quando comparados com os 38 HID-INDELS.

A probabilidade de *match*, que é a probabilidade de dois indivíduos compartilharem o mesmo genótipo na população, foi $(6,4 \times 10^{-15})$ para o conjunto 38 HID-INDELS e $(8,7 \times 10^{-24})$ para o conjunto de 20 STRs autossômicos.

Os marcadores INDELS são bialélicos (dois alelos possíveis), enquanto os STRs são multialélicos (diferentes números de alelos possíveis), isto gera maior poder de discriminação para cada marcador STR, no entanto, na análise de amostras de DNA com alto nível de degradação onde perfis incompletos são obtidos com os STRs, a sua combinação com marcadores INDELS pode ser uma ótima alternativa para aumentar o poder de discriminação.

Para que STRs e INDELS possam ser combinados, é importante conhecer o desequilíbrio de ligação entre os marcadores na população em que estes sejam utilizados, pois apesar da distância no genoma, o desequilíbrio pode ser específico para cada população (SZIBOR et al., 2005).

Fondevila et al. (2012) citaram dois pares de marcadores STR-INDEL muito próximos no genoma para serem considerados como marcadores independentes: D7S820-rs2307978 e D12S391-rs1610919. Estes dois INDELS são analisados no método de 38 HID-INDELS e os dois STRs são analisados pelo kit PowerPlex 21 (Promega).

O teste exato para avaliação do desequilíbrio de ligação foi realizado para todos os marcadores (20 STRs e 38 INDELS) e não foi identificado valor de p significativo após a correção de Bonferroni ($p < 0,000029$) para nenhum par de marcadores. Foram obtidos os valores $p = 0,43988$ para o par D7S820-rs2307978 e $p = 0,99606$ para o par D12S391-rs1610919, indicando que não há desequilíbrio de ligação entre estes marcadores na população estudada do estado de São Paulo e que podem ser utilizados para identificação humana como marcadores independentes.

5.4 Análise de 46 AIM-INDELS

Nesta parte, são discutidos os dados obtidos da análise de 46 AIM-INDELS nas populações dos estados de São Paulo e Espírito Santo.

5.4.1 Implantação e avaliação da técnica

Apesar das amostras terem sido analisadas durante o estágio sanduíche no exterior utilizando o equipamento ABI 3130 (Thermo Fisher Scientific), foi realizado o ajuste do protocolo descrito por Pereira, R. et al. (2012b) para o ABI 3500 pelos mesmos motivos descritos no item 5.1.1.

As diferentes posições de leitura em pb são mostradas a seguir (Tabela 10).

Tabela 10 - Comparação das posições em pares de base (pb) no eletroferograma dos alelos de cada um dos 46 marcadores com os respectivos fluorocromos no ABI 3130 e no ABI 3500 (Thermo Fisher Scientific). 1 (alelo curto), 2 (alelo longo) e 3 (terceiro estado alélico).

Marcador	Fluorocromo	ABI 3130 (polímero POP-7)			ABI 3500 (polímero POP-4)		
		Pereira, R. et al., (2012)			Este trabalho		
		1 (pb)	2 (pb)	3* (pb)	1 (pb)	2 (pb)	3* (pb)
1470	Azul	62,55	67,49		55,59	61,37	
777	Azul	71,3	74,1		66,75	69,93	
196	Azul	81,62	84,74		76,34	79,45	
881	Azul	89,78	93,68		84,58	88,77	
3122	Azul	99,03	102,63		94,68	99,02	
548	Azul	108,34	110,34		104,86	106,83	
659	Azul	117,55	119,34		113,39	115,4	
2011	Azul	128,06	132,92		124,19	129,02	
2929	Azul	151,56	153,66		145,82	147,88	
593	Azul	157,16	159,41		151,98	154,36	
798	Azul	169,27	175,04		164,04	170,15	
1193	Azul	181,54	183,75		176,78	179,24	
1871	Azul	191,06	193,89		187,89	190,08	

17	Azul	200,99	204,8		196,25	200,26	
2538	Azul	210,76	214,57		207,1	210,94	
1644	Azul	223,46	225,22		219,09	220,85	
3854	Verde	56,94	61,06		49,69	54,3	
2275	Verde	71,13	77,84		66,59	74,06	
94	Verde	91,29	94,35		87,59	90,69	
3072	Verde	106,51	113,13		102,54	109,53	
772	Verde	118,44	121,38		115,65	118,74	
2313	Verde	128,26	137,51		124,66	133,83	
397	Verde	164,52	168,66		159,69	164,6	
1636	Verde	174,68	176,51		170,61	172,42	
51	Verde	185,09	190,11		179,94	185,09	
2431	Verde	210,24	214,35		206,09	210,06	
2264	Verde	225,96	229,9	228,9	221,61	225,79	224,79
2256	Amarelo	59,64	62,53		52,54	56,1	
128	Amarelo	68,81	71,76		63,63	66,92	
15	Amarelo	79,13	90,24		74,16	85,17	
2241	Amarelo	108,06	116,17		104,64	113,27	
419	Amarelo	121,08	127,91		117,84	124,69	
943	Amarelo	157,51	161,4		151,96	156	
159	Amarelo	169,1	174,22		164,76	170,15	
2005	Amarelo	185,43	191,11		179,89	185,96	
250	Amarelo	201,58	203,58		197,33	199,43	
1802	Amarelo	214,48	217,32		209,62	212,64	
1607	Amarelo	222,19	223,69		217,36	219,46	

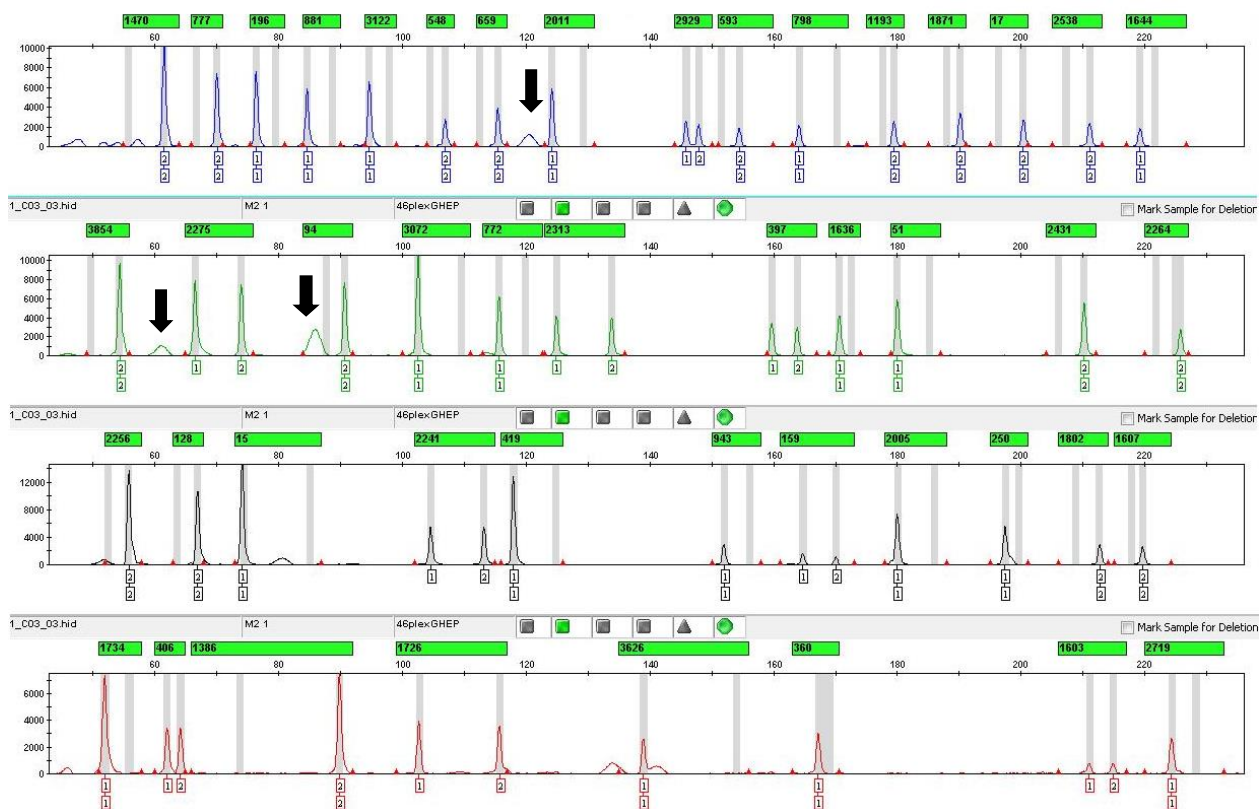
1734	Vermelho	59,53	62,83		51,89	56,5	
406	Vermelho	67,68	69,46		62,05	64,21	
1386	Vermelho	73,3	94,69		68,41	89,9	
1726	Vermelho	105,97	118,4		102,78	115,72	
3626	Vermelho	142,95	159,01		138,92	154,14	
360	Vermelho	170,76	172,55	171,63	167,26	169,05	168,12
1603	Vermelho	214,82	218,62		211	214,68	
2719	Vermelho	228,36	232,26		224,26	228,17	

Fonte: Este trabalho.

*Dois marcadores deste conjunto podem apresentar o alelo 3. Esta característica será discutida posteriormente no ítem 5.4.4.

Após o ajuste de leitura para correta identificação dos alelos, verificamos que a reação de amplificação foi eficiente para análise dos 46 INDELS (Figura 10).

Figura 10 - Eletroferograma de uma reação multiplex dos 46 AIM-NDELS. O eixo Y indica a intensidade de fluorescência em RFU (*relative fluorescence units*) e o eixo X, a posição do fragmento em pares de base. As setas indicam alguns artefatos da amplificação, que não são prejudiciais à identificação do perfil.



Fonte: Este trabalho

Destaca-se na figura alguns artefatos causados na reação de amplificação; estes artefatos podem ser gerados por excesso de álcool no processo de purificação ou fluorescência residual da purificação. Neste caso, não foi realizada a purificação do produto de PCR previamente à EC, pois, os artefatos visualizados não interferem na identificação correta do perfil da amostra.

5.4.2. Diversidade gênica e distância genética entre populações

O painel de linhagem celular para diversidade do genoma humano (HGDP-CEPH) é uma coleção de 1064 amostras de DNA de indivíduos de populações do mundo todo (CANN et al., 2002). As amostras de DNA do painel de diversidade estão publicamente disponíveis para estudos de variação genética, formando uma

base considerável para pesquisa em genética humana (CAVALLI-SFORZA, 2005). Pereira, R. et al. (2012) analisaram as amostras do HGDP-CEPH para os 46 AIM-INDELS e geraram dados para populações africanas, europeias, asiáticas e nativo-americanas.

Os dados obtidos no presente estudo, para as populações dos estados de São Paulo e Espírito Santo, foram comparados com dados de populações africanas, europeias e nativo-americanas do HGDP-CEPH (PEREIRA, R. et al., 2012). Não apresentamos aqui a comparação com os dados de populações asiáticas, pois: a história de formação populacional do Brasil indica uma miscigenação entre europeus, africanos e nativo-americanos (BUENO, 2003); trabalhos que utilizaram estes mesmos 46 AIM-INDELS em populações miscigenadas brasileiras não apresentam a comparação de dados com os de populações asiáticas (MANTA et al., 2012b), assim, seguimos o mesmo parâmetro para compararmos nossos dados com os de outras populações brasileiras; e além disso, não identificamos em nossa amostragem das populações de São Paulo e Espírito Santo nenhum indivíduo com evidente proporção de ancestralidade asiática.

Os dados também foram comparados com os das seguintes populações brasileiras, provenientes das cinco principais regiões do país: Santa Isabel e Manaus (Norte); Pernambuco e Alagoas (Nordeste); Mato Grosso do Sul e Terena (Centro-Oeste); Minas Gerais, Rio de Janeiro, São Paulo e Espírito Santo (Sudeste) e Paraná, Santa Catarina e Rio Grande do Sul (Sul) (MANTA et al., 2012b) (Tabela 11). Lembrando que as amostras de São Paulo e Espírito Santo analisadas no presente trabalho são diferentes das últimas mencionadas, e foram comparadas visando a identificação de possíveis variações provenientes de efeito amostral.

Na análise de estimativa de distância genética (F_{st}) entre as populações estudadas de São Paulo e Espírito Santo, o valor obtido para a distância genética foi $F_{st} = 0,00179$, $p = 0,09940$, indicando que as populações são bastante próximas geneticamente, como observado também nos dados dos 38 HID-INDELS.

Tabela 11 - Distâncias genéticas entre as populações (abaixo da diagonal) e valores de p correspondentes (acima da diagonal).

*amostras populacionais de Espírito Santo e São Paulo estudadas neste trabalho.

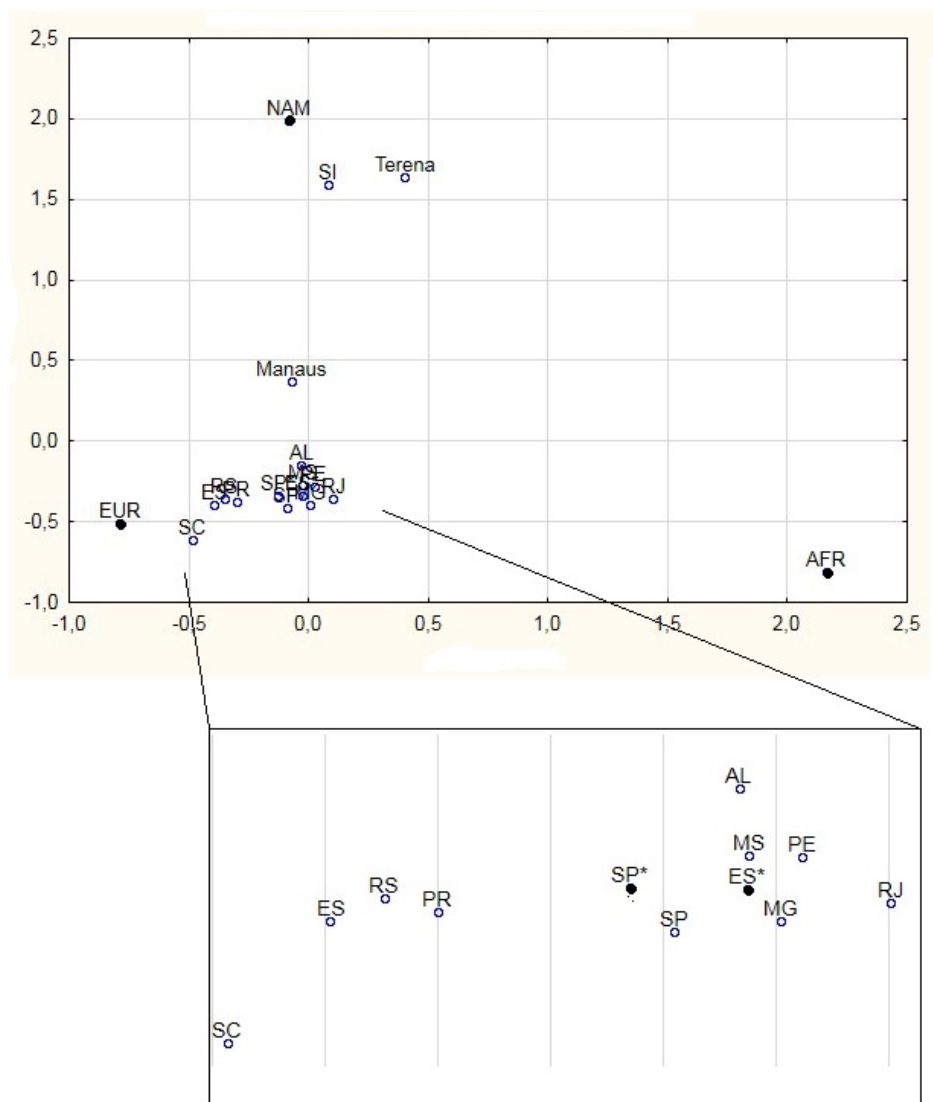
	AFR	EUR	NAM	Sta. Isabel	Manaus	PE	AL	MS	Terena	MG	ES	RJ	SP	PR	SC	RS	ES*	SP*
AFR	*	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
EUR	0,36515	*	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00149	0,02633	0,00000	0,00000	0,00000
NAM	0,44273	0,29768	*	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
Sta. Isabel	0,38605	0,22896	0,01989	*	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
Manaus	0,28378	0,08085	0,12616	0,06718	*	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
PE	0,19928	0,04762	0,21225	0,15387	0,03067	*	0,07425	0,58875	0,00000	0,47995	0,00000	0,12128	0,37392	0,03366	0,00000	0,00010	0,65706	0,04039
AL	0,21443	0,04950	0,19384	0,13834	0,02030	0,00147	*	0,43867	0,00000	0,01020	0,00000	0,00139	0,05514	0,05811	0,00000	0,00109	0,02416	0,01604
MS	0,21612	0,04507	0,21900	0,15813	0,03032	-0,00027	0,00009	*	0,00000	0,72676	0,00000	0,27077	0,14454	0,18959	0,00000	0,04188	0,76567	0,10593
Terena	0,39929	0,24800	0,02536	0,01655	0,08480	0,16179	0,14511	0,16385	*	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
MG	0,20782	0,04337	0,23940	0,17813	0,03948	0,00006	0,00335	-0,00084	0,18570	*	0,00000	0,57856	0,66112	0,04623	0,00069	0,03673	0,93011	0,04703
ES	0,29589	0,01014	0,25249	0,18361	0,04479	0,01422	0,01530	0,01281	0,19722	0,01249	*	0,00000	0,00564	0,72914	0,11930	0,67904	0,00000	0,00010
RJ	0,18190	0,05176	0,22025	0,16468	0,03604	0,00081	0,00343	0,00056	0,17494	-0,00019	0,01868	*	0,18266	0,01168	0,00000	0,00020	0,16691	0,00089
SP	0,23356	0,03689	0,25761	0,18514	0,04066	0,00050	0,00297	0,00199	0,19439	-0,00079	0,00563	0,00136	*	0,47471	0,00347	0,13682	0,44481	0,51836
PR	0,29837	0,01248	0,26107	0,17656	0,03232	0,00686	0,00573	0,00335	0,19253	0,00717	-0,00145	0,00943	0,00045	*	0,48916	0,73765	0,16484	0,35214
SC	0,34219	0,00807	0,30523	0,21891	0,06608	0,02698	0,02713	0,02537	0,24345	0,02143	0,00545	0,03081	0,01729	0,00315	*	0,28562	0,00010	0,00149
RS	0,30366	0,01774	0,25299	0,17310	0,03493	0,01320	0,01169	0,00656	0,18759	0,00727	-0,00078	0,01542	0,00508	-0,00102	0,00707	*	0,02049	0,11702
ES*	0,21479	0,04051	0,22869	0,16555	0,03263	-0,00038	0,00246	-0,00087	0,17651	-0,00163	0,00977	0,00084	0,00026	0,00382	0,02194	0,00777	*	0,09940
SP*	0,22667	0,03709	0,22349	0,16026	0,03318	0,00205	0,00299	0,00178	0,16934	0,00261	0,00830	0,00413	-0,00005	0,00176	0,01671	0,00491	0,00179	*

Nesta tabela, é possível perceber que as distâncias genéticas entre as populações parentais, e até mesmo para as populações dos estados brasileiros aqui estudados, obtidas com os 46 AIM-INDELS, são maiores do que as obtidas com os 38 HID-INDELS.

Isto é esperado, pois, as características dos marcadores analisados em cada sistema são diferentes. Enquanto os 38 HID-INDELS apresentam alta frequência alélica nas populações visando à identificação de indivíduos, os 46 AIM-INDELS apresentam diferenças alélicas entre as populações, o que permite diferenciá-las com maior eficiência.

Os dados de distância genética (F_{st}) entre as populações foram plotados em escala multidimensional (MDS-plot), conforme a figura 11 a seguir:

Figura 11 - MDS-plot da estimativa de distância genética entre europeus (EUR), nativo-americanos (NAM), africanos (AFR), Santa Isabel (SI), Terena, Manaus, Santa Catarina (SC), Rio Grande do Sul (RS), Paraná (PR), Alagoas (AL), Mato Grosso do Sul (MS), Pernambuco (PE), Minas Gerais (MG), Rio de Janeiro (RJ), São Paulo (SP) e Espírito Santo (ES). As populações EUR, AFR, NAM e as de SP e ES analisadas neste estudo estão marcadas em preto.



Fonte: Este trabalho.

As populações de Espírito Santo e São Paulo analisadas neste trabalho são semelhantes às demais populações miscigenadas brasileiras disponíveis na literatura, mais próximas geneticamente de populações da Europa do que África e Nativo-americanos, com exceção de Silvânia e a tribo Terena. Entretanto, a

população de Espírito Santo se mostrou mais próxima das demais populações da região sudeste do Brasil em comparação com os dados da amostragem do Espírito Santo disponíveis na literatura (MANTA et al., 2012b) que, segundo os autores, pode ter sofrido um efeito amostral devido à forma de seleção das amostras no momento da coleta, que foi realizada de professores e alunos de uma universidade.

Nas tabelas 12 e 13, são mostradas as frequências alélicas dos 46 AIM-INDELS e os dados obtidos nos cálculos de heterozigosidade observada (H_o), heterozigosidade esperada (H_e) e teste do equilíbrio de Hardy-Weinberg (P-HWE) nas populações de Espírito Santo e São Paulo.

Tabela 12 - Frequências alélicas dos 46 AIM-INDELS nas populações de Espírito Santo e São Paulo.

Marcador	Espírito Santo			São Paulo		
	1	2	3	1	2	3
1470	0,39899	0,60101		0,39691	0,60309	
777	0,35859	0,64141		0,38144	0,61856	
196	0,47980	0,52020		0,48454	0,51546	
881	0,73737	0,26263		0,80208	0,19792	
3122	0,93434	0,06566		0,86082	0,13918	
548	0,31818	0,68182		0,30928	0,69072	
659	0,12121	0,87879		0,13402	0,86598	
2011	0,64141	0,35859		0,53093	0,46907	
2929	0,72222	0,27778		0,65464	0,34536	
593	0,18182	0,81818		0,24742	0,75258	
798	0,44949	0,55051		0,50000	0,50000	
1193	0,16162	0,83838		0,15979	0,84021	
1871	0,26768	0,73232		0,23711	0,76289	
17	0,47980	0,52020		0,46392	0,53608	

2538	0,30808	0,69192		0,37113	0,62887	
1644	0,73232	0,26768		0,78866	0,21134	
3854	0,21717	0,78283		0,18041	0,81959	
2275	0,20202	0,79798		0,24742	0,75258	
94	0,16162	0,83838		0,21134	0,78866	
3072	0,72222	0,27778		0,77835	0,22165	
772	0,93434	0,06566		0,93299	0,06701	
2313	0,21212	0,78788		0,29381	0,70619	
397	0,76263	0,23737		0,69588	0,30412	
1636	0,63636	0,36364		0,60309	0,39691	
51	0,49495	0,50505		0,54639	0,45361	
2431	0,09091	0,90909		0,18041	0,81959	
2264	0,44444	0,51515	0,04040	0,36598	0,58247	0,05155
2256	0,22222	0,77778		0,25773	0,74227	
128	0,27273	0,72727		0,24742	0,75258	
15	0,41919	0,58081		0,48969	0,51031	
2241	0,32828	0,67172		0,37629	0,62371	
419	0,82828	0,17172		0,79381	0,20619	
943	0,64141	0,35859		0,67368	0,32632	
159	0,68182	0,31818		0,70103	0,29897	
2005	0,42424	0,57576		0,52222	0,47778	
250	0,73737	0,26263		0,72165	0,27835	
1802	0,13131	0,86869		0,12371	0,87629	
1607	0,27273	0,72727		0,29381	0,70619	
1734	0,76768	0,23232		0,75258	0,24742	

406	0,57576	0,42424		0,61856	0,38144	
1386	0,25758	0,74242		0,20103	0,79897	
1726	0,52020	0,47980		0,56701	0,43299	
3626	0,48485	0,51515		0,57732	0,42268	
360	0,74747	0,23737	0,01515	0,73196	0,25258	0,01546
1603	0,36364	0,63636		0,37629	0,62371	
2719	0,36364	0,63636		0,46392	0,53608	

Fonte: Este trabalho.

Tabela 13 - Heterozigosidade observada (Ho), heterozigosidade esperada (He) e teste exato para o equilíbrio de Hardy-Weinberg (P-HWE) calculados para os 46 AIM-INDELS nas populações de Espírito Santo e São Paulo.

Marcador	Espírito Santo			São Paulo		
	Ho	He	P-HWE	Ho	He	P-HWE
1470	0,41414	0,48203	0,20835	0,38144	0,48122	0,05425
777	0,49495	0,46234	0,51840	0,55670	0,47433	0,13151
196	0,47475	0,50172	0,68855	0,49485	0,50211	100000
881	0,38384	0,38927	100000	0,33333	0,31915	100000
3122	0,11111	<u>0,12331</u>	0,34354	0,23711	0,24085	100000
548	0,43434	0,43609	100000	0,43299	0,42946	100000
659	0,22222	0,21412	100000	0,24742	0,23332	100000
2011	0,43434	0,46234	0,66104	0,52577	0,50067	0,68486
2929	0,45455	0,40327	0,31430	0,38144	0,45452	0,12356
593	0,28283	0,29903	0,73393	0,30928	0,37434	0,10155
798	0,51515	0,49741	0,83935	0,50515	0,50259	100000

1193	0,26263	0,27237	0,71350	0,23711	0,26991	0,25386
1871	0,37374	0,39404	0,61489	0,37113	0,36366	100000
17	0,59596	0,50172	0,06998	0,57732	0,49997	0,15157
2538	0,45455	0,42850	0,63801	0,47423	0,46921	100000
1644	0,33333	0,39404	0,13017	0,27835	0,33508	0,12266
3854	0,37374	0,34174	0,55250	0,25773	0,29726	0,18599
2275	0,32323	0,32405	100000	0,35052	0,37434	0,58607
94	0,24242	0,27237	0,27225	0,31959	0,33508	0,75957
3072	0,35354	0,40327	0,22196	0,31959	0,34683	0,55416
772	0,13131	<u>0,12331</u>	100000	0,09278	<u>0,12569</u>	0,05084
2313	0,32323	0,33595	0,76296	0,38144	0,41713	0,46306
397	0,37374	0,36389	100000	0,36082	0,42546	0,15389
1636	0,36364	0,46516	0,03199	0,40206	0,48122	0,13650
51	0,42424	0,50249	0,16005	0,37113	0,49826	0,01475
2431	0,14141	0,16613	0,17231	0,31959	0,29726	0,72917
2264	0,50505	<u>0,53817</u>	0,78679	0,58763	<u>0,52684</u>	0,58751
2256	0,34343	0,34743	100000	0,41237	0,38459	0,59624
128	0,40404	0,39871	100000	0,39175	0,37434	0,78619
15	0,47475	0,48941	0,83747	0,54639	0,50238	0,42113
2241	0,45455	0,44327	100000	0,38144	0,47182	0,08186
419	0,28283	0,28590	100000	0,26804	0,32904	0,11212
943	0,45455	0,46234	100000	0,48421	0,44199	0,48243
159	0,39394	0,43609	0,35983	0,41237	0,42135	0,81257
2005	0,44444	0,49100	0,41116	0,35556	0,50180	0,00650
250	0,44444	0,38927	0,19575	0,37113	0,40382	0,45064

1802	0,18182	0,22930	0,05844	0,20619	0,21794	0,63148
1607	0,44444	0,39871	0,31145	0,44330	0,41713	0,62753
1734	0,32323	0,35851	0,39624	0,41237	0,37434	0,41490
406	0,54545	0,49100	0,30635	0,43299	0,47433	0,39707
1386	0,37374	0,38440	0,79429	0,34021	0,32290	0,75538
1726	0,53535	0,50172	0,54810	0,55670	0,49356	0,22024
3626	0,40404	0,50208	0,06972	0,43299	0,49057	0,29767
360	0,42424	0,38666	0,50574	0,38144	0,40228	0,66725
1603	0,40404	0,46516	0,20003	0,38144	0,47182	0,08271
2719	0,46465	0,46516	100000	0,43299	0,49997	0,22161

Fonte: Este trabalho.

*Os maiores e menores valores de H_o de cada população estão em negrito e grifados.

Não foram identificados valores de P-HWE significativo após correção por Bonferroni ($p < 0,0010$) ou de desequilíbrio de ligação entre os pares de marcadores ($p < 0,000048$).

5.4.3 Proporções de ancestralidade

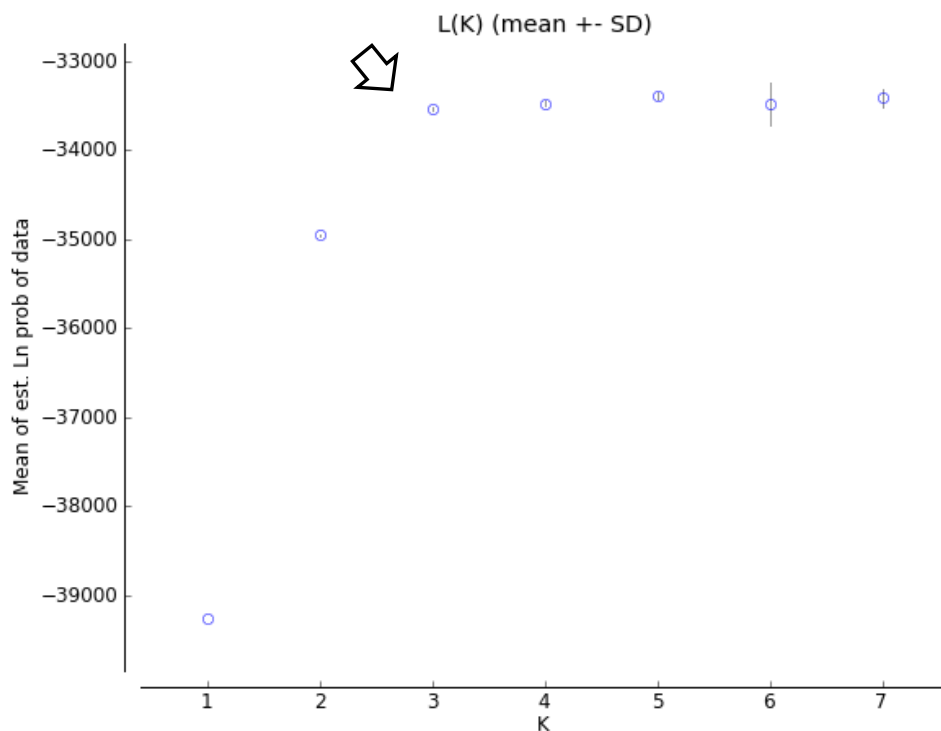
Foi realizada a comparação das populações estudadas de São Paulo e Espírito Santo com as populações africanas, europeias e nativo-americanas (PEREIRA, R. et al., 2012) para identificação de proporção de ancestralidade destes três grupos. Estes testes foram realizados com o software STRUCTURE v. 2.3.4 (FALUSH; STEPHENS; PRITCHARD, 2003) com os seguintes parâmetros: 100.000 (*burnin length*) e 100.000 (repetições Markov Chain Monte Carlo-MCMC), o mínimo de 3 corridas independentes foram realizadas para cada valor de K testado (assumindo que K é o número de grupos formados na análise).

Inicialmente, foi realizado um teste sem qualquer informação das origens das amostras usando o modo *Admixture*, assumindo teoricamente que todas eram

desconhecidas, para verificar qual seria o K ideal para análise destas populações, testando K= 1 a K= 7. Neste contexto, K pode ser definido como o número ideal de grupos que conseguem ser formados com os dados fornecidos ao software.

A estimativa de probabilidade de dados ($2\ln P(D)$) foi plotada através do site STRUCTURE Harvester v. 0.6.6 (<http://taylor0.biology.ucla.edu/structureHarvester/>) (Figura 12).

Figura 12 - Estimativa de probabilidade de dados obtida com o software STRUCTURE v. 2.3.4 e plotado usando o site STRUCTURE Harvester v. 0.6.6. A seta indica o K= 3, onde a diferenciação de grupos foi ideal.



Fonte: Este trabalho.

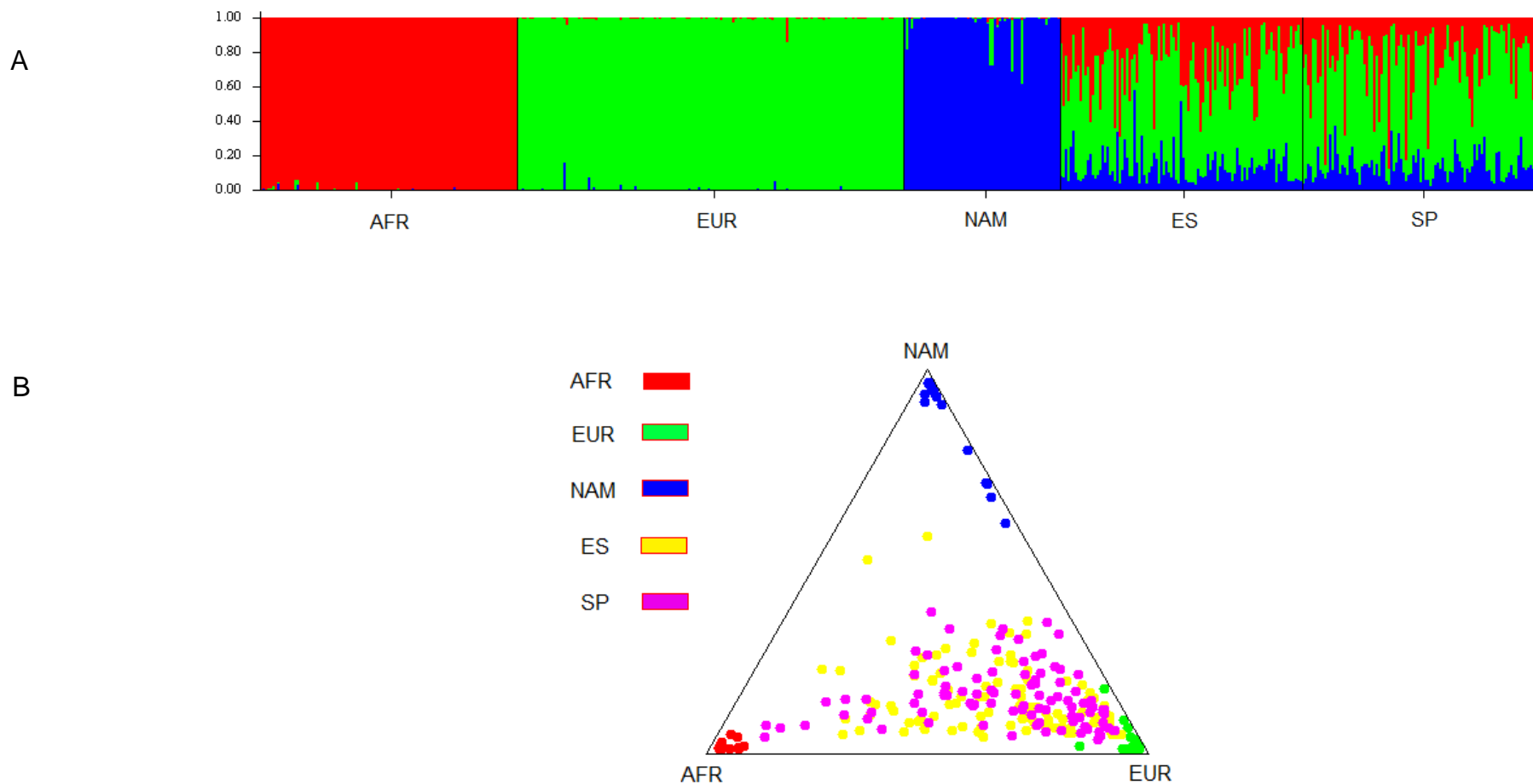
Verificamos que o K= 3 é o ideal para a análise deste grupo de populações, pois foi onde houve o maior nível de diferenciação, ou seja, três grupos foram formados com estes dados. Como esperado, os três grupos diferenciados foram os africanos, europeus e nativo-americanos, e as populações de São Paulo e Espírito Santo apresentando proporções destes três grupos principais.

Posteriormente, as amostras africanas, europeias e nativo-americanas foram utilizadas como referência para as populações de São Paulo e Espírito Santo

através da opção “*Use Population Information*” do software. Nesta análise três comparações independentes foram realizadas apenas para $K= 3$, conforme os resultados das análises iniciais.

Na figura 13 (A) são apresentados os dados obtidos desta última análise, onde é possível visualizar a proporção de ancestralidade africana, europeia e nativo-americana nas populações de São Paulo e Espírito Santo, e na figura 13 (B) como os indivíduos destas duas populações se distribuem entre as populações de referência.

Figura 13 - Resultados gerados com o software STRUCTURE. A - Proporção de origem ancestral para Espírito Santo (ES) e São Paulo (SP) a partir das amostras referência HGDP-CEPH de africanos (AFR), europeus (EUR) e nativo-americanos (NAM); B - Dispersão dos indivíduos analisados entre os três grupos populacionais de referência.



Pode ser observado que as proporções de ancestralidade das populações estudadas são muito semelhantes, apresentando uma maior contribuição europeia e menor contribuição nativo-americana, em ambas (Tabela 14).

Tabela 14 - Proporção de ancestralidade (%) para as amostras testadas de populações de africanos, europeus e nativo-americanos (PEREIRA, R. et al., 2012b) e as do Espírito Santo e de São Paulo.

	Africanos	Europeus	Nativo-americanos
Africanos	99,6	0,3	0,2
Europeus	0,2	99,4	0,4
Nativo-americanos	0,2	2,5	97,3
Espírito Santo	27	59,2	13,9
São Paulo	25	61	14

Fonte: Este trabalho.

As populações dos estados de São Paulo e Espírito Santo tem formação histórica semelhante à de várias regiões do Brasil, por isto apresentam alta miscigenação de nativo-americanos, europeus e africanos, pois, a região era habitada por povos indígenas desde aproximadamente 12.000 anos a.C. Posteriormente, os colonizadores portugueses desembarcaram na terra que para eles era desconhecida. Por volta dos anos 1800, inicia-se uma chegada maciça de imigrantes europeus, principalmente portugueses, espanhóis e italianos; além da frequente presença dos negros trazidos como escravos (BUENO, 2003).

Segundo dados do Instituto Brasileiro de Geografia e Estatística (IBGE) para raça ou cor de pele, 64% da população da região sudeste do Brasil se autodeclarou como brancos, 6,7% como pretos, 28,4% como pardos, 0,8% como amarelos e 0,1% como indígenas (<http://www.ibge.gov.br/home/estatistica/populacao/condicaodevida/indicadoresminimos/tabela1.shtm#a112> – consulta realizada em 23/02/2015).

Os dados de informação de ancestralidade obtidos neste trabalho, referentes aos estados de São Paulo e Espírito Santo, pertencentes à região sudeste do Brasil, são muito próximos dos dados publicados pelo IBGE, se considerarmos a porcentagem de indivíduos que se autodeclararam pretos e pardos como sendo de origem ancestral africana.

É importante considerar que pode haver divergência entre a autodeclaração de um indivíduo sobre sua cor de pele e/ou ancestralidade, em comparação com a proporção de ancestralidade obtida pela análise de DNA, mesmo assim, ambos os dados confirmam a alta miscigenação da população brasileira.

5.4.4 Identificação do alelo 3 nas amostras analisadas

Dois marcadores deste conjunto podem apresentar um terceiro estado alélico (alelo 3), que foi confirmado por sequenciamento pelos autores da metodologia. MID-360 apresenta um terceiro alelo devido à inserção de um “T” associado com o alelo curto, 8 bases antes da região do polimorfismo. MID-2264 apresenta um terceiro alelo que corresponde a deleção de um “T” no alelo longo. Interessante que o alelo variante do MID-360 foi encontrado somente em amostras africanas e o alelo variante de MID-2264 encontrado somente em europeus, contribuindo para a diferenciação destes dois grupos (PEREIRA, R. et al., 2012).

Para as duas populações analisadas neste trabalho, o alelo 3 foi identificado em 5 amostras no marcador MID-360 e 8 amostras no marcador MID-2264, sendo que praticamente todas estas amostras foram tipadas com alguma proporção ancestral africana e europeia, respectivamente, e nenhuma amostra foi identificada com o alelo 3 nos dois marcadores (Tabela 15).

A identificação do alelo 3 nestes dois marcadores, nas amostras utilizadas no presente estudo, certamente contribuiu para a formação dos dados de proporção de ancestralidade.

Tabela 15 - Amostras em que o alelo 3 foi identificado para os marcadores MID-360 ou MID-2264 e respectiva proporção ancestral (%) africana (AFR), europeia (EUR) e nativo-americana (NAM).

Amostra	AFR	EUR	NAM	
ES190	21,9	69,4	8,7	
ES202	35,1	62,7	2,2	
ES221	35,7	50,6	13,8	MID-360
SP032	61,5	32,1	6,4	
SP089	47,3	47,1	5,6	
ES013	32,7	59,8	7,5	
ES106	15,6	66,8	17,6	
ES179	2,5	90,6	6,9	
ES232	2,1	87,3	10,6	
SP010	87,8	7,3	4,9	MID-2264
SP017	16	70,9	13,1	
SP092	14,1	74,1	11,8	
SP094	14,2	74,7	11,1	

Fonte: Este trabalho.

5.5 Comparação de dados do cromossomo Y, DNA mt e AIM-INDELS na população do estado de Espírito Santo

A análise do cromossomo Y e do DNA mt são ferramentas muito importantes na identificação humana e no estudo de genética de populações uma vez que, por não se recombinarem, permitem a identificação das linhagens paterna e materna, respectivamente. Ao contrário, o DNA autossômico se recombina e quando utilizado

para análise de origem ancestral de uma amostra ou população, pode proporcionar uma estimativa geral de ancestralidade inclusive com proporção de miscigenação.

Estudos realizados na população do estado de Espírito Santo indicam uma superioridade de ancestralidade europeia (85,88%) na análise do cromossomo Y (FIGUEIREDO et al., 2015), e ao contrário, maior ancestralidade africana (43%) na análise do DNA mt (SANCHES et al., 2014). A média de ancestralidade dos marcadores uniparentais em uma população tende a gerar um resultado semelhante aos AIM-INDELS (Tabela 16).

Tabela 16 - Comparação entre as proporções de origem ancestral africana (AFR), europeia (EUR) e nativa-americana (NAM) obtidas da análise do cromossomo Y, DNA mt e AIM-INDELS para a população do Espírito Santo em diferentes trabalhos. N: número de amostras analisadas.

	N	AFR	EUR	NAM
Cromossomo Y (FIGUEIREDO et al., 2015)	253	11,37%	85,88%	2,75%
DNA mt (SANCHES et al., 2014)	97	43%	33%	24%
Média uniparentais	-	27,19%	59,44%	13,38%
AIM-INDELS (Este trabalho)	106	27%	59,2%	13,9%
AIM-INDELS (MANTA et al., 2012b)	92	13,4%	74,1%	12,5%

Fonte: Este trabalho.

Também é possível observar que a amostra populacional do Espírito Santo (MANTA et al., 2012b) apresenta maior índice de ancestralidade europeia em comparação com a amostragem do mesmo estado utilizada no presente trabalho, possivelmente pela escolha diferente de amostras em cada caso.

CONCLUSÕES

6 CONCLUSÕES

- Os sistemas 38 HID-INDELS e 46 AIM-INDELS tiveram as posições de leitura dos alelos corrigidas, levando em consideração a diferença de mobilidade dos fragmentos entre o analisador genético ABI 3130, que foi utilizado pelos autores no desenvolvimento da metodologia, e o ABI 3500, que foi utilizado neste trabalho;
- O método para a análise de 38 HID-INDELS foi eficiente, e a avaliação da eficiência forense mostrou que o sistema foi altamente discriminativo para as populações estudadas;
- Os cálculos estatísticos indicaram que os marcadores INDELS e STRs analisados não estão em desequilíbrio de ligação e podem ser utilizados combinadamente como marcadores independentes na população de São Paulo;
- A inserção do marcador amelogenina no sistema multiplex 38 HID-INDELS possibilitou a identificação do sexo das amostras analisadas, sendo uma alternativa interessante na análise de amostras com DNA degradado;
- O método para a análise de 46 AIM-INDELS foi eficiente, possibilitando a avaliação de ancestralidade dos indivíduos das duas populações estudadas;
- As populações estudadas dos estados de São Paulo e Espírito Santo apresentam grande diversidade para os polimorfismos analisados e confirmam a miscigenação brasileira.

REFERÊNCIAS

BASTOS-RODRIGUES, L.; PIMENTA, J. R.; PENA, S. D. The genetic structure of human populations studied through short insertion-deletion polymorphisms. **Annals of Human Genetics**, v. 70, p. 658-665, 2006.

BRAGANHOLI, D. F.; POLVERARI, F. S.; AMBROSIO, I. B.; CICARELLI, R. M. B. Insertion of amelogenin marker in the 38 Hid-Indels multiplex reaction. **Forensic Science International: Genetics Supplement Series**, 2015. In press.
doi:<http://dx.doi.org/10.1016/j.fsigss.2015.09.066>.

BUDOWLE, B.; van DAAL, A. Forensically relevant SNP classes. **Biotechniques**, v. 44, p. 603-608, 2008.

BUDOWLE, B.; MORETTI, T. R.; NIEZGODA, S. J.; BROWN, B. L. CODIS and PCRbased short tandem repeat loci: law enforcement tools. In: EUROPEAN SYMPOSIUM ON HUMAN IDENTIFICATION, 2nd, 1998, Madison. **Proceedings...** Madison: Promega Corporation, 1998, p. 73-88.

BUENO, E. **Brasil: uma história**. 2. ed. São Paulo: Ática, 2003. 447 p.

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

BUTLER, J. M. Short Tandem Repeat (STR) loci and kits. In_____. **Advanced topics in forensic DNA typing: methodology**. London: Elsevier Academic Press, c2012. Chap. 5, p. 99-139.

CANN, H. M.; de TOMA, C.; CAZES, L.; LEGRAND, M. F.; MOREL, V.; PLOUFFRE, L.; BODMER, J.; BODMER, W. F.; BONNE-TAMIR, B.; CAMBON-THOMSEN, A.; CHEN, Z.; CHU, J.; CARCASSI, C.; CONTU, L.; DU, R.; EXCOFFIER, L.; FERRARA, G. B.; FRIEDLAENDER, J. S.; GROOT, H.; GURWITZ, D.; JENKINS, T.; HERRERA, R. J.; HUANG, X.; KIDD, J.; KIDD, K. K.; LANGANEY, A.; LIN, A. A.; MEHDI, S. Q.; PARHAM, P.; PIAZZA, A.; PISTILLO, M. P.; QIAN, Y.; SHU, Q.; XU, J.; ZHU, S.; WEBER, J. L.; GREELY, H. T.; FELDMAN, M. W.; THOMAS, G.; DAUSSET, J.; CAVALLI-SFORZA, L. L. A human genome diversity cell line panel. **Science**, v. 296, p. 261-262, 2002.

CAVALLI-SFORZA, L. L. The human genome diversity project: past, present and future. **Nature Reviews Genetics**, v. 6, p. 333-340, 2005.

CLIFFORD, R.; EDMONSON, M.; HU, Y.; NGUYEN, C.; SCHERPBIER, T.; BUETOW, K. H. Expression-based genetic physical maps of single nucleotide polymorphisms identified by the cancer genome anatomy project. **Genome Research**, v. 10, p. 1259-1265, 2000.

DAWSON, E.; CHEN, Y.; HUNT, S.; SMINK, L. J.; HUNT, A.; RICE, K.; LIVINGSTON, S.; BUMPSTEAD, S.; BRUSKIEWICH, R.; SHAM, P.; GANSKE, R.; ADAMS, M.; KAWASAKI, K.; SHIMIZU, N.; MINOSHIMA, S.; ROE, B.; BENTLEY, D.; DUNHAM, I. A SNP resource for human chromosome 22: extracting dense clusters of SNPs from the genomic sequence. **Genome Research**, v. 11, p. 170-178, 2001.

EXCOFFIER, L.; LAVAL, G.; SCHNEIDER, S. Arlequin version 3.0: an integrated software package for population genetics data analysis. **Evolutionary Bioinformatics Online**, v. 1, p. 47-50, 2005.

FALUSH, D.; STEPHENS, M.; PRITCHARD, J. K. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. **Genetics**, v. 164, p. 1567-1587, 2003.

FIGUEIREDO, R. F.; AMBROSIO, I. B.; BRAGANHOLI, D. F.; CHEMALE, G.; MARTINS, J. A.; GOMES, V.; GUSMÃO, L.; CICARELLI, R. M. B. Male-specific contributions to the Brazilian population of Espirito Santo. **International Journal of Legal Medicine**, 2015. doi:10.1007/s00414-015-1214-2.

FONDEVILA, M.; PHILLIPS, C.; SANTOS, C.; PEREIRA, R.; GUSMÃO, L.; CARRACEDO, A.; BUTLER, J. M.; LAREU, M. V.; VALLONE, P. M. Forensic performance of two insertion–deletion marker assays. **International Journal of Legal Medicine**, v. 126, p. 725-737, 2012.

FRANÇA, G. V. Antropologia médico-legal. In: _____. **Medicina legal**. 9. ed. Rio de Janeiro: Guanabara-Koogan, 2011. Cap. 3, p. 48-82.

HIRSCHHORN, J. N.; GAJDOS, Z. K. Genome-wide association studies: results from the first few years and potential implications for clinical medicine. **Annual Review of Medicine**, v. 62, p. 11-24, 2011.

JAKOBSSON, M.; SCHOLZ, S. W.; SCHEET, P.; GIBBS, J. R.; VANLIERE, J. M.; FUNG, H. C.; SZPIECH, Z. A.; DEGNAN, J. H.; WANG, K.; GUERREIRO, R. Genotype, haplotype and copy-number variation in worldwide human populations. **Nature**, v. 451, p. 998-1003, 2008.

JIANG, Y.; TURINSKY, A. L.; BRUDNO, M. The missing indels: an estimate of indel variation in a human genome and analysis of factors that impede detection. **Nucleic Acids Research**, v. 43, p. 7217-7228, 2015.

JOBLING, M. A.; GILL, P. Encoded evidence: DNA in forensic analysis. **Nature Reviews Genetics**, v. 5, p. 739-751, 2004.

KAYSER, M.; de KNIJFF, P. Improving human forensics through advances in genetics genomics and molecular biology. **Nature Reviews Genetics**, v. 12, p. 179-192, 2011.

KENT, J. In-silico PCR. In: UCSC GENOME bioinformatics. Santa Cruz: University of California, [2001]. Disponível em: <<https://genome.ucsc.edu/cgi-bin/hgPcr>>. Acesso em: 16 dez. 2015.

KRENKE, B. E.; TEREBA, A.; ANDERSON, S. J.; BUEL, E.; CULHANE, S.; FINIS, C. J.; TOMSEY, C. S.; ZACHETTI, J. M.; MASIBAY, A.; RABBACH, D. R.; AMIOTT, E. A.; SPRECHER, C. J. Validation of a 16-locus fluorescent multiplex system. **Journal of Forensic Sciences**, v. 47, p. 773-785, 2002.

LANDER, E. S. Initial impact of the sequencing of the human genome. **Nature**, v. 470, p. 187-197, 2011.

LITT, M.; LUTY, J. A. A hipervariável microssatélite revelado por in vitro amplificação de um dinucleotídeo repetido dentro do gene da actina muscular cardíaca. **American Journal of Human Genetic**, v. 44, p. 397-401, 1989.

MANTA, F.; CAIAFA, A.; PEREIRA, R.; SILVA, D.; AMORIM, A.; CARVALHO, E. F.; GUSMÃO, L. Indel markers: genetic diversity of 38 polymorphisms in Brazilian populations and application in a paternity investigation with post mortem material. **Forensic Science International: Genetics**, v. 6, p. 658-661, 2012a.

MANTA, F. S. N.; PEREIRA, R.; VIANNA, R.; ARAÚJO, A. R. B.; GITAÍ, D. L. G.; DA SILVA, D. A. da; WOLFGRAMM, E. V.; PONTES, I. M.; AGUIAR, J. I.; MORAES, M. O.; CARVALHO, E. F. de; GUSMÃO, L. Revisiting the genetic ancestry of Brazilians using autosomal AIM-Indels. **PLoS One**, v. 8, 2012b.
doi: 10.1371/journal.pone.0075145

MARTINS, J. A. **Estudo de polimorfismos do cromossomo X na população da região sudeste do Brasil**. 2011. 109 f. Tese (Doutorado em Biociências e Biotecnologia Aplicadas à Farmácia) - Faculdade de Ciências Farmacêuticas, Universidade Estadual Paulista, Araraquara, 2011.

MILLS, R. E.; LUTTIG, C. T.; LARKINS, C. E.; BEAUCHAMP, A.; TSUI, C.; PITTARD, W. S. An initial map of insertion and deletion (INDEL) variation in the human genome. **Genome Research**, v. 16, p. 1182-1190, 2006.

MONTEIRO, E. H. G. **Desenvolvimento e aplicação de polimorfismos de inserção/deleção do cromossomo X em genética forense**. 2007. 162 f. Dissertação (Mestrado em Ciências Genômicas e Biotecnologia) - Universidade Católica de Brasília, Brasília, DF, 2007.

MULLANEY, J. M.; MILLS, R. E.; PITTARD, W. S.; DEVINE, S. E. Small insertions and deletions (INDELS) in human genomes. **Human Molecular Genetics**, v. 19, p. 131-136, 2010.

NACHMAN, M. W.; CROWELL, S. L. Estimate of the mutation rate per nucleotide in humans. **Genetics**, v. 156, p. 297-304, 2000.

PEREIRA, F.; CARNEIRO, J.; MATTHIESEN, R.; van ASCH, B.; PINTO, N.; GUSMAO, L.; AMORIM, A. Identification of species by multiplex analysis of variable-length sequences. **Nucleic Acids Research**, v. 38, 2010. doi: 10.1093/nar/gkq865.

PEREIRA, R. **Bridging the gap between SNPs and STRs**: insertion deletion polymorphisms in forensic genetics; principles and applications. 2011. 259 f. Tese (Doutorado em Ciências Forenses) - Faculdade de Medicina e Odontologia, Universidade de Santiago de Compostela, Santiago de Compostela, 2011.

PEREIRA, R.; GUSMÃO, L. Capillary electrophoresis of 38 non-coding biallelic mini-Indels for degraded samples and as complementary tool in paternity testing. In: ALONSO, A. (Ed.). **DNA electrophoresis protocols for forensic genetics**. New York: Humana Press, 2012. Chap. 10, p. 141-158.

PEREIRA, R.; PHILLIPS, C.; ALVES, C.; AMORIM, A.; CARRACEDO, A.; GUSMÃO, L. A new multiplex for human identification using insertion/deletion polymorphisms. **Electrophoresis**, v. 30, p. 3682-3690, 2009.

PEREIRA, R.; PHILLIPS, C.; PINTO, N.; SANTOS, C.; SANTOS, S. E. B.; AMORIM, A.; CARRACEDO, A.; GUSMÃO, L. Straightforward inference of ancestry and admixture proportions through ancestry-informative insertion deletion multiplexing. **PLoS One**, v. 7, 2012. doi:10.1371/journal.pone.0029684.

PHILLIPS, C.; FONDEVILA, M.; GARCIA-MAGARINOS, M.; RODRIGUEZ, A.; SALAS, A.; CARRACEDO, A.; LAREU, M. V. Resolving relationship tests that show ambiguous STR results using autosomal SNPs as supplementary markers. **Forensic Science International: Genetics**, v. 2, p. 198-204, 2008.

PIMENTA, J. R.; PENA, S. D. Efficient human paternity testing with a panel of 40 short insertion-deletion polymorphisms. **Genetics and Molecular Research**, v. 9, p. 601-607, 2010.

RIBEIRO-RODRIGUES, E. M.; SANTOS, N. P. dos; SANTOS, A. K. R. dos; PEREIRA, R.; AMORIM, A.; GUSMÃO, L.; ZAGO, M. A.; SANTOS, S. E. B. Assessing interethnic admixture using an X-linked insertion-deletion multiplex. **American Journal of Human Biology**, v. 21, p. 707-709, 2009.

ROSENBERG, N. A.; LI, L. M.; WARD, R.; PRITCHARD, J. K. Informativeness of genetic markers for inference of ancestry. **American Journal of Human Genetics**, v. 73, p. 1402-1422, 2003.

SANCHES, N. M.; PANETO, G. G.; FIGUEIREDO, R. F.; MELLO, A. O.; CICARELLI, R. M. B. Mitochondrial DNA control region diversity in a population from Espírito Santo State, Brazil. **Molecular Biology Reports**, v. 41, p. 6645-6648, 2014.

SANTOS, N. P.; RIBEIRO-RODRIGUES, E. M.; SANTOS, A. K. C. R. dos; PEREIRA, R.; GUSMÃO, L.; AMORIM, A.; GUERREIRO, J. F.; ZAGO, M. A.; MATTE, C.; HUTZ, M. H.; SANTOS, S. E. B. Assessing individual interethnic admixture and population substructure using a 48-insertion-deletion (INDEL) ancestry-informative marker (AIM) panel. **Human Mutation**, v. 31, p. 184-190, 2010.

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

SOBRINO, B.; BRION, M.; CARRACEDO, A. SNPs in forensic genetics: a review on SNP typing methodologies. **Forensic Science International**, v. 154, p. 181-194, 2005.

SZIBOR, R. et al. Haplotyping os STR cluster DXS6801-DXS6809-DXS6789 on Xq21 provides a powerful tool for kinship testing. **International Journal of Legal Medicine**, v. 119, p. 363-369, 2005.

TISHKOFF, S. A.; REED, F. A.; FRIEDLAENDER, F. R.; EHRET, C.; RANCIARO, A.; FROMENT, A.; HIRBO, J. B.; AWOMOYI, A. A.; BODO, J. M.; DOUMBO. O. The genetic structure and history of Africans and African Americans. **Science**, v. 324, p. 1035-1044, 2009.

WEBER, J. L.; DAVID, D.; HEIL, J.; FAN, Y.; ZHAO, C.; MARTH, G. Human diallelic insertion/deletion polymorphisms. **American Journal of Human Genetics**, v. 71, p. 854-862, 2002.

WEEDEN, V. W.; SWARNEN, S. L. Exames forenses de identificação por análises do DNA. In: HENRY, J. B. **Diagnósticos clínicos e tratamento por métodos laboratoriais**. 19. ed. São Paulo: Manole, 1998. Cap. 63, p. 1427-1438.

YANG, N.; LI, H.; CRISWELL, L. A.; GREGERSEN, P. K.; ALARCON-RIQUELME, M. E.; KITTLES, R.; SHIGETA, R.; SILVA, G.; PATEL, P. I.; BELMONT, J. W.; SELDIN, M. F. Examination of ancestry and ethnic affiliation using highly informative diallelic DNA markers: application to diverse and admixed populations and implications for clinical epidemiology and forensic medicine. **Human Genetics**, v. 118, p. 382-392, 2005.

ZIDKOVA, A.; HORINEK, A.; KEBRDLOVA, V.; KORABECNA, M. Application of the new insertion–deletion polymorphism kit for forensic identification and parentage testing on the Czech population. **International Journal of Legal Medicine**, v. 127, p. 7-10, 2011.

ANEXO



Contents lists available at ScienceDirect

Forensic Science International: Genetics Supplement Series

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



Insertion of amelogenin marker in the 38 Hid-Indels multiplex reaction

D.F. Braganholi*, F.S. Polverari, I.B. Ambrosio, R.M.B. Cicarelli

* Laboratory of Paternity Investigation, NAC/PCF–UNESP, Araraquara, SP, Brazil

ARTICLE INFO

Article history:
Received 4 August 2015
Accepted 16 September 2015
Available online xxx

Keywords:
Amelogenin marker
INDELS
Degraded DNA

ABSTRACT

Pereira et al. [3] reported a method for multiplex analysis of 38 INDELS autosomes known polymorphic in different populations of the world which can be used in the human identification, but this methodology not originally introduced marker for the analysis of amelogenin. In laboratory of paternity of UNESP/ Araraquara, Brazil, kinship analysis are performed using samples with degraded DNA, where the amplification of 38HID-INDELS was efficient compared to STRs in these samples. For paternity tests also analyze STRs, and the verification of the sex of samples is done by the amelogenin marker analysis present in commercial kits, but the bad amplification of STRs may interfere in the correct reading of amelogenin in the electropherogram. Trying to solve this problem we decided to add the amelogenin marker in multiplex reaction with 38HID-INDELS.

© 2015 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Ancient DNA may be to present difficulties or fails in STR amplification related to preservation and contamination, restricting the correct typing of the sample [1]. INDEL markers can overcome this problem because presents small fragment to be amplified in the PCR (about 200 bp) [2]. Therefore, we use the analysis method for 38 autosomal INDELS [3] and insert the amelogenin marker in the multiplex reaction, to complement the STR results when necessary.

2. Objectives

Insert the amelogenin marker in the multiplex reaction with 38HID-INDELS for use in the analysis of the degraded samples.

3. Methodology

Through literature search and *in silico* analysis we select the pair of primers described by Krenke et al. [4], and we made a small alteration in the sequence of the reverse primer, for the amplification of a fragment with 6 pb additional for correct

reading of the amelogenin marker and 38HID-INDELS in the electropherogram. The sequences of the primers are (5'–3'): "NED-CCCTGGGCTCTGTAAAGAA" (forward) and "CCAACCATCAGAGCT-TAAACTGGG" (reverse) and were used in the final concentration of 0.2 μM in the PCR, with reagents and cycling by the protocol [3].

4. Results and discussion

Until this moment, the multiplex reaction 38HID-INDELS with amelogenin marker (Fig. 1) was used in the analysis of fifteen cases of biological kinship with degraded DNA samples helping to solve them.

5. Conclusion

The alteration of the reverse primer allowed the multiplex analysis of the amelogenin marker with 38HID-INDELS without interaction between the primers.

Conflict of interest

None.

* Corresponding author. Fax: +55 16 33016940.
E-mail address: danielobraganholi@hotmail.com (D.F. Braganholi).

<http://dx.doi.org/10.1016/j.fsigs.2015.09.066>
1875-1768/© 2015 Elsevier Ireland Ltd. All rights reserved.

Please cite this article in press as: D.F. Braganholi, et al., Insertion of amelogenin marker in the 38 Hid-Indels multiplex reaction, Forensic Sci. Int. Gene. Suppl. (2015), <http://dx.doi.org/10.1016/j.fsigs.2015.09.066>

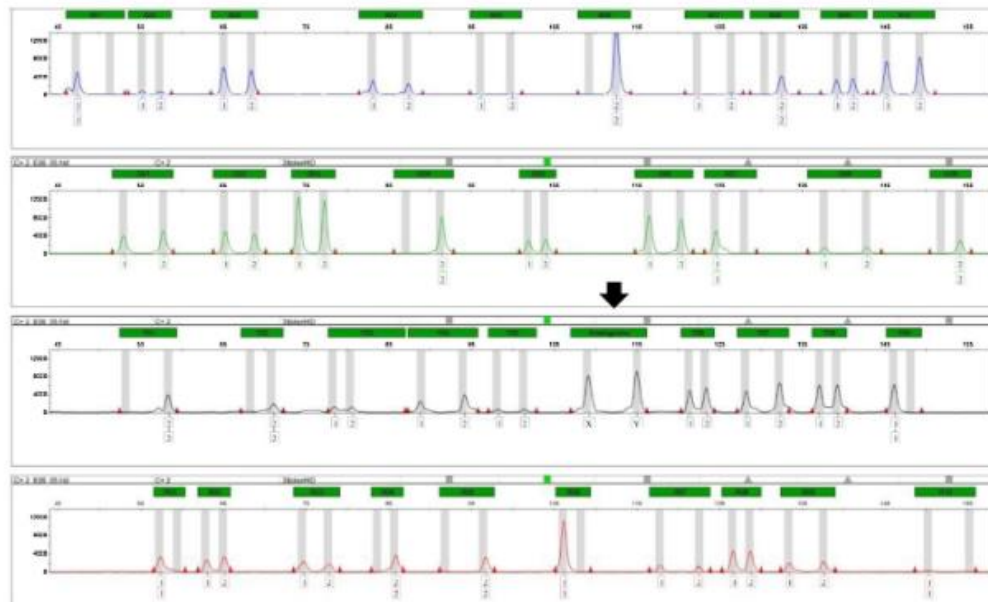


Fig. 1. Electropherogram of 38 HID-INDELS with amelogenin marker (indicated by arrow).

Acknowledgments

The authors would like to thank Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) by financial support and Rui Pereira by scientific support.

References

[1] C. Romanini, M. Romero, M.S. Puerto, L. Catelli, C. Phillips, R. Pereira, L. Gusmão, C. Vullo, Ancestry informative markers: inference of ancestry in aged bone

samples using an autosomal AIM-Indel multiplex, *Forensic Sci. Int. Genet.* 16 (2015) 58–63.
[2] M. Fondevila, C. Phillips, N. Naverán, L. Fernandez, M. Cerezo, A. Salas, A. Carracedo, M.V. Lareu, Case report: identification of skeletal remains using short-amplicon marker analysis of severely degraded DNA extracted from a decomposed and charred femur, *Forensic Sci. Int. Genet.* 2 (2008) 212–218.
[3] R. Pereira, C. Phillips, C. Alves, A. Amorim, A. Carracedo, L. Gusmão, A new multiplex for human identification using insertion/deletion polymorphisms, *Electrophoresis* 30 (2009) 3682–3690.
[4] B.E. Krenke, A. Tereba, S.J. Anderson, E. Buel, S. Culhane, C.J. Finis, C.S. Tomsey, J. M. Zschetti, A. Masibay, D.R. Rabbach, E.A. Amiot, C.J. Sprecher, Validation of a 16-locus (fluorescent multiplex system), *J. Forensic Sci.* 47 (4) (2002) 773–785.

Male-specific contributions to the Brazilian population of Espírito Santo

Raquel de F. Figueiredo¹ · Isabela B. Ambrosio¹ · Danilo F. Braganholi¹ · Gustavo Chemale² · Joyce A. Martins¹ · Veronica Gomes³ · Leonor Gusmão^{3,4} · Regina M. B. Cicarelli¹

Received: 29 January 2015 / Accepted: 5 June 2015
© Springer-Verlag Berlin Heidelberg 2015

Abstract Y chromosome markers have been widely studied due to their various applications in the fields of forensic and evolutionary genetics. In this study, 35 Y-SNPs and 17 Y-STRs were genotyped in 253 males from the State of Espírito Santo, Brazil. A total of 18 haplogroups and 243 haplotypes were detected; the haplogroup and haplotype diversities were 0.7794 and 0.9997, respectively. Genetic distance analysis using the Y-STR data showed no statistically significant differences between Espírito Santo and other admixed populations from Brazil. The classification of paternal lineages based on haplogroups showed a predominant European contribution (85.88 %), followed by African (11.37 %) and Amerindian (2.75 %) contributions.

Keywords Y chromosome · Y-SNP haplogroup · Y-STR haplotype · Brazil

Brazil has a large and heterogeneous population originated from the admixture of Native Americans, Europeans, and Africans. As in most South American countries, the process of admixture in Brazil was very diverse depending on the country region, which created a highly sub-structured population. Therefore, the definition of the limits of a database to be used for forensic purposes is a complex task and implies the need for a full comprehension of all genetic variation that exists across the country.

Many studies of Y chromosome distribution in the diverse regions of the country [1–5] have been performed to fill some gaps in knowledge about the demographic events involved in the formation of Brazilian populations. Nevertheless, none of these studies have focused on the population of Espírito Santo State, making the present work the first report of Y-STRs combined with Y-SNPs in this population.

Blood samples were collected from 253 unrelated males from the State of Espírito Santo, located in the southeastern Brazil.

All samples were typed to 35 Y-SNP markers (see Fig. S1 for markers information and details on the methods used), as well as to a set of 17 Y-STR loci included in the Y-Filer (Life Technologies), following the manual instructions.

This project was approved by the Ethical Committee of FCFAR-UNESP (n. 12/2010) and all participants have signed an informed consent.

The data were deposited in the Y Chromosome Haplotype Reference Database—YHRD (<http://www.yhrd.org>; [6]) under the accession number YA003849.

Haplotypes and haplogroup diversities and population pairwise genetic distances (F_{ST}) were calculated using ARLEQUIN software version 3.1 [7].

Raquel de F. Figueiredo, Isabela B. Ambrosio and Danilo F. Braganholi contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s00414-015-1214-2) contains supplementary material, which is available to authorized users.

✉ Regina M. B. Cicarelli
cicarell@fcfar.unesp.br

¹ Laboratório de Investigação de Paternidade, Núcleo de Atendimento a Comunidade - Faculdade de Ciências Farmacêuticas, São Paulo State University (UNESP), Rodovia Araraquara-Jaú, km 01, 14801-902 Araraquara, São Paulo, Brazil

² Laboratório de Genética Forense, Departamento de Polícia Federal, Instituto Nacional de Criminalística, Diretoria Técnico-Científica, Brasília, Distrito Federal, Brazil

³ IPATIMUP, Institute of Molecular Pathology and Immunology of the University of Porto, 4200-465 Porto, Portugal

⁴ DNA Diagnostic Laboratory (LDD), State University of Rio de Janeiro (UERJ), Rio de Janeiro, Brazil

The haplotype data for the 17 Y-STR loci in the population of the State of Espírito Santo (Brazil) are presented in Table S1. Besides those usually present at DYS385, three rare cases of duplicated alleles were detected at DYS19 locus.

A high haplotype diversity was estimated (99.97 %); for a total of 243 different haplotypes identified in our sample, 234 of which being unique and 9 observed more than once in our sample (Table S2). The most frequent haplotype in our sample, present in three individuals from haplogroup R1b1-P25 (ID 13, ID 98, and ID 140), contains the core set of alleles that was previously described having a decreasing W–E gradient in Europe, with its highest frequency in the Iberian Peninsula [8].

Pairwise F_{ST} genetic distance analysis performed between admixed population samples from the five geopolitical states forming the Southeastern region in Brazil [3, 5] showed non-differentiation p values between Espírito Santo and São Paulo (0.00018), Minas Gerais (0.00014) and Rio de Janeiro (0.00024), for a significance level of 0.0083, obtained after applying the Bonferroni's correction. Therefore, the Espírito Santo sample was added to the Southeastern region and compared with populations from other Brazilian regions, accessible through the YHRD [6]: Northern (YP000574, $n=400$), Northeastern (YP000663, $n=205$), Central-Western (YP000793, $n=445$ and YP000754, $n=50$), and Southern (YP000759, $n=243$ and YP000757, $n=77$). The results showed no significant differentiation for Y-STR haplotype data between Southeastern and all the other regions (Table S3).

The frequencies of the 18 Y-SNP haplogroups identified in the 253 Brazilian male individuals in this study (Table S1) are presented in Fig. S1. As expected for an admixed population, high haplogroup diversity was observed (0.7794 ± 0.0229). This value was greater than those observed in Brazilian populations from Alagoas [2] and Rio de Janeiro [1, 5].

Among the European haplogroups, R1b1-P25 was the most frequently detected haplogroup in our sample, followed by haplogroup I-M170*(xM26) (Fig. S1). These haplogroups, together with lineages within paragroups I-M170 and J-12f2, are the most common in Europe and in the Brazilian populations from Rio de Janeiro [1, 5] and Alagoas [2]. Eight samples were classified in the haplogroup R1-M173. This European haplogroup was identified for the first time in a Brazilian population.

The African haplogroups E1a-M33, E2-M85, E1b1a-M191, and E1b1a-M154 were also detected in our sample (Fig. S1). One sample was classified as Y-A4*, the most basal clade of the Y chromosome phylogenetic tree, which is almost restricted to the African continent.

A low frequency of the main Amerindian haplogroup Q1a2-M3 was detected in this study (Fig. S1; Table S1).

Based on the detected Y-SNP haplogroups, we conclude that Europe is the main source of paternal lineages of the present-day population in the State of Espírito Santo (85.88 %). African and Amerindian sources only account for 11.37 and 2.75 % of the lineages, respectively.

Haplogroup frequencies were used to calculate population pairwise genetic distances (F_{ST}) between the studied sample and other 15 populations from Rio de Janeiro, Brazil [1], Alagoas, Brazil [2], Colombia [9], Argentina [9], Spain [9], Portugal [8], Germany [9], Denmark [9], Turkey [9], Mozambique [9], Somalia [9], Japan [9], Thailand [9], China [9], and Greenland [9] (Supplementary Table S4). The results were represented in a MDS plot (Fig. S2). This figure shows that the population from Espírito Santo is more similar to other Brazilian populations, and to populations from Iberian Peninsula (Spain and Portugal) and Latin America. South American populations tend to cluster together with European populations, which is in accordance with the known sex-biased formation of these populations, in which Europeans has the greatest contribution to paternal lineages.

Acknowledgments The authors wish to thank the “Centro de Hematologia e Hemoterapia do Espírito Santo” for assistance in sample collection. This study was supported by the PADCF/FCF - UNESP (proc. 2010/24-I) and Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) (proc. 2010/17220-5). RFF is supported by a fellowship from the “Conselho Nacional de Desenvolvimento Científico e Tecnológico” (CNPq); DFB, IBA, and JAM are supported by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). IPATIMUP is an Associate Laboratory of the Portuguese Ministry of Education and Science and is partially supported by FCT. VG is supported by Fundação para a Ciência e Tecnologia through grant SFRH/BPD/76207/2011.

References

1. Silva DA, Carvalho E, Costa G, Tavares L, Amorim A, Gusmão L (2006) Y-chromosome genetic variation in Rio de Janeiro population. *Am J Hum Biol* 18:829–837
2. de Azevedo DA, da Silva LAF, Gusmão L, de Carvalho EF (2009) Analysis of Y chromosome SNPs in Alagoas, Northeastern Brazil. *Forensic Sci Int Genet Suppl Ser* 2:421–422
3. Palha T, Gusmão L, Ribeiro-Rodrigues E, Guerreiro JF, Ribeiro-Dos-Santos A, Santos S (2012) Disclosing the genetic structure of Brazil through analysis of male lineages with highly discriminating haplotypes. *PLoS One* 7:e40007
4. Chemale G, Freitas JM, Badaraco JL, Rosa LD, Martins JA, Martins AD et al (2013) Y-Chromosomal STR haplotypes in a sample from Brasília, Federal District, Brazil. *Forensic Sci Int Genet*. doi:10.1016/j.fsigen.2013.06.005
5. Oliveira AM, Domingues PM, Gomes V, Amorim A, Jannuzzi J, Carvalho EF, Gusmão L (2014) Male lineage strata of Brazilian population disclosed by the simultaneous analysis of STRs and SNPs. *Forensic Sci Int: Genet* 13:264–268
6. Willuweit S, Roewer L (2015) The new Y chromosome haplotype reference database. *Forensic Sci Int Genet* 15:43–48

7. Excoffier L, Laval G, Schneider S (2005) Arlequin v. 3.0: an integrated software package for population genetics data analysis. *Evol Bioinform Online* 1:47–50
8. Gusmão L, Sánchez-Diz P, Alves C, Beleza S, Lopes A, Carracedo A, Amorim A (2003) Grouping of Y-STR haplotypes discloses European geographic clines. *Forensic Sci Int* 134:172–179
9. Brión M, Sanchez JJ, Balogh K, Thacker C, Blanco-Verea A, Borsting C et al (2005) Introduction of a single nucleotide polymorphism-based “Major Y-chromosome haplogroup typing kit” suitable for predicting the geographical origin of male lineages. *Electrophoresis* 26: 4411–4420



Contents lists available at ScienceDirect

Forensic Science International: Genetics Supplement Series

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



Mitochondrial DNA 30-SNP data confirm high prevalence of African lineages in the population of Espirito Santo, Brazil

Isabela Brunelli Ambrosio, Danilo Faustino Braganholi, Fernanda Silva Polverari, Regina Maria Barretto Cicarelli*

Laboratório de Investigação de Paternidade, Núcleo de Atendimento à Comunidade—Faculdade de Ciências Farmacêuticas, São Paulo State University (UNESP), Rodovia Araraquara-Jaú, km 01, 14801-902 Araraquara, São Paulo, Brazil

ARTICLE INFO

Article history:

Received 22 August 2015
Accepted 14 September 2015
Available online xxx

Keywords:

DNA mitochondrial
SNPs
SNaPshot
Haplogroups
Espirito Santo-Brazil

ABSTRACT

Mitochondrial DNA (mtDNA) analysis is usually a last resort in routine forensic DNA casework especially when nuclear DNA analysis cannot be applied. Because of the Brazilian population is highly admixed it is necessary to select highly discriminatory SNPs. The goal of this article is to characterize the mtDNA haplogroups in 100 unrelated individuals from the state of Espirito Santo, Brazil, by using 30-SNPs of mtDNA in two multiplex SNaPshot. The population was classified according to its origin in 43% of African, 26% of Native American, 30% of European and 1% Asian in 19 different haplogroups. These data confirmed therefore that this population is also highly admixed because of the history of country colonization.

© 2015 Published by Elsevier Ireland Ltd.

1. Introduction

Mitochondrial DNA (mtDNA) sequence analysis has become a powerful tool in the field of forensic sciences for missing persons, victims of mass disasters comparing the mother's genotype or a relative who has maternal relationship with the individual to be identified [1,2]. The Brazilian population is highly admixed, resulting mostly from interethnic crosses between Native Americans, Europeans and Africans [2]. Classification in haplogroups has been an important tool used in population genetics in forensic cases and genetic studies of disease [3]. Paneto et al. [4] analysed 42-SNPs (single nucleotide polymorphisms) of mtDNA to classify individuals in the population of state of São Paulo by SNaPshot. In this article we used 30-SNPs of mtDNA in two multiplex SNaPshot to characterize the haplogroups in individuals from the state of Espirito Santo, Brazil, a population which has no data about it.

2. Material and methods

Blood samples from 100 unrelated individuals who are residents in the state of Espirito Santo were used. All the

participants gave their written informed consent. 30 of the 42-SNPs described in Paneto et al. [4] were used in this study (Fig. 1). We used blood spotted on filter paper and DNA was extracted from two discs of 1.2 mm using Chelex (BioRad). PCR and SNaPshot reactions were made in two separate reactions, Multiplex I (21 SNPs) and Multiplex II (9 SNPs). All steps were performed according to the protocol described by Paneto et al. [8]. The capillary electrophoresis was performed on ABI 3500 Genetic Analyzer (Applied Biosystems by Life Technologies). The samples were classified into haplogroups using the phylogenetic tree described by Paneto et al. [4], and confirmed by van Oven and Kayser [5].

3. Results and discussion

It was found 19 different haplogroups (Fig. 2) classified according to the origin of the population: 43% of Africa, 26% of Native American, 30% of European and 1% Asian. Our results are in agreement with the history of Brazil colonization: it was populated by Native Americans; after discovery (1500), European immigration, mainly Portuguese, who arrived between 1500 and 1808 followed by Africans reflecting the period of slavery. Asian immigration also occurred, but to a lesser extent [2]. These data confirm, therefore, that the Brazilian population is highly racially mixed, result of ethnic crossings.

* Corresponding author. Fax: +55 16 3322 0073.
E-mail address: paternidade.fcfa@gmail.com (R.M.B. Cicarelli).

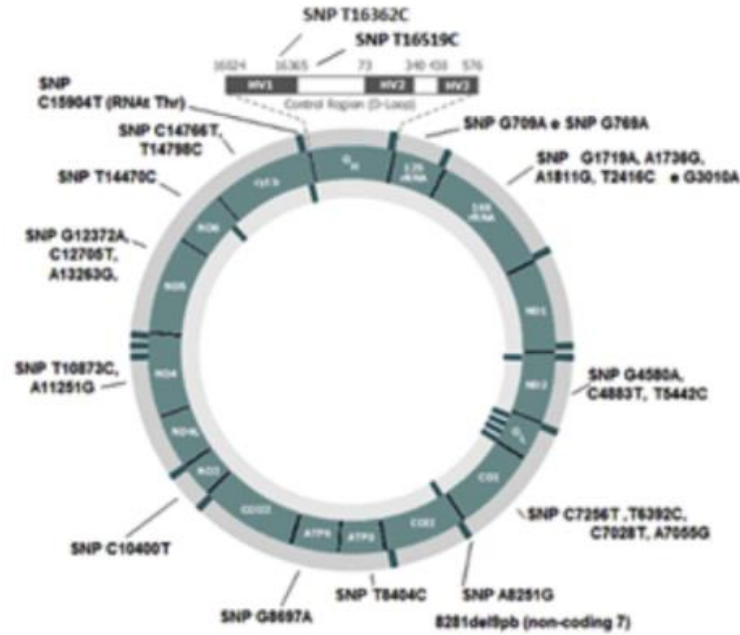


Fig. 1. General location of the mtDNA 30-SNPs.

Frequency of Haplogroups

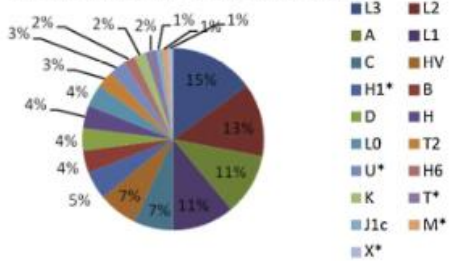


Fig. 2. Frequency of haplogroups in the population of Espirito Santo, Brazil.

References

- [1] G. Chemale, G.G. Paneto, M.A.M. Menezes, et al., Development and validation of a D-loop mtDNA SNP assay for the screening of specimens in forensic casework, *Forensic Sci. Int.: Genet.* 7 (2013) 353–358.
- [2] J. Alves-Silva, M.S. Santos, P.E.M. Guimarães, et al., The ancestry of Brazilian mtDNA lineages, *Am. J. Hum. Genet.* 67 (2000) 444–461.
- [3] M.D. Coble, R.S. Just, J.E. O'Callaghan, et al., Single nucleotide polymorphisms over the entire mtDNA genome that increase the power of forensic testing in Caucasians, *Int. J. Legal Med.* 118 (2004) 137–146.
- [4] G.G. Paneto, S. Köhnenmann, J.A. Martins, et al., A single multiplex PCR and SNaPshot minisequencing reaction of 42 SNPs to classify admixture populations into mitochondrial DNA haplogroups, *Mitochondrion* 11 (2011) 296–302.
- [5] M. van Oven, M. Kayser, Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation, *Hum. Mutat.* 118 (2004) 14–18.

Acknowledgments

This study was supported by PADCF/CF-UNESP, FAPESP and CAPES.

Please cite this article in press as: I.B. Ambrosio, et al., Mitochondrial DNA 30-SNP data confirm high prevalence of African lineages in the population of Espirito Santo, Brazil, *Forensic Sci. Int. Gene. Suppl.* (2015), <http://dx.doi.org/10.1016/j.fsigs.2015.09.137>



Analysis of SNP (single nucleotide polymorphism) multiplex markers related to sudden cardiac death in Brazilian families

D.F. Braganholi^{1,2} and R.M.B. Cicarelli¹

¹Laboratório de Investigação de Paternidade-NAC, Faculdade de Ciências Farmacêuticas, Universidade Estadual Paulista, Araraquara, SP, Brasil

²Instituto de Química, Universidade Estadual Paulista, Araraquara, SP, Brasil

Corresponding author: D.F. Braganholi
E-mail: danilobraganholi@hotmail.com

Genet. Mol. Res. 14 (4): 14348-14355 (2015)
Received June 8, 2015
Accepted August 12, 2015
Published November 13, 2015
DOI <http://dx.doi.org/10.4238/2015.November.13.20>

ABSTRACT. Sudden cardiac death (SCD) is a major public health concern worldwide, and genetic analysis may be useful in identifying the cause of death as well as in determining the possible genetic risk factors for SCD. This study analyzed eight SNPs (single nucleotide polymorphisms) highly correlated with cardiac sudden death in samples (blood and bone) from six Brazilian families with a history of cardiovascular diseases. Individuals with no family history of cardiovascular diseases were recruited as controls. Y chromosomes and mtDNA haplogroups belonging to these subjects were verified as well. We found that SNP rs16857031 showed significant differences between the group with a family history of cardiovascular diseases and the control group. Furthermore, the data obtained were compatible with known frequencies of SNPs for the haplogroups in which the samples were classified. A possible hereditary factor was identified for SNP rs4725982 in one family. These preliminary results suggest that identification of certain SNPs could be used to assess risk factors for SCD.

Key words: Cardiopathy; Sudden cardiac death; SNaPshot; SNP; Post mortem analysis

INTRODUCTION

Sudden cardiac death (SCD) is one of the most common causes of death in many countries. It is a major health issue, as sudden loss of cardiac function may lead to death in as little as a few minutes after the onset of symptoms. It is not due to trauma or violence, and can affect anyone from newborns to adults (Márban, 2002; Tester and Ackerman, 2006).

Population studies have shown that a significant number of SCDs are associated with hereditary channelopathies (Tester and Ackerman, 2006). It is recognized that changes in cardiac ion channel genes can cause sudden death without any morphological abnormalities in the heart (Márban, 2002). This can make it difficult to identify the cause of death in legal medicine reviews (Rodríguez-Calvo et al., 2008).

Post mortem genetic analyses (molecular autopsy) can potentially substantiate the pathogenic basis for the SCD, and may prevent further deaths in the same family as family history is known to be a significant risk factor (Rodríguez-Calvo et al., 2008).

In 2009, two teams of international researchers have identified several SNPs to be highly associated with QT interval syndromes (Newton-Cheh et al., 2009), a major cause of SCD. In this study, we chose eight SNPs, and examined their polymorphisms in Brazilian families. Comparisons were made between individuals with and without clinically diagnosed family history of cardiovascular diseases. In order to further complement our findings, subjects were classified into evolutionary haplogroups by polymorphisms in mitochondrial DNA and / or Y chromosome via sequencing and SNaPshot analysis, respectively. We assessed whether there may be a correlation between the occurrences of certain polymorphisms with different haplogroups. This analysis was also conducted on human remains in order to check the efficiency of this technique in deceased individuals.

MATERIAL AND METHODS

Selection of SNPs and design of primers

In the study by Newton-Cheh et al. (2009), 14 SNPs were listed as having the most associations with QT interval syndromes. In this study, we selected 8 of these SNPs (Table 1) for analyses. As Brazil has a heterogeneous population, we took into account any prior knowledge in the literature on the correlation of these genes with SCD in different populations (Manta et al., 2012).

Primers for PCR and SNaPshot reactions were designed and analyzed using the programs PerlPrimer version 1.1.19 (Marshall, 2004), AutoDimer version 1 (Vallone and Butler, 2004), and the online tool PCR *in silico* on the website <http://genome.ucsc.edu/> (Kent et al., 2002).

Amplification and detection of SNPs

PCR reactions were performed using 2 U Platinum Taq DNA polymerase (Invitrogen by Life Technologies); 1X Goldstar Buffer (Promega); and appropriate concentration of each primer (Tables S1 and S2); 0.5 ng DNA in a final volume of 11 μ L.

The following cycling parameters were performed on the Veriti thermal cycler (Applied Biosystems by Life Technologies): 95°C for 10 minutes; 35 cycles at 95°C for 30 s; 51°C for 45 s (increasing 0.3°C per cycle from the fourth cycle); 72°C for 30 s; holding at 4°C.

PCR products were purified prior to completion of the SNaPshot reaction via addition of 1 μ L EXO-SAP IT (GE Healthcare) with the cycling parameter as follows: 37°C for 95 min; 78°C for

15 min; hold at 4°C.

SNaPshot reactions were performed using the ABI PRISM SNaPshot Multiplex Kit (Applied Biosystems by Life Technologies) as follows: PCR Buffer II to 0.38X (Applied Biosystems by Life Technologies); 1.25 µL SNaPshot mix; and appropriate concentration of each primer (Tables S3 and S4); 1 µL purified PCR products in a final volume of 6.5 µL. The cycling parameters are as follows: 30 cycles at 96°C for 3 min; 55°C for 5 s; 60°C for 30 s; hold at 4°C.

The reaction product was purified by addition of 1 µL SAP (GE Healthcare) with the following cycling protocol: 37°C for 95 min; 78°C for 15 min; hold at 4°C.

Products of SNaPshot reactions were prepared for capillary electrophoresis (CE) by the addition of 1 µL purified SNaPshot reaction products to 10 µL HI-DI formamide and LIZ 120 (12.5 µL LIZ 120 to 1000 µL formamide HI-DI, Applied Biosystems by Life Technologies). CE was performed on an ABI 3500 Genetic Analyzer (Applied Biosystems by Life Technologies) using the filter dye set E5 and POP-7 polymer. Analyses were performed using the GeneMapper software - IDX version 1.2 (Applied Biosystems by Life Technologies). Patent application for this methodology was filed to the INPI (National Institute of Industrial Property) under number: BR102012028699-8.

Table 1. Location and related genes on chromosomes of 8 SNPs selected for this study.

SNP	Chromosome	Gene	Ancestral allele	Derived allele
rs16857031	1q	Intron <i>NOS1AP</i>	C	G
rs2074238	11p	Intron <i>KCNQ1</i>	C	T
rs12576239	11p	Intron <i>KCNQ1</i>	C	T
rs4725982	7q	Downstream <i>KCNH2</i>	C	T
rs1805128	21q	Missense <i>KCNE1</i>	G	A
rs12053903	3p	Intron <i>SCH5A</i>	C	T
rs2074518	17q	Intron <i>LIG3</i>	G	A
rs2988884	7q	Downstream <i>KCNH2</i>	A	G

Samples

Blood samples from individuals of five Brazilian families (N = 28) were collected on FTA cards (Whatman). Subjects were clinically diagnosed with a history of heart problems, Coronary Disease, Hypertrophic Cardiomyopathy, or had a family member who suffered a heart attack.

Samples belonging to male individuals were classified into evolutionary haplogroups by SNPs analysis of the Y chromosome (Brion et al., 2005). Blood samples from individuals who had no family history of heart diseases were used as controls (N = 30). This study was approved by the Ethics Committee of the Faculty of Pharmaceutical Sciences - Araraquara by Protocol: 24/2009.

DNA samples from bones and teeth were obtained from five exhumed corpses belonging to the same family, three of which with cardiovascular disorder as the cause of death. These samples were named as MPV, HV, MLV, CLV and OVB: MPV, 42 years, cardiac injury; HV, 42 years, pleuropneumonia; MLV, 79 years, myocardial infarction; CLV, 82 years, head trauma; OVB, 52 years, acute myocardial infarction/coronary insufficiency.

These samples were classified into haplogroups through evolutionary analysis of the hypervariable regions I and II of the mtDNA (Paneto et al., 2007). DNA extraction from blood samples was performed with Chelex resin (Biorad) (Singer-Sam et al., 1989).

For DNA extraction, samples of bones and teeth were subjected to demineralization, (Loreille et al., 2007) followed by centrifugation using the AMICON ULTRA filter (Millipore). Extraction procedures were carried out with the DNA IQ Kit (Promega) according to manufacturer instructions.

RESULTS AND DISCUSSION

SNPs analysis related to SCD

The amplified SNPs are shown in the electropherogram (Figure 1). Some artifacts were present, but did not have an effect on the analyses of the expected peaks for the SNPs. The primer sequences and the concentrations at which they were used are described in the supplementary material.

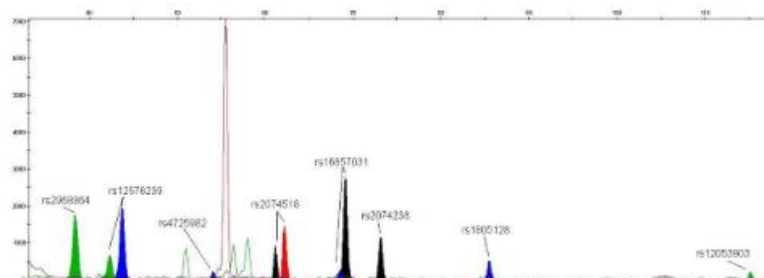


Figure 1. Electropherogram of SNaPshot reaction of eight SNPs analyzed in multiplex system.

We identified the presence of the derivative allele "T/T", which was identical for the SNP rs12053903 in all samples of from both groups. (Figure 2).

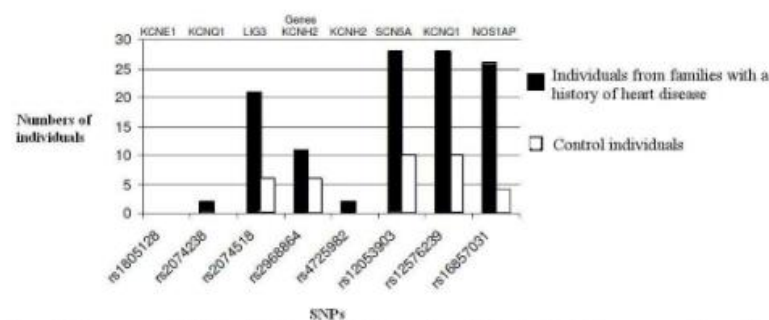


Figure 2. Occurrence of 8-SNPs with their respective frequencies which the derived alleles are present (regardless of whether homo or heterozygous) in samples of individuals from families with a history of heart disease and controls.

For the SNP rs12576239, in the group with a family history of cardiovascular diseases, 20 samples showed the alleles "C/T" and 8 samples showed the alleles "T/T". However, in the control group, we identified 14 samples with the alleles "T/T" and 16 with "C/T" (Figure 3).

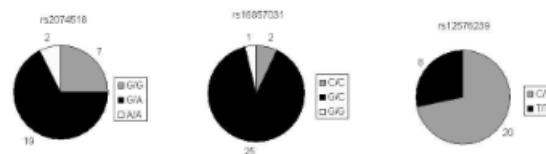


Figure 3. Profiles observed in samples of individuals with a family history of heart disease for SNPs rs2074518, rs16857031 and rs12576239.

Twenty-five samples of the family history group presented tested for the SNP rs16857031 were heterozygous for that locus, namely the original or the ancestor base (C), and the derivative base which G. In addition, one sample presented with the homozygous "G/G", while two samples alleles "C/C". In the control group, the alleles "G/C" was found in nine samples, one sample showed the profile "G/G", and 20 samples showed no exchange of bases, i.e., the alleles "C/C" (Figure 3).

In summary, no significant differences were found between individuals with family history of heart disease and healthy controls. However, the SNP rs16857031 was identified in a higher percentage of individuals with family history of cardiovascular complications compared to the controls. This polymorphism is located in the NOS1AP gene, which was previously described as being closely associated with heart diseases and SCD (Eijgelsheim et al., 2009). The SNP rs2968864 was identified in greater proportion of the control group, which was contrary to what was expected. However, it may be explained by the frequency of this polymorphism in the population, as will be discussed later.

While these results are still preliminary, we have demonstrated that SNP sequencing can be used efficiently in a clinical setting.

Bone and tooth samples

We were unable to analyze the eight SNPs as a multiplex with the bone and teeth samples. The SNP rs16857031 is the only polymorphism displayed on the electropherograms of the SNaPshot reactions for the five samples, and SNP rs2074238 was only amplified in one sample (MPV).

For the samples OVB, CLV, HV, and MLV, the SNP rs16857031 was identified as homozygous with the derivative allele "G/G" and for MPV, the heterozygous "G/C" was observed. It is possible that the SNP rs16857031 may be related to the cause of death for MPV, MLV and OVB, who died from heart failure.

In the MPV sample, the alleles for the rs2074238 SNP was observed to be the homozygous "C/C", which are the ancestral alleles.

Correlation between SNPs related to SDC and mitochondrial DNA haplogroups in bones and teeth samples

The SNP rs16857031 with "G/G" alleles has a low frequency in Europeans and was identified in the samples OVB, CLV, HV, and MLV, indicating a possible hereditary factor. The alleles "C/C", which was identified in the MPV sample, have a median frequency. As mentioned previously, the results were suggestive that these alleles may not be related to the population of origin, but are rather with cardiac irregularities. Therefore, the samples OVB, MLV, CLV, and HV

were classified by mtDNA into the "H" haplogroup, whereas the MPV sample was classified as the "T2d" haplogroup, both frequent in Europeans (Babalini et al., 2005).

SNP rs2074238 with the derived allele "T" was not identified in the MPV sample, which showed the profile "C/C". This form was found to be very common in Europeans, which correlated with the fact that MPV was born in Portugal whereas the others were born in Brazil. Figure 4 presents the heredogram of this family.

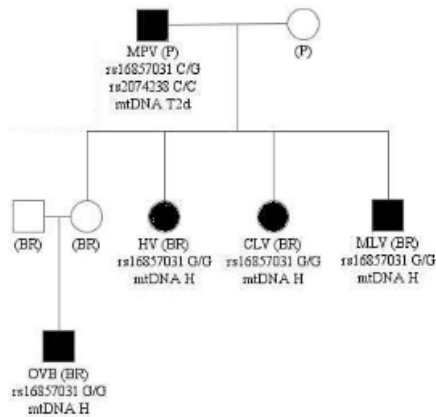


Figure 4. Ancient remains heredogram. In black, the remains analyzed in this paper showing their respective alleles for SNPs rs2074238, rs16857031 and mtDNA haplogroup. It is indicated the birth country (P = Portugal; BR = Brazil).

Correlation between SNPs related to SDC and Y chromosome haplogroups

The genetic characteristics that affect the QT interval may be inherited in approximately 35% of the general population (Eijgelsheim et al., 2009). The genetic inheritance can be illustrated in the analysis of the SNP rs4725982 in two male individuals from family eight. Both the father and son were presented with the polymorphic base "T" for this SNP. Following analyses of the Y chromosome, these individuals were classified into the rare T haplogroup, which is found in low frequencies in Europe and parts of the Middle East, North and West Africa (King et al., 2007).

The frequency of the alleles "T/T" in Europeans and African is also relatively low (<http://www.ncbi.nlm.nih.gov/snp/>) which may indicate some correlation between the onset of heart disease or SCD in male individuals belonging to this particular haplogroup. Confirmation of this hypothesis requires larger sample of individuals with the syndrome.

The profile "C/C" in the same locus found in samples belonging to the other families have a high frequency in both Europeans and Africans, which correlates with the populations included in this study. Almost all subjects were classified into the European haplogroups once the frequency of haplogroup J2 was identified in Spain and Germany, and absent in China and Japan (Brion et al., 2005); R1b1 is highly common in Europe; haplogroup B is almost entirely restricted to sub-Saharan Africa (Karafet et al., 2008), C is quite found in Asia (Hammer et al., 2001), as haplogroup D, and

it is not found anywhere outside of Asia (Karafet et al., 2001) and haplogroup E is found in high frequency in Africa and moderate frequencies in Europe (Jobling and Tyler-Smith, 2000).

The results presented in this study are in agreement with the literature as it recognized the predominance of European lineages for in the Brazilian populations (Carvalho-Silva et al., 2001).

The finding that the SNP rs1805128 was not identified with the derived allele "A" in any of the samples and rs2074238 with the derived allele "T" in only two samples can be explained by the high frequency of "G/G" and "C/C" alleles in Europeans.

On the other hand, the rs12576239 SNP with allele derivative "T" was identified in all samples and has a median frequency for the profile "C/T" and almost zero for the profile "T/T" in Europeans. The rs16857031 SNP, which was also identified in all samples with the allele "G", has low to medium frequency for the profiles "C/G" and "G /G", respectively. These evidences suggest that these alleles are not population-specific, but rather, are associated with a family history of cardiac irregularities.

Frequency of the profiles "A/A" and "A/G" in SNP rs2968864 is high in Europeans. This agrees with the results from our study, which showed high percentage of the "A/G" alleles in the individuals classified into the European haplogroups.

The frequency of the allele "A" in SNP rs2074518 was found to be slightly higher in the group with a family history of cardiac diseases as compared to the controls. In addition, Europeans show high occurrences of profile "A/G", and median occurrences for the alleles "A/A" and "G/G" (<http://www.ncbi.nlm.nih.gov/snp/>).

The possibility that ancestral origin may influence the development of cardiovascular diseases is an interesting notion (Fridman et al., 2011). In this preliminary study, we have established a reliable method for SNP analysis related to SCD, which can be used in a larger number of samples to confirm the association between ancestral origin and disease.

CONCLUSIONS

The data are consistent with the known frequencies of the SNPs; SNP rs16857031 presented the polymorphic base with greater frequency in individuals with a family history of heart failure compared with the control group.

Lastly, we recommend the use of *monoplex* for analysis of SNP samples from partially degraded DNA obtained from ancient remains (exhumed bones and/ or teeth).

Conflicts of interest

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

The authors would like to thank José Antonio Caracciolo, medical of CEDCOR clinic, for their assistance in sample collection. Research supported by the Coordenação de aperfeiçoamento de pessoal de nível superior (#CAPES) and Programa de apoio ao desenvolvimento científico (#PADC-FCFAr: 2015/14-III).

[Supplementary material](#)

REFERENCES

- Babalini C, Martinez-Labarga C, Tolk HV, Kivisild T, et al. (2005). The population history of the Croatian linguistic minority of Molise (southern Italy): a maternal view. *Eur. J. Hum. Genet.* 13: 902-912.
- Brion M, Sobrino B, Blanco-Verea A, Lareu MV, et al. (2005). Hierarchical analysis of 30 Y-chromosome SNPs in European populations. *Int. J. Legal Med.* 119: 10-15.
- Carvalho-Silva DR, Santos FR, Rocha J and Pena SDJ (2001). The Phylogeography of Brazilian Y-Chromosome Lineages. *Am. J. Hum. Genet.* 68: 281-286.
- Eljgishelm M, Aarnoudse AL, Rivadeneyra F, Kors JA, et al. (2009) Identification of a common variant at the *NOSTAP* locus strongly associated to QT-Interval duration. *Hum. Mol. Genet.* 18: 347-357.
- Fridman C, Cardena M, Krieger JM and Pereira AC (2011). Evaluation of the relationship between mitochondrial haplogroup and development of heart failure in Brazilian sample. *Forensic Sci. Int. Genet.* 3: 77-78.
- Hammer MF, Karafet TM, Redd AJ, Jarjanazi H, et al. (2001). Hierarchical patterns of global human Y-chromosome diversity. *Mol. Biol. Evol.* 18: 1189-1203.
- Jobling MA and Tyler-Smith C (2000). New uses for new haplotypes: the human Y chromosome, disease, and selection. *Trends Genet.* 16: 356-362.
- Karafet T, Mendez FL, Melleran MB, Underhill PA, et al. (2005). New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Res.* 15: 830-838.
- Karafet T, Xu L, Du R, Wang W, et al. (2001). Paternal population history of East Asia: sources, patterns, and microevolutionary processes. *Am. J. Hum. Genet.* 69: 615-628.
- Kent WJ, Sugnet CW, Furey TS, Roskin KM, et al. (2002). The human genome browser at UCSC. *Genome Res.* 12: 996-1006.
- King TE, Bowden GR, Salaresque PL, Adams SM, et al. (2007). Thomas Jefferson's Y chromosome belongs to a rare European lineage. *Am. J. Phys. Anthropol.* 132: 583-589.
- Loreille OM, Diegoli TM, Irwin JA, Coble MD, et al. (2007). High efficiency DNA extraction from bone by total demineralization. *Forensic Sci. Int.* 1: 191-195.
- Manta F, Calafá A, Pereira R, Silva D, et al. (2012). Indel markers: Genetic diversity of 38 polymorphisms in Brazilian populations and application in a paternity investigation with post mortem material. *Forensic Sci. Int. Genet.* 6: 658-661.
- Marbán E (2002). Cardiac channelopathies. *Nature* 415: 213-218.
- Marshall OJ (2004). PeriPrimer: cross-platform, graphical primer design for standard, bisulphite and real-time PCR. *Bioinformatics* 20: 2471-2472.
- Newton-Cheh C, Eljgishelm M, Rice KM, de Bakker PI, et al. (2009). Common variants at ten loci influence QT interval duration in the QTGEN Study. *Nat. Genet.* 41: 399-406.
- Paneto GG, Martins JA, Longo LV, Pereira GA, et al. (2007). Heteroplasmy in hair: Differences among hair and blood from the same individuals are still a matter of debate. *Forensic Sci. Int.* 173: 117-121.
- Rodríguez-Calvo MS, Brion M, Allegue C, Conchello L, et al. (2008). Molecular genetics of sudden cardiac death. *Forensic Sci. Int.* 182: 1-12.
- Singer-sam J, Tanguay R and Riggs AD (1989). Use of Chelex to improve the PCR signal from a small number of cells. *Amplifications* 3: 11.
- Tester DJ and Ackerman J (2006). The role of molecular autopsy in unexplained sudden cardiac death. *Curr. Opin. Cardiol.* 21: 166-172.
- Vallone PM and Butler JM (2004). AutoDimer: a screening tool for primer-dimer and hairpin structures. *BioTechniques* 37: 226-231.



DNA human identification for the family rights: One year of agreement between UNESP/Laboratory of Paternity and Public Defender Service in São Paulo State



J.A. Martins, D.F. Braganholi, J.H.P. Felix, C.F. de Almeida, I. Brunelli, V.L.S. Alvarenga, R.M.B. Cicarelli*

Faculdade de Ciências Farmacêuticas, UNESP – Univ Estadual Paulista, Laboratório de Investigação de Paternidade, Araraquara, SP, Brazil

ARTICLE INFO

Article history:
Received 29 August 2013
Accepted 2 October 2013

Keywords:
Family right
DNA
UNESP
Public Defender Service
São Paulo State
Brazil

ABSTRACT

The universities have realized the importance of extending their knowledge to the population through the provision of services. Thus, this paper presents the data obtained in an agreement between UNESP/Laboratory of Paternity and Public Defender Service in São Paulo State to make DNA paternity tests.

© 2013 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

In Brazil, there is a high number of children with incomplete birth registration data. Our laboratory initiated the activities in March 2001 and recently, in October 2011, signed an agreement with the Public Defender Service of the State of São Paulo (DPESP), aiming to provide individuals without financial conditions with DNA tests. These individuals should be entitled to legal assistance and living in a ratio of 100 km from the Araraquara region (Headquarters), São Paulo State.

2. Materials and methods

Blood samples were collected on FTA cards and genomic DNA was extracted using the FTA method (Whatman). PCR amplification of autosomal (AS) short tandem repeat (STR) markers was carried out using the PowerPlex[®]16 System (Promega) from December/2011 to February/2012; after that PowerPlex[®]18D System (Promega) was used from March to July and PowerPlex[®]21 System (Promega) from August to December of 2012. When necessary, 12 Y-STR (PowerPlex[®] Y System, Promega) and 10 X-STR [1] markers were analyzed. The latter were amplified as proposed by Martins et al. [2]. The PCR products were separated on

an Applied Biosystems 3500 Genetic Analyzer and genotyping was performed using GeneMapper[®] ID-X Software (Applied Biosystems). Biostatistical analysis was carried out using the Familias program version 1.81 [3]. Each month, a list of the services provided was sent to DPESP for the payment to be processed.

3. Results and discussion

In the period from December/2011 to December/2012, 196 examinations were performed typing 605 people, including 165 trios (mother, son and alleged father); nine duo (child and alleged father); eleven reconstructions (five solved by AS and X STRs and six by AS and Y STRs); four with mother, son, registry father and alleged father; six involving mother, two sons and alleged father, and one with alleged mother (Fig. 1). Non-exclusions were obtained in 130 of 196 cases (6633%) and, not considering the

Table 1
Genetic inconsistencies observed in complete trios.

Cases	STR	Mother	Child	Alleged father
Case 1	D18S51	15/21	16/20	16/20
Case 2	D18S51	14/19	15/19	14/16
Case 3	D8S1179	13/14	13/15	14
Case 4	FGA	21/23	21/26	21/25
	D12S391	18.3/23	18.3/26	17/25
Case 5	D18S51	14/21	21/28	15/27

Bold numbers indicate genetic inconsistencies between the child and the alleged father or mother.

* Corresponding author. Tel.: +55 16 3301 6950; fax: +55 16 3301 6940.
E-mail address: paternidade.fcfar@gmail.com (R.M.B. Cicarelli).

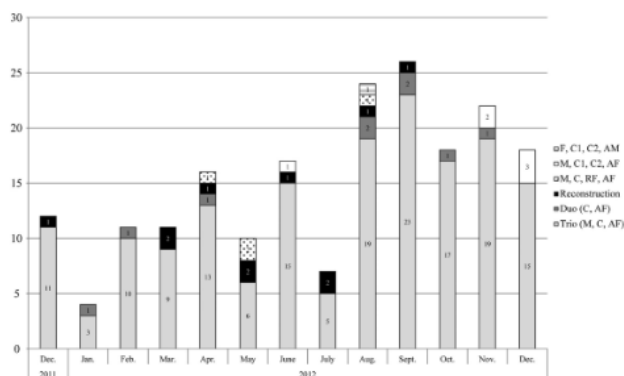


Fig. 1. Cases of biological relationships performed from December/2011 to December/2012. Note: F, father; C, Child; AM, alleged mother; M, mother; RF, registry father; AF, alleged father.

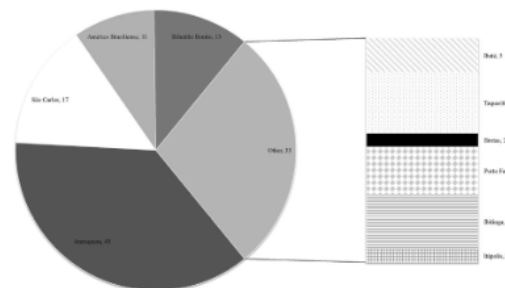


Fig. 2. Relation of cities to number of cases required by Court.

reconstructions, the paternity index (PI) ranged between 139.521.43 for PowerPlex[®]16 System and 270.247.652.546.228.00 for PowerPlex[®]21 System. A total of six genetic inconsistencies were found: three to D18551 and one to FGA, D8S1179 and D12S391 (Table 1). DNA tests were required by public defenders from Araraquara (n = 55) and São Carlos (n = 24) and Court from several cities (n = 117), mainly from Araraquara (n = 49) and São Carlos (n = 17) (Fig. 2).

4. Conclusion

In conclusion, the Laboratory has a university outreach community service program provided by highly qualified human resources with a team including two academic supports, one Master, one PhD, one specialist student, one specialist professional and one coordinating professor.

Role of funding

J.H.P. Felix and C.F. de Almeida were supported by BAAEII-PROEX-REITORIA-UNESP; I. Brunelli is supported by CAPES; D.F. Braganholi is supported by CNPq; J.A.M. is supported by FUNDECIF.

The funding sources had no involvement in the development of the paper or decisions related to the paper.

Conflict of interest

None.

Acknowledgements

The authors would like to thank Public Defender Service in São Paulo State by agreement signed with UNESP/Laboratory of Paternity.RMBC was partially supported by PADC 2013/18-II and Fundunesp 0035/015/13-PROPE/CDC.

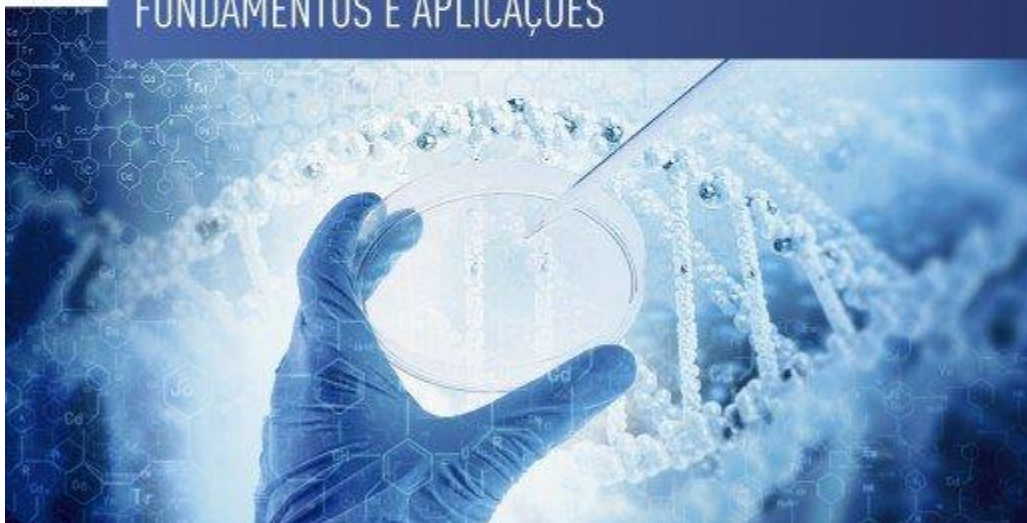
References

- [1] L. Gusmão, P. Sánchez-Diz, C. Alves, et al., A CEP-ISFG collaborative study on the optimization of an X-STR decaplex: data on 15 Iberian and Latin American populations, *Int. J. Legal Med.* 123 (2009) 227–234.
- [2] J.A. Martins, J.C. Costa, G.C. Pinheiro, et al., Genetic profile characterization of 10 X-STRs in four populations of the southeastern region of Brazil, *Int. J. Legal Med.* 124 (2010) 427–432.
- [3] T. Egeland, P.F. Mostad, B. Mervåg, et al., Beyond traditional paternity and identification cases. Selecting the most probable pedigree, *Forensic Sci. Int.* 110 (2000) 47–59.

3

BIOTECNOLOGIA APLICADA À SAÚDE

FUNDAMENTOS E APLICAÇÕES



organizador

**RODRIGO RIBEIRO
RESENDE**

colaboradores

**SILVIA GUATIMOSIM
MARCUS VINICIUS GOMEZ
CARLOS R. SOCCOL**

Blucher



20

CAPÍTULO

APLICAÇÕES FORENSES DO DNA: FUNDAMENTOS, MÉTODOS E LIMITAÇÕES

Isabela Brunelli Ambrósio
Danilo Faustino Braganholi
Regina Maria Barretto Cicarelli

20.1 INTRODUÇÃO

A identificação humana é o processo científico por meio do qual se determina a identidade de uma pessoa, através de um conjunto de caracteres que a individualize, fazendo-a igual apenas a si mesma¹. Esta identificação por meio do DNA constitui um dos produtos mais revolucionários da genética moderna.

A genética forense, também conhecida como DNA Forense, é a área que utiliza os conhecimentos e técnicas de genética e de biologia molecular no auxílio à justiça, sendo aceita rotineiramente em processos judiciais em todo o mundo.

Apesar de ser a identificação humana pelo DNA o ramo mais desenvolvido da genética forense, bem como sua aplicação mais popular ser o teste de paternidade, ela não se limita a isso, e pode ser aplicada na identificação ou individualização de animais, plantas e micro-organismos.

A genética forense iniciou-se quando foram utilizadas pela primeira vez características genéticas para a definição de paternidade, ajudando a justiça na solução de um caso criminal^{2,3}. Sua fase moderna teve início na década de 1980, quando pesquisadores descobriram regiões altamente variáveis do DNA, capazes de individualizar uma pessoa^{2,3,4}. Em 1985, sir Alec Jeffreys⁵ apelidou as características únicas do DNA de uma pessoa de “impressões digitais do DNA”. O perfil genético de um indivíduo, comumente utilizado na identificação humana, é baseado na combinação de diversos marcadores que são herdados de seus progenitores. Esses marcadores referem-se geralmente a diferenças nas sequências de DNA entre os indivíduos (polimorfismos).

Os marcadores genéticos utilizados em identificação humana podem ser biparentais (metade dos cromossomos herdados da mãe e a outra metade herdada do pai) ou uniparentais (cromossomo Y herdado apenas do pai em indivíduos do sexo masculino e DNA mitocondrial herdado apenas da mãe em todos os indivíduos). Os marcadores biparentais autossômicos permitem uma identificação individual porque um determinado perfil genético é específico de um só indivíduo, sendo apenas partilhado por gêmeos monozigóticos. Os marcadores uniparentais não permitem a identificação individual, pois todos os indivíduos aparentados por via masculina partilharão a mesma linhagem de cromossomo Y, enquanto todos os indivíduos aparentados por via feminina partilharão a mesma linhagem de DNA mitocondrial, salvo nos casos de ocorrência de mutações durante a transmissão.

20.2 POLIMORFISMOS STRS (AUTOSSÔMICOS E DOS CROMOSSOMOS SEXUAIS, X E Y)

Desde que se desenvolveu a tecnologia do DNA recombinante, em 1973, dividindo em pedaços o DNA genômico para estudar os genes, houve um avanço na biologia molecular, e diversos genes foram descritos e isolados, muitos deles relacionados com a identificação humana⁵. Com o uso de enzimas de restrição e sondas de DNA, foram exploradas as variabilidades ou polimorfismos do DNA que variavam de indivíduo para indivíduo. Descobriram-se, também, regiões constituídas por vários pares de bases, repetidas inúmeras vezes, com alelos variando quanto ao número de sequências, denominados minissatélites (*variable number tandem repeats* – VNTR). Inicialmente, os VNTRs foram utilizados como marcadores genéticos informativos, pois estes são compostos por cem ou mais alelos com diferentes comprimentos na população humana

Outros marcadores altamente polimórficos foram encontrados no genoma humano: os microssatélites (*short tandem repeats*, STR) (Figura 20.1). Os STRs são repetições curtas de 1 a 5 pares de bases encontrados em eucariotos. Tais repetições são encontradas tanto em regiões codificantes quanto não codificantes do DNA, cuja função ainda é desconhecida⁵.

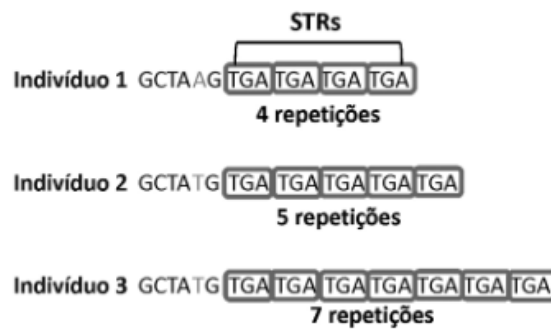


Figura 20.1 Esquema representando variações dos polimorfismos *short tandem repeats* (STRs). O número de repetições das pequenas sequências configuram um alelo que pode variar entre indivíduos.

Inovação tecnológica revolucionária na biologia molecular surgida em 1985, a reação da polimerase em cadeia ou *polymerase chain reaction* (PCR), tornou-se de uso universal. Por meio dessa reação, quantidades muito pequenas de DNA podem ser amplificadas para obtenção de praticamente qualquer quantidade desejada^{6,7,8}. A técnica utiliza o mesmo princípio pelo qual o DNA é normalmente copiado na célula, exceto pelo fato de que somente um pequeno segmento do DNA no cromossomo será amplificado. Isso possibilitou processar quantidades muito pequenas de DNA frequentemente deixadas como evidência de um crime e aumentou enormemente a sensibilidade dos sistemas disponíveis usados nas análises forenses, auxiliando o sistema judiciário no julgamento de crimes. Graças à PCR, diminutas quantidades de DNA podem ser extraídas de fios de cabelo, selos postais, pontas de cigarro, xícaras de café e outras fontes de DNA.

Os STRs substituíram os polimorfismos de comprimento dos fragmentos de restrição (do inglês, *restriction fragment length polymorphisms* – RFLPs) como principal método de identificação genômica e tornaram-se popular na tipagem de DNA para análises forenses. São sequências curtas que variam de

duas a quatro bases, podendo repetir-se até dezessete vezes (de sete a quinze alelos por locus), comumente amplificadas por PCR. Os métodos para tipagem de STRs estão atualmente automatizados, com detecção envolvendo fluoróforos⁹.

Com o desenvolvimento de reações da PCR em multiplex (Figura 20.2), ou seja, amplificação de várias regiões do DNA em um único tubo, empresas internacionais iniciaram a comercialização de kits contendo um *mix* ou mistura com diversos iniciadores (*primers*), que permitem amplificar até 24 regiões do DNA genômico, sendo 23 *loci* STR mais a amelogenina para determinação do sexo. Entretanto, para aplicação e uso nos diferentes laboratórios forenses, há a necessidade de se conhecer as frequências dos respectivos alelos presentes na população daquela região geográfica em que os testes estão sendo realizados. Tais frequências devem ser publicadas em revistas indexadas para que possam ser utilizadas nas análises estatísticas dos resultados dos perfis que farão parte da conclusão do laudo pericial.

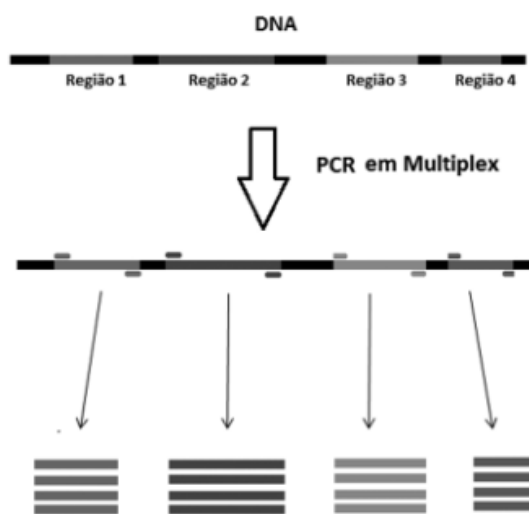


Figura 20.2 Esquema representando uma reação de PCR multiplex, na qual são amplificadas diferentes regiões do DNA em uma única reação.

Tratando-se de amostras obtidas em cenas de crimes, casos de violação sexual (estupro), desastres em massa ou de restos mortais de exumação de cadáver, uma questão fundamental, concernente ao uso do DNA como evidência, está na validação científica dos métodos de análise dos STRs. Em outras palavras, é preciso ter garantias científicas de que os testes podem, inequivocamente, identificar coincidências ou não coincidências para cada marcador genético utilizado, na comparação das amostras questionada e de referência. Inicialmente, a credibilidade dos testes deve partir da natureza das amostras biológicas utilizadas². Muito frequentemente as amostras são encontradas em superfícies não esterilizadas, podendo sofrer danos após contato com a luz solar, micro-organismos e solventes. Existem procedimentos que podem minimizar a ação desses fatores de degradação do DNA. Entretanto, muitos cuidados devem ser tomados para evitar equívocos na interpretação dos resultados dos perfis alélicos. Ressalte-se que a cadeia de custódia das amostras deve ser mantida até o momento de serem processadas no laboratório.

A amplificação por PCR pode produzir falhas e artefatos devido à baixa qualidade do material biológico. Amostras de DNA parcialmente degradadas podem proporcionar, por exemplo, a amplificação preferencial de alguns alelos e/ou o surgimento das bandas fantasmas (*stutter bands*) (Figura 20.3)². No primeiro caso, pode-se ter a amplificação de um alelo em detrimento do outro, o que pode gerar a falsa impressão de se tratar de um indivíduo homocigoto em vez de heterocigoto para o locus em estudo. As bandas fantasmas ocorrem em virtude de falhas no processo de amplificação, gerando bandas com uma unidade de repetição a menos que a do alelo original. Desse modo, pode-se equivocadamente interpretar o resultado como um falso heterocigoto ou identificar um alelo erroneamente.

Apesar dessas dificuldades, a genética é a área da ciência forense que mais tem avançado na atualidade. Há alguns anos, a Secretaria Nacional de Segurança Pública do Brasil está implementando o Banco de Dados Nacional Criminal de Perfis Genéticos, similarmente ao americano CODIS (Combined DNA Index System), que armazena dados de criminosos condenados, e o europeu Fenix, que contém o perfil genético de milhares de pessoas desaparecidas. Tais ferramentas tornam mais ágil a troca de informações entre as instituições espalhadas pelo vasto território no país, como no caso do CODIS nos Estados Unidos da América¹⁰, facilitando a resolução de diversos casos.

No Brasil, a implementação desse banco de dados trouxe o aumento da demanda nos laboratórios de perícia, uma vez que ele permite, por exemplo,

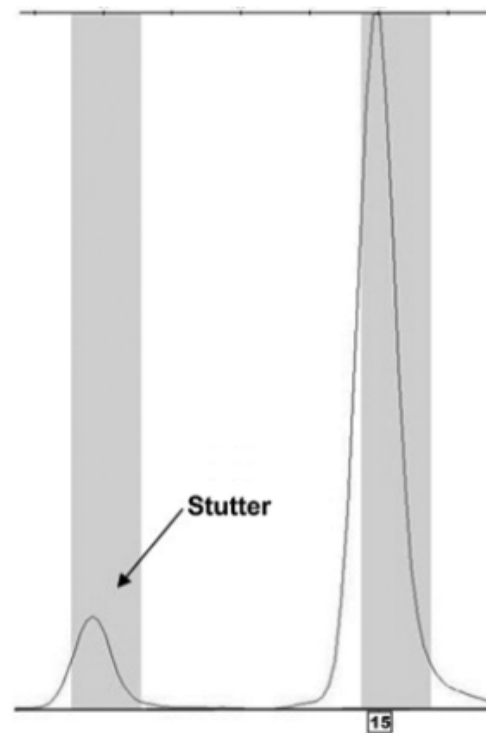


Figura 20.3 Esquema representando um alelo e um *stutter*. A seta indica o *stutter*. Nota-se a grande diferença entre as intensidades de sinais entre *stutter* e alelo verdadeiro.

identificar o criminoso pela análise de uma única gota de sangue encontrada no local do crime.

Em 2006, alguns peritos judiciais publicaram artigos^{11, 12} revelando as possibilidades de erros em exames de DNA, o que acarretou um convite da revista *Scientific American Brasil* para escrever um artigo sobre o estado da arte da genética forense no Brasil¹³. Em decorrência dessas publicações, vários laboratórios brasileiros começaram a ser condenados por erros em exames de DNA.

Na falta de um órgão nacional credenciado para promover a acreditação de laboratórios de DNA forense no Brasil, como é o caso, por exemplo, da American Association of Blood Banks (AABB) nos Estados Unidos, o

Ministério da Justiça solicitou que o Instituto Nacional de Metrologia, Qualidade e Tecnologia (INMETRO) viesse a padronizar os exames de DNA realizados por laboratórios da polícia científica¹⁴. Mais recentemente, a Agência Nacional de Vigilância Sanitária (ANVISA) publicou a Resolução RDC n. 11, de 16 de fevereiro de 2012, estabelecendo normas para o funcionamento de laboratórios que realizam testes e análises técnicas, o que abrange também os laboratórios que realizam investigação de paternidade.

Em associação aos STRs autossômicos, que muito contribuem para a identificação do indivíduo tanto quanto para a investigação de paternidade, os Y-STRs relacionados com o cromossomo Y, bem como os X-STRs do cromossomo X, são também ferramentas importantes na elucidação das análises de DNA forense ou casos de relação biológica. Os primeiros são importantes na determinação de linhagem patrilinear, mas não servem para incluir o indivíduo como o suposto pai da criança do sexo masculino, uma vez que o cromossomo Y é transmitido de pai para filho e deste para seu filho de maneira haplotípica, à semelhança da transmissão do DNA mitocondrial (herança materna), levando a que toda a descendência do sexo masculino, daquele primeiro indivíduo compartilhe os mesmos haplotipos Y-STRs; ou seja, todos os descendentes do sexo masculino apresentarão o mesmo padrão haplotípico do seu ancestral. Por essa razão, não se pode associar (ou multiplicar) os valores obtidos a partir das frequências dos alelos do perfil dos STRs autossômicos, denominados de índice de verossimilhança, com aquele da frequência do haplotipo dos Y-STRs. Entretanto, da análise do perfil Y-STR é possível excluir um indivíduo como pai biológico de uma criança do sexo masculino ou como doador de uma mancha de sangue ou sêmen.

Em relação ao cromossomo X (X-STRs), este é transmitido pelo pai para todas as filhas do sexo feminino, que recebem o outro cromossomo X da mãe biológica. Como neste caso o padrão de transmissão é muito semelhante aos STRs autossômicos, em geral os índices obtidos através de cálculos estatísticos utilizando as respectivas frequências dos alelos de ambos, STRs e X-STRs, podem ser multiplicados, aumentando, assim, as chances de favorecimento de paternidade, quando se trata de filha questionada na ausência do suposto pai, por exemplo (Figura 20.4).

A equipe da autora deste capítulo elaborou um banco de dados muito interessante e bastante completo para estudo de vários parâmetros de análise do cromossomo X*, os quais foram obtidos a partir das publicações nacionais sobre esse tema e de dados do próprio laboratório. Esforços estão sendo

* Ver <http://bgbx.com.br>. Número de registro de programa de *software*: 12166-1.

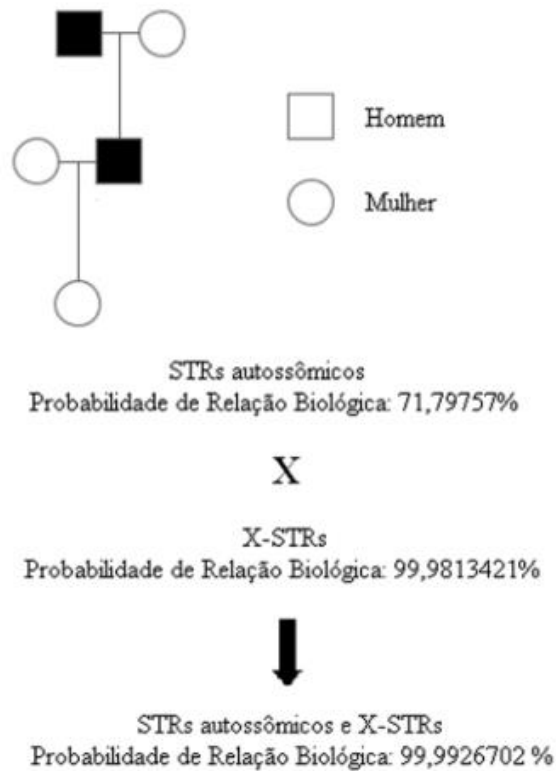


Figura 20.4 Heredograma de um caso de investigação de paternidade em que foi utilizada a análise de X-STRs. Os indivíduos em preto não foram analisados. Foi utilizada neste caso a análise da mãe biológica do suposto pai, uma vez que este estava ausente. O esquema mostra a probabilidade de relação biológica com os STRs autossômicos e X-STRs; a multiplicação dos índices aumenta a probabilidade de relação biológica entre os indivíduos analisados.

investidos no sentido de disponibilizar também as fórmulas para realizar os cálculos estatísticos mais importantes a serem incluídos nos laudos periciais em que há a necessidade da associação dos STRs autossômicos e X-STRs, garantindo, desse modo, um índice de paternidade ou de verossimilhança mais robusto.

A utilização das análises dos marcadores de Y e X-STRs é importante em casos de confirmação de exclusão de paternidade e, especialmente, em casos

de reconstrução, quando o suposto pai está ausente. Em casos criminais, estas também são ferramentas importantes para elucidar a origem de manchas de sangue e/ou sêmen.

A técnica de amplificação de STRs (autossômicos, Y e X) baseia-se na realização de uma reação da PCR multiplex, isto é, na qual vários *primers* são adicionados a um mesmo tubo de 0,2 mL contendo um mix da PCR (tampão, cloreto de magnésio, dNTPs e enzima Taq polimerase) e acrescentando-se à mistura o DNA em estudo.

Duas reações para controle das condições de amplificação são necessárias e imprescindíveis: um controle positivo, contendo DNA humano conhecido e convenientemente diluído (concentração de uso 0,5 ng/uL), e outro negativo, geralmente realizado com água destilada esterilizada. Os tubos são levados ao termociclador para a amplificação em ciclagem adequada e compatível com a reação multiplex.

Atualmente, o uso de kits comerciais são os mais recomendáveis para as análises forenses e de paternidade por serem validados nacional e internacionalmente. A seguir, a Figura 20.5 apresenta um esquema dos procedimentos básicos para a reação.

O sucesso da análise dos STRs está diretamente relacionado com a qualidade do DNA obtido. O DNA está presente em todas as células nucleadas e, portanto, pode ser extraído de diferentes materiais biológicos deixados na cena de um crime. Em geral, as amostras testadas nos laboratórios forenses são tipicamente sangue e manchas de sêmen, mas podem existir muitas outras fontes de DNA.

Para a comparação dos perfis de DNA, sempre há a necessidade de amostras de referência (sangue ou *swab* oral). Todas as amostras devem ser cuidadosamente manuseadas para a extração de DNA, de modo a evitar qualquer tipo de contaminação. Por essa razão, muitos laboratórios processam as amostras de evidência (aquelas obtidas na cena do crime) em momentos diferentes ou mesmo em locais diferentes das amostras de referência. Existem diversos métodos de extração de DNA; entretanto, os mais utilizados atualmente são orgânico, Chelex e papel FTA. Todos eles permitem a obtenção de DNA de boa qualidade para análise de STRs. A Figura 20.6 ilustra um perfil alélico amplificado a partir de uma amostra de DNA humano.

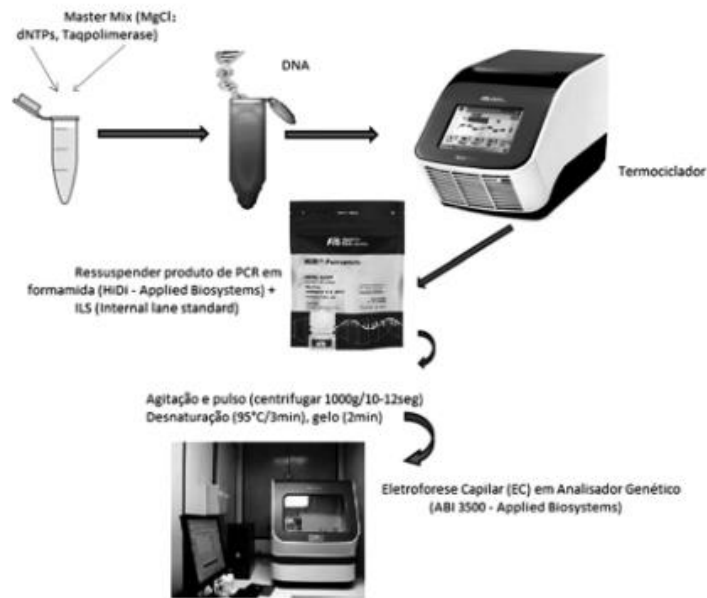


Figura 20.5 Esquema ilustrativo da reação de PCR multiplex para amplificação de STRs (autossômicos, Y ou X-STRs).

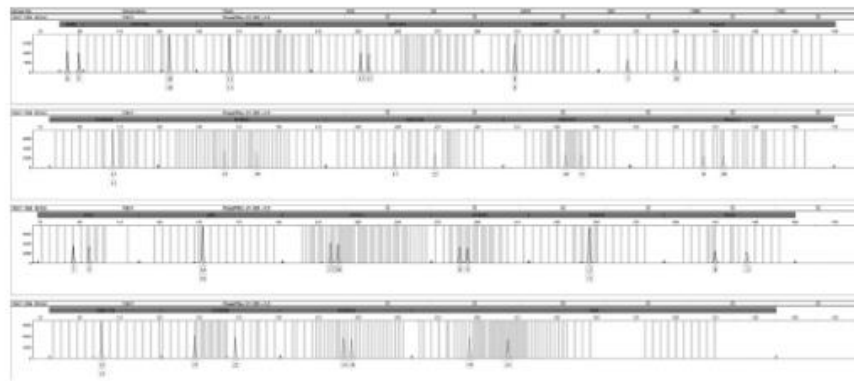


Figura 20.6 Eletróferograma obtido no analisador genético (ABI 3500 - Applied Biosystems) e analisado no GeneMapper ID-X apresentando os alelos amplificados na reação multiplex de 21 regiões do DNA humano (STRs autossômicos). A amostra pertence a um indivíduo do sexo masculino (X/Y). Os picos representam os alelos obtidos em cada marcador, que são repetições de pequenas sequências de bases, como mostrado na Figura 20.1.

20.3 POLIMORFISMOS DE INSERÇÃO/DELEÇÃO (INDELS) E SUAS APLICAÇÕES

O estudo das variações genéticas, utilizando os polimorfismos do DNA, permite uma melhor compreensão da história e da diversidade das populações humanas, além de proporcionar um sistema para a identificação genética de indivíduos. A aplicação do DNA decorre do seu alto poder de discriminação gerado por seus polimorfismos, e além dos STRs já estudados anteriormente, destacam-se outros dois: *single nucleotide polymorphisms* (SNPs) e *insertion-deletion polymorphisms* (INDELS)¹⁵.

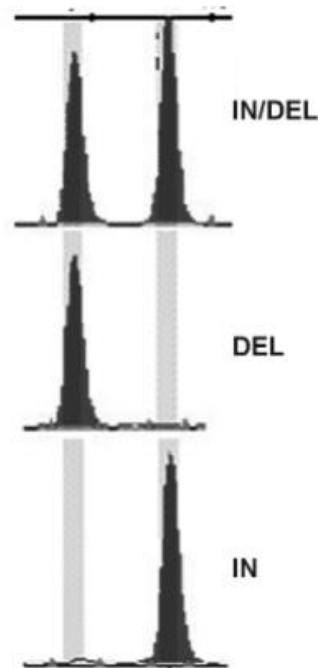
Weber e colaboradores (2002)¹⁶ foram pioneiros, pois identificaram e caracterizaram aproximadamente 2 mil INDELS no genoma humano, com diferentes tamanhos e frequências alélicas em europeus, africanos, japoneses e nativo-americanos, e devido à abundância e facilidade de análise, destacaram a sua utilidade para estudos genéticos. INDELS chegam a representar cerca de 16% de todos os polimorfismos do DNA humano e se apresentam muito distribuídos pelo genoma, em média, um INDEL a cada 7,2 Kb¹⁷.

INDELS são polimorfismos de comprimento, caracterizados pela inserção ou deleção de um ou mais nucleotídeos em uma determinada região do genoma; por exemplo, o polimorfismo rs16363 é identificado pela deleção da sequência "TGTTT" na localização cromossômica 22q13.1. O polimorfismo ao ser inserido no National Center for Biotechnology Information (NCBI) recebe um código único descrito com as letras rs (*reference sequence*), seguido da numeração específica para sua identificação correspondente à sua posição no genoma. Na localização cromossômica, o primeiro número refere-se ao cromossomo; no exemplo citado acima, o polimorfismo está localizado no cromossomo 22; a letra refere-se aos braços do cromossomo; a letra *p* indica o braço curto e *q* o braço longo; o número seguinte representa a posição no braço do cromossomo, no exemplo, banda 13, sub-banda 1.

A Figura 20.7 a seguir ilustra a identificação do INDEL "ATCG" no eletroferograma.

20.4 UTILIZAÇÃO DOS INDELS NA ANÁLISE DE DNA DEGRADADO

O DNA pode ser danificado ou destruído devido a condições ambientais adversas. A exposição ao ambiente degrada as moléculas de DNA, dividindo-o aleatoriamente em pequenos pedaços. Água, oxigênio, radiação



Deleção: TTAGATG ___ GTCGAA

Inserção: TTAGATGATCGGTCGAA

Figura 20.7 Eletroferograma identificando o INDEL (In/Del) com o alelo curto (deleção-Del) e o alelo longo (inserção-In-ATCG).

ultravioleta e enzimas nucleases são alguns dos agentes naturais que atuam nessa degradação¹⁸.

A habilidade de recuperar e analisar sequências de DNA de amostras como ossos e dentes, expostos ao longo do tempo a uma variedade de condições ambientais, tornou-se uma ferramenta valiosa para a identificação de indivíduos ou amostras desconhecidas¹⁹.

Pouco DNA endógeno, ação do meio ambiente, micro-organismos e presença de inibidores são os principais fatores que fazem com que a

recuperação e análise do DNA deste tipo de amostra sejam um desafio na área de identificação humana²⁰.

Ainda que a análise de STR (*short tandem repeats*) seja o principal polimorfismo utilizado como marcador genético em identificação humana¹⁹, conforme já foi mencionado anteriormente, apresenta dificuldades quando as amostras possuem DNA degradado e/ou em pouca quantidade, devido ao tamanho relativamente grande dos fragmentos a serem amplificados (150 pb a 500 pb)²¹.

Por essa razão, aumentou de modo considerável recentemente o interesse na utilização de INDELS como marcadores genéticos nas áreas de identificação humana e genética forense, devido à facilidade de análise, ampla distribuição no genoma, por apresentarem fragmento de pequeno tamanho, cerca de 100 pb e poderem ser detectados por diferentes técnicas. As frequências alélicas apresentam diferenças significativas entre populações geograficamente distintas, sendo potenciais marcadores para estudo de ancestralidade. Apresentam menor taxa de mutação em comparação com os STRs (Figura 20.8)²².



Figura 20.8 Esquema comparando STRs com INDELS e mostrando os respectivos tamanhos em pares de bases (pb) obtidos na amplificação.

O fato de o fragmento de DNA amplificado ser pequeno é muito interessante, facilitando a análise em amostras biológicas não mantidas em boas condições de preservação, por exemplo, ossos e dentes coletados em casos de exumação de cadáveres ou de acidentes em massa. Nessas situações, a alternativa é a combinação de marcadores STRs e INDELS para a obtenção de uma análise mais conclusiva e robusta.

Em testes de paternidade ou outra relação biológica, a maioria dos laboratórios utiliza um conjunto de dezesseis marcadores STRs que oferecem um poder discriminatório alto e geram um resultado de exclusão de paternidade (entre 0,9999983 e 0,9999998) ou fortemente indicativo de probabilidade

de inclusão (índices de paternidade entre 522.000 e 4.110.000). No entanto, alguns casos apresentam resultados ambíguos ou inconclusivos, em que a relação reivindicada não pode ser confirmada por uma probabilidade suficientemente alta, ou uma exclusão é sugerida por incompatibilidade alélica em apenas um ou dois destes marcadores, sendo que para uma análise resultar em exclusão de paternidade há a necessidade de não coincidência entre o alelo paterno obrigatório (filho/a) e o alelo presente no suposto pai em, no mínimo, três marcadores autossômicos²³. Na Figura 20.9, estão representados os eletroferogramas parciais de três regiões de STRs autossômicos, mostrando duas exclusões. É importante notar que incompatibilidades podem também ser identificadas por ação de mutações, uma vez que os STRs apresentam elevada taxa de mutação (cerca de 2×10^{-3}).

Resultados ambíguos surgem frequentemente, quando, por exemplo, o irmão biológico de primeiro grau do pai biológico é indicado como suposto pai de uma criança, de modo que a taxa de exclusão torna-se muito reduzida (poucas não coincidências nos alelos) e o índice de paternidade, usando uma razão de probabilidade de um homem aleatório na população, não se aplica; estes resultados são difíceis de serem interpretados. Um recurso encontrado é a análise de STRs adicionais para melhorar a probabilidade ou fornecer claras e inequívocas exclusões; no entanto, além daqueles que já são utilizados nos kits comerciais, poucos STRs são validados para que possam ser aplicáveis²⁴.

Estes resultados também são observados nos casos em que o suposto pai é falecido e há a necessidade de analisar amostras provenientes de exumação, como ossos e dentes, uma vez que esse tipo de amostra pode apresentar baixa quantidade de DNA e alto nível de degradação e contaminação, o que na grande maioria das vezes dificulta a análise de um número adequado de STRs. Nesses casos, analisando-se um número relativamente grande de INDELS (de 30 a 40), que apresentam menor taxa de mutação em comparação aos STRs, pode-se encontrar novas incompatibilidades, o que favorece a hipótese de exclusão, ou, ao contrário, caso não sejam encontradas incompatibilidades, aumentar o índice de paternidade a valores que suportem a confirmação de uma relação biológica.

20.5 TÉCNICAS PARA ANÁLISE DE INDELS AUTOSSÔMICOS

A técnica para análise de INDELS também é uma reação da PCR em multiplex, conforme mencionado anteriormente para amplificação de STRs.

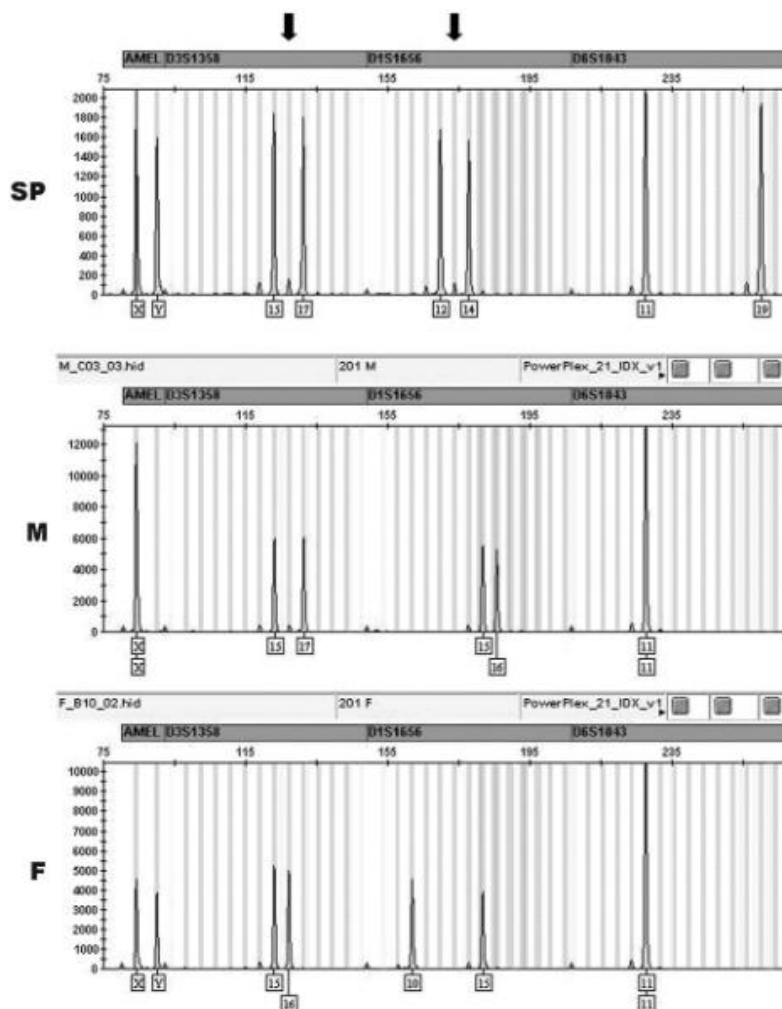


Figura 20.9 Comparação de três eletroferogramas parciais no software GeneMapper ID-X para verificar compatibilidade de alelos em um caso de investigação de paternidade. SP, suposto pai; M, mãe; F, filha. As setas indicam dois marcadores onde não houve compatibilidade, ou seja, o F não herdou o alelo obrigatório do SP.

Podem ser analisados trinta marcadores utilizando-se o kit *InvestigatorDIPplex* (QIAGEN), que também amplifica a amelogenina em uma única reação multiplex²⁵ (Figura 20.10).

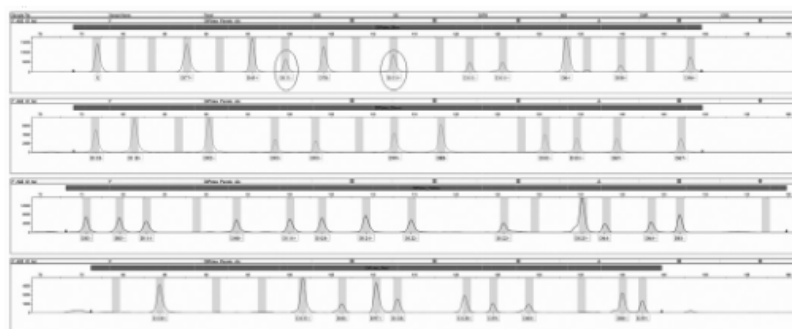


Figura 20.10 Eletroferograma da reação com o kit *InvestigatorDIPplex* (QIAGEN). (Trata-se de uma amostra feminina.) Os picos representam os alelos obtidos para cada marcador, sendo o alelo curto o que apresenta a deleção e o alelo longo o que apresenta a inserção, conforme exemplificado na Figura 20.3. Neste eletroferograma, os alelos são encontrados em regiões separadas, sendo divididos entre as quatro cores de fluoróforos. Estão identificados dois alelos de um mesmo marcador, separados no eletroferograma.

Devido ao fato de que os INDELS apresentam menor taxa de recombinação em comparação com os STRs, o número de INDELS analisados deverá ser maior para atingir um poder de discriminação aceitável; neste caso, os trinta polimorfismos são suficientes segundo o fabricante.

Outra reação foi padronizada por Pereira e colaboradores (2009)¹⁵ e permite analisar 38 INDELS autossômicos em multiplex (Figura 20.11). Segundo os autores, esses 38 polimorfismos são suficientes para atingir um poder de discriminação aceitável.

Comparando-se as duas reações, verifica-se que a reação de Pereira e colaboradores (2009)¹⁵ apresenta oito polimorfismos a mais do que aquela do kit *InvestigatorDIPplex* (QIAGEN), não havendo nenhum polimorfismo em comum. Em contrapartida, a reação de Pereira e colaboradores (2009)¹⁵ não amplifica a amelogenina.

Outra avaliação interessante nos eletroferogramas demonstra que, nos 38-INDELS, os marcadores podem apresentar os alelos na mesma região, enquanto no *InvestigatorDIPplex* (QIAGEN), os alelos dos marcadores se

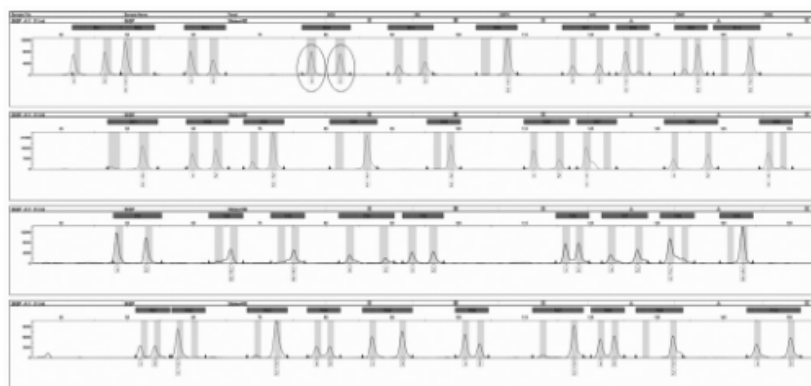


Figura 20.11 Eletroferograma da reação padronizada por Pereira e colaboradores (2009)¹⁵. (Amostra diferente da apresentada na Figura 20.10.) Os picos representam os alelos obtidos para cada marcador, sendo o alelo curto o que apresenta a deleção e o alelo longo o que apresenta a inserção, conforme exemplificado na Figura 20.3. Ao contrário da Figura 20.10, neste eletroferograma os alelos podem ser identificados na mesma região.

intercalam na mesma cor de fluoróforo. Isto dificulta a análise e a comparação entre os alelos de um mesmo marcador. A análise de amostras é realizada pelo *software* GeneMapper.

Como apresentado na Figura 20.12, os mesmos passos são realizados para análise de INDELS por ambas as reações multiplex, sendo que o passo de purificação pode ser opcional conforme a necessidade (quando o eletroferograma não discrimina adequadamente os alelos). De maneira geral, a metodologia para análise de INDELS é prática e rápida e muito similar à metodologia para análise de STRs, o que agrega praticidade ao trabalho em laboratórios de identificação humana.

20.6 APLICAÇÕES DO DNA MITOCONDRIAL

As mitocôndrias são organelas citoplasmáticas que apresentam em sua composição um genoma extracromossômico separado e distinto do genoma nuclear, o DNA mitocondrial (DNAm_t), composto por 16.569 nucleotídeos e apresentando-se como uma dupla fita circular. Sua composição química

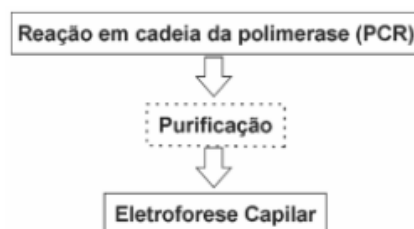


Figura 20.12 Fluxograma para análise de INDELS por ambas as reações. (O passo de purificação é opcional.)

em nada difere da composição do DNA nuclear; entretanto, ele possui um código genético próprio.

O DNAm_t evoluiu pelo acúmulo de mutações nas linhagens maternas e acredita-se fixar novas mutações dez vezes mais rápido do que o DNA nuclear²⁶. Este fato ocorre devido à baixa fidelidade da DNA polimerase mitocondrial, juntamente com a falta de mecanismos de reparo durante sua replicação, o que acaba sendo de grande interesse no campo forense.

Em alguns casos pode não ser possível a análise do DNA nuclear de algumas amostras (casos em que o DNA apresente degradações ou quando o material biológico não apresentar DNA nuclear), e então a melhor alternativa é a análise do DNA mitocondrial. Assim, o DNAm_t é utilizado na rotina forense para analisar ossos, dentes, fios de cabelos e outros tipos de amostras biológicas, sempre que se necessite identificar o indivíduo.

A principal vantagem do DNAm_t é que ele está presente em uma média de quinhentas a 2 mil cópias por célula. Esta abundância aumenta significativamente as chances de que algumas cópias de DNAm_t permanecerão estáveis em ambientes que apresentam condições não favoráveis à sua conservação, além do que a natureza circular do DNAm_t o torna menos suscetível à degradação por exonucleases²⁷.

O genoma mitocondrial possui herança estritamente materna, permitindo investigar relações de parentesco nesta linhagem, sendo, portanto, haploide e, por isso, não é submetido a processos de recombinação. Por essas duas razões, o DNAm_t não pode ser utilizado na investigação de paternidade. Vale ressaltar que tanto os filhos como as filhas compartilham o DNAm_t da mãe biológica.

Este genoma possui natureza monoclonal, ou seja, todo o DNAm_t de um indivíduo apresenta a mesma sequência, o que simplifica a interpretação

dos resultados durante o sequenciamento do DNAm. Porém, em alguns casos pode haver indivíduos que apresentam uma condição denominada de heteroplasmia, na qual uma mesma pessoa apresenta mais de um tipo de DNAm, podendo ser heteroplasmia de sequência, quando há diferença em apenas um nucleotídeo, ou heteroplasmia de comprimento, quando há deleção ou inserção de uma citosina em regiões com repetições mononucleotídicas (conhecidas como poli-C). Nos dois casos, as análises devem ser muito cuidadosas para a interpretação da heteroplasmia. Portanto, há necessidade de treinamento do perito para análise do DNAm em casos forenses.

20.7 UTILIZAÇÃO DO DNA MITOCONDRIAL EM ANÁLISES FORENSES

O genoma mitocondrial é composto de duas regiões (Figura 20.13): a região codificadora e a região controle, esta também conhecida como região não codificadora, região hipervariável ou *D-loop*, que controla a replicação e transcrição do DNAm, a qual, devido ao acúmulo de mutações pontuais, apresenta uma maior taxa de mutação quando comparada à região codificadora.

A região controle é formada por aproximadamente 1.200 pares de base e, por ser altamente polimórfica, é muito utilizada para o propósito da genética forense. Esta região do DNAm inclui três sub-regiões denominadas HV1 (formada por 342 pb), HV2 (268 pb) e HV3 (137 pb), sendo esta última região um segmento que pode apresentar um tamanho variável de pares de base, pelo fato de conter repetições dinucleotídicas “CA”, cujo número pode variar de um indivíduo para outro²⁸. Entre essas regiões existem outras duas denominadas inter-regiões, a primeira entre HV1 e HV2 e a segunda entre HV2 e HV3. A maior parte da variação de sequência entre indivíduos é encontrada nos segmentos HV1 e HV2. Por isso, estes são os mais utilizados na rotina forense para análise do DNAm. Entretanto, a região HV3 também pode auxiliar as análises²⁹.

A identificação baseia-se na análise dos polimorfismos presentes nesses segmentos (SNPs, inserções e deleções/*indels*), que são comparados a uma sequência de referência (*revised Cambridge Reference Sequence – rCRS*), sequenciada em 1981 por Anderson e colaboradores³⁰, reanalisada e revisada por Andrews e colaboradores em 1999³¹. No entanto, tais regiões oferecem um limitado poder de discriminação em um contexto forense³². Assim, a região HV3, junto com as inter-regiões, vem sendo explorada nos últimos

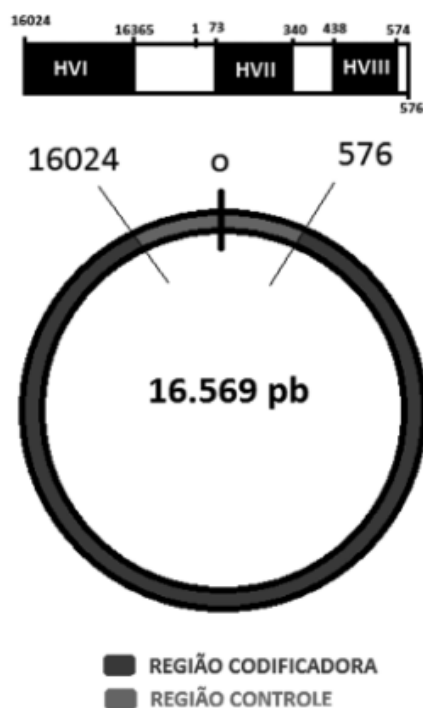


Figura 20.13 Esquema do DNA mitocondrial: região controle, com três sub-regiões (HV1, HV2 e HV3) e região codificadora.

anos, pois apesar de apresentarem baixos números de polimorfismos, informações complementares são obtidas quando adicionadas às regiões HV1 e HV2, aumentando o poder discriminativo, o que é importante na solução de casos forenses e, principalmente, em estudos de frequência populacional^{33, 34, 35}. Essas características fazem com que o sequenciamento completo da região não codificadora do DNAm_t seja mais adequado tanto para testes de identidade forense como para estudos evolucionários.

O sequenciamento da região controle do DNAm_t pode ser feito a partir de DNA obtido de diversos materiais biológicos, como ossos, dentes, tecidos, sangue, sêmen e fios de cabelo. A partir do DNA extraído, realiza-se uma PCR utilizando os *primers* (*forward* e *reverse*) que flanqueiam as regiões de

interesse, podendo ser toda a região não codificadora ou cada uma das suas sub-regiões.

Os produtos da PCR são, então, purificados para retirada do excesso de *primers* e nucleotídeos que não foram incorporados durante a PCR, e submetidos à reação de sequenciamento. Os produtos obtidos a partir do sequenciamento também deverão ser purificados previamente à eletroforese capilar realizada em analisador genético (ABI3500 – Applied Biosystems). A análise dos resultados é realizada com a ajuda de *softwares* de bioinformática, e as sequências obtidas são comparadas com a rCRS.

Outra técnica de análise do DNAm é a técnica de SNaPshot. Essa reação baseia-se na análise de polimorfismos de base única (SNPs), que representam a maior classe de polimorfismos humanos. Essa técnica está em uso crescente devido à detecção rápida de vários polimorfismos em um único ensaio. A capacidade multiplex é particularmente importante, especialmente no que diz respeito à análise de DNA forense, uma vez que pode reduzir o consumo de amostras, enquanto aumenta o rendimento do processamento das amostras e análise de dados. Esta técnica é realizada a partir da utilização de um kit comercial (ABI PRISM SNaPshot multiplex kit – Applied Biosystems) produzido pela empresa Life Technologies. A análise de SNPs do DNAm, tanto da região não codificadora como da região codificadora, permite a classificação de amostras em haplogrupos.

20.8 APLICAÇÕES DO SEQUENCIAMENTO DO DNAMT

Alguns fatos importantes comprovam como a análise do DNAm é de grande importância no contexto forense: foi a primeira ferramenta utilizada na identificação de ossadas de soldados americanos que lutaram na Guerra do Vietnã³⁶ e na identificação das vítimas da tragédia de 11 de setembro de 2001 no World Trade Center (EUA).

Outro exemplo interessante da utilização do DNA mitocondrial foi realizado na década de 1990 em um estudo de parentesco, no qual cientistas americanos foram convidados a examinar ossos e dentes de oito esqueletos de uma vala na Sibéria, onde se acreditava estar enterrada a família real russa Romanov, que fora executada pelos bolcheviques. Neste caso, os resultados do sequenciamento do DNAm elucidou que havia vínculo genético entre as ossadas³⁷.

A utilização do DNAm estendeu-se para estudos da evolução humana, migrações e formação da história de diversas populações, não permanecendo

apenas na rotina forense. O emprego do DNAm_t como ferramenta para investigações desta natureza surgiu com os estudos pioneiros do grupo liderado por Allan Wilson nos anos 1980³⁸.

20.9 BANCO DE DADOS

A análise do DNAm_t é rotineiramente utilizada em diversos laboratórios como ferramenta em análises forenses e estudos genéticos evolutivos. No entanto, é preciso um conhecimento aprofundado das várias características do DNAm_t a fim de analisar e interpretar corretamente os resultados. A forma correta de interpretar um resultado de DNAm_t ainda não foi padronizada, existindo apenas recomendações, como, por exemplo, aquelas adotadas pelo Federal Bureau of Investigation (FBI)³⁹ ou publicadas pelo Grupo de Habla Española y Portuguesa – International Society for Forensic Genetics (GHEP-ISFG)*.

A maioria dos laboratórios que utilizam tipagem do DNAm_t baseiam-se nos polimorfismos (SNPs, deleções e inserções) presentes na sequência de nucleotídeos nas regiões HV1 e HV2, comparando a amostra questionada com a sequência referência (rCRS) para a anotação das diferenças (polimorfismos)²⁹.

De acordo com o FBI, se essas sequências diferirem em dois ou mais polimorfismos, uma exclusão inequívoca pode ser feita. Se as sequências corresponderem uma à outra, existe a possibilidade da amostra questionada pertencer ao indivíduo analisado ou demonstrar vínculo materno, podendo, no entanto, pertencer a outro indivíduo não aparentado; por isso, torna-se necessário verificar a frequência com que um conjunto de polimorfismo ocorre em uma população (frequência do haplótipo), gerando a necessidade da criação de um banco de dados populacional de DNAm_t.

O termo “haplogrupo” refere-se ao conjunto de haplótipos (conjunto de alelos), que são designados por letras do alfabeto e derivados por descendência de uma mesma molécula ancestral, as quais apresentam um padrão basal de mutações. Os haplogrupos são definidos por um conjunto de haplótipos que são constituídos de nucleotídeos particulares presentes nas regiões controle e codificadora do DNAm_t. A relação entre haplótipos e haplogrupos do DNAm_t humano nos fornece informações mais detalhadas sobre a estrutura da variação genética. No site PhyloTree** encontra-se a árvore filogené-

* Ver <http://www.gep-isfg.org>.

** Ver www.phylotree.org.

tica mitocondrial humana que está representada de modo simplificado na Figura 20.14.

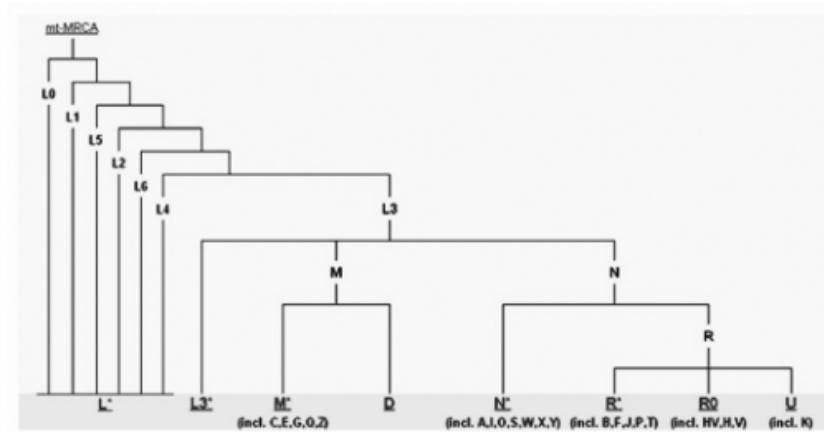


Figura 20.14 Árvore filogenética humana simplificada do DNA mitocondrial, mostrando os principais haplogrupos.

A criação de banco de dados de DNAm_t para estimar a frequência dos haplótipos em determinada população é indispensável para permitir sua utilização. Entretanto, pode estar sujeito a vários tipos de erros, principalmente humanos (na leitura e digitação de dados, por exemplo), havendo a necessidade de se estabelecerem regras de controle de qualidade tanto na geração quanto na compilação de resultados³⁹. Atualmente, existem ferramentas que podem ser utilizadas para que esses erros humanos sejam evitados como, por exemplo, MITOMAP (*mtDNA control region sequence polymorphisms*)* e IAN LOGAN**, as quais permitem comparar se todos os polimorfismos encontrados já foram descritos na literatura.

Diversos pesquisadores geram milhares de dados de genomas mitocondriais completos de seres humanos em todos os continentes, bem como uma nomenclatura detalhada dos tipos de DNAm_t ou haplogrupos em todo o mundo e uma enorme árvore filogenética dos haplogrupos observados, sub-haplogrupos e macro-haplogrupos⁴⁰. No campo da identificação humana,

* Ver <http://www.ianlogan.co.uk/mtDNA.htm>.

** Ver <http://www.ianlogan.co.uk/mtDNA.htm>.

as revistas científicas de maior impacto já exigem que os dados de trabalhos envolvendo análises de DNAmT ou outros tipos de análises gerado por pesquisadores sejam adicionadas aos bancos de dados para serem publicados.

Um exemplo de banco de dados que foi desenvolvido com rigorosos parâmetros de qualidade é o banco de dados mundial de DNAmT denominado EMPOP. Para manter um padrão de qualidade, é necessário ter uma padronização da geração, análise e transferência de dados, além do uso de *softwares* para a visualização de dados, interpretação e detecção de erros.

20.10 TÉCNICA PASSO A PASSO

Protocolo para a análise de 38 INDEls autossômicos

Extração de DNA

A extração do DNA do interior das células é o primeiro passo para sua análise. Descreveremos a seguir um protocolo⁴⁴ para a extração de DNA de sangue armazenado em papel de filtro FTA (Whatman), utilizando a resina Chelex 100 (Biorad) a 5%:

μ

- 1) Colocar em um tubo de 1,5 mL 2 discos (3 mm cada) de papel de filtro com sangue.
- 2) Adicionar 50 μL de H₂O, vortexar rapidamente e descartar a H₂O.
- 3) Adicionar 50 μL de H₂O e incubar a temperatura ambiente por 30 minutos.
- 4) Descartar a H₂O e adicionar 100 μL da solução homogeneizada da resina chelex 100 (Biorad) a 5%.
- 5) Incubar por 1 hora a 56 °C.
- 6) Vortexar rapidamente e incubar por 8 minutos a 100 °C.
- 7) Centrifugar a 14.000 RPM por 3 minutos.
- 8) Armazenar o sobrenadante.

Reação em cadeia da polimerase (PCR)

Segue a descrição do preparo da PCR para amplificação dos 38 INDELS padronizado por Pereira et al. (2009)¹⁵, utilizando o *Multiplex PCR kit* (QIAGEN):

H ₂ O	3 µL
Qiagen <i>Multiplex PCR master mix</i> (2×)	5 µL
<i>Primer mix</i> (10×)	1 µL
DNA 0,3 a 5 ng/µ L	1 µL
Volume total	10 µL

Em seguida, é utilizada a ciclagem em termociclador:

95 °C – 15 min.	
94 °C – 30 seg.	
60 °C – 90 seg.	10 ciclos
72 °C – 60 seg.	
94 °C – 30 seg.	
58 °C – 90 seg.	20 ciclos
72 °C – 60 seg.	
72 °C – 60 min.	
4 °C – até a retirada do termociclador.	

Observação: um passo adicional de purificação do produto da reação de PCR pode ser utilizado se necessário.

Análise do produto da reação de PCR

O produto da PCR deve ser preparado para eletroforese capilar pela adição de 1 µL do produto amplificado em 9,7 µL de formamida HI-DI (Applied Biosystems) e 0,3 µL de GeneScan 600 LIZ (Applied Biosystems). A separação e detecção são realizadas por eletroforese capilar.

20.11 CONCLUSÕES

As ferramentas para identificação humana utilizando o DNA são, atualmente, bastante eficazes na discriminação entre indivíduos e/ou vestígios e na investigação de paternidade, gerando índices de verossimilhança bastante elevados. Entretanto, cabe ressaltar que o exame de DNA na genética

forense constitui-se em uma evidência a mais na elucidação dos casos do Poder Judiciário, favorecendo ou não a hipótese de acusação ou de defesa. O perito deve analisar os resultados obtidos de maneira cuidadosa, isenta e ética, relatando as evidências científicas encontradas através das análises dos perfis genéticos de maneira clara e objetiva, valorando os dados com os cálculos estatísticos adequados, mas deixando ao juiz a tarefa de avaliação final da importância de tais evidências no julgamento do referido caso.

REFERÊNCIAS

1. França GV. Medicina Legal. 6. ed. Rio de Janeiro: Guanabara-Koogan; 2001. 579 p.
2. Jeffreys AJ, Wilson V, Thein SL. Hypervariable minisatellite regions in human DNA. *Nature*. 1985;314:67-72.
3. Jeffreys AJ, Wilson V, Thein SL. Individual specific "fingerprints" of human DNA. *Nature*. 1985;316:75-79.
4. Wyman AR, White R. A highly polymorphic locus in human DNA. *Proc Natl Acad Sci USA*. 1980;77:6754-8.
5. Jeffreys AJ. Genetic fingerprinting. *Nature Medicine*. 2005;11(10):XIV-XVIII.
6. Saiki RK, Scharf S, Faloona F, Mullis K, Horn G, Erlich HA, et al. Enzymatic amplification of β -globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science*. 1985;230:1350-4.
7. Saiki RK, Gelfand DH, Stoffel S, Scharf S, Higuchi RH, Horn GT, et al. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*. 1988;239:487-91.
8. Mullis K, Faloona F. Specific synthesis of DNA in vitro via a polymerase catalyzed chain reaction. *Method Enzymol*. 1987;155: 335-50.
9. Watson JD. DNA: o segredo da vida. São Paulo: Companhia das Letras; 2005. 470 p.
10. The FBI Federal Bureau of Investigation. Combined DNA Index System (Codis) [Internet]. Available from: <http://www.fbi.gov/about-us/lab/biometric-analysis/codis>.
11. Consultor Jurídico. Técnica usada na tipagem genética não está livre de erros [Internet]. 2006 Jun 13. Available from: http://www.conjur.com.br/2006-jun-13/tecnica_usada_tipagem_genetica_ao_livre_errores.
12. Paradela ER, Figueiredo AL dos S, Smarra ALS. A identificação humana por DNA: aplicações e limites [Internet] [Cited Nov 2013]. In: *Âmbito Jurídico*, Rio Grande, IX, n. 30, jun 2006. Available from: http://www.ambito-juridico.com.br/site/index.php?artigo_id=1175&cn_link=revista_artigos_leitura.
13. Houck MM. A realidade do CSI. *Scientific American Brasil*. Available from: http://www.uol.com.br/sciam/reportagens/a_realidade_do_csi.html.
14. Bassette F. País padroniza técnicas para perícia. *Folha de S. Paulo*. 27 set 2010:Cotidiano.
15. Pereira R, Phillips C, Alves C, Amorim A, Carracedo A, Gusmao L. A new multiplex for human identification using insertion/deletion polymorphisms. *Electrophoresis*. 2009;30:3682-90.
16. Weber JL, David D, Heil J, Fan Y, Zhao C, Marth G. Human diallelic insertion/deletion polymorphisms. *Am J Hum Genet*. 2002;71(4):854-862.

17. Mills RE, Luttig CT, Larkins CE, Beauchamp A, Tsui C, Pittard WS, Devine SE. An initial map of insertion and deletion (INDEL) variation in the human genome. *Genome Res.* 2006;16(9):1182-90.
18. Butler J. *Advanced Topics in Forensic DNA Typing: Methodology*. 1st ed. Waltham/San Diego/London:Academic Press; 2011. 704 p.
19. Hochmeister MN, Budowle B, Borer UV, Eggmann U, Comey CT, Dirnhofer R. Typing of deoxyribonucleic acid (DNA) extracted from compact bone from human remains. *J Forensic Sci.* 1991;36:1649-61.
20. Loreille OM, Diegoli TM, Irwin JA, Coble MD, Parsons TJ. High efficiency DNA extraction from bone by total demineralization. *Forensic Sc International.* 2007;1:191-5.
21. Zidkova A, Horinek A, Kebrdlova V, Korabecna M. Application of the new insertion-deletion polymorphism kit for forensic identification and parentage testing on the Czech population. *Int J Legal Méd.* 2011;127(1):7-10.
22. Phillips C, Fondevila M, Garcia-Margarinos M, Rodriguez A, Salas A, Carracedo A, Lareu MV. Resolving relationship tests that show ambiguous STR results using autosomal SNPs as supplementary markers. *Forensic Science International Genetics.* 2008;2:198-204.
23. Mullaney JM, Mills RE, Pittard WS, Devine SE. Small insertions and deletions (INDELs) in human genomes. *Hum Mol Genet.* 2010;19(2):131-6.
24. Yang N, Li H, Criswell LA, Gregersen PK, Alarcon-Riquelme ME, Kittles R, et al. Examination of ancestry and ethnic affiliation using highly informative diallelic DNA markers: application to diverse and admixed populations and implications for clinical epidemiology and forensic medicine. *Hum Genet.* 2005;118(3-4):382-92.
25. Larue BL, Ge J, King JL, Budowle B. A validation study of the Qiagen Investigator DIPplex® kit; an INDEL-based assay for human identification. *Int J Legal Med.* 2012;126(4):533-40.
26. Wallace DC, Ye J, Neckelmann SN, et al. Sequence Analysis of cDNAs for the Human and Bovine ATP Synthase Beta Subunit: Mitochondrial Genes Sustain Seventeen Times more Mutations. *Current Genetics.* 1987;12:81-90.
27. Kashyap VK, Sitalaximil T, Chattopadhyay P, Trivedi R. DNA Profiling Technologies in Forensic Analysis. *International Journal of Human Genetics.* 2004;3: 11-30.
28. Lutz S, Wittig H, Weisser HJ, Heizmann J, Jungle A, Simonin ND, et al. Is it possible to differentiate mtDNA by means of HVIII in samples that cannot be distinguished by sequencing the HVI and HVII regions? *Forensic Science International.* 2000;113:97-101.
29. Paneto GG, et al. Heteroplasmy in Hair: Study of Mitochondrial DNA Third Hypervariable Region in Hair and Blood Samples. *J Forensic Sci.* 2010;55:715-7.
30. Anderson S, Bankier AT, Barrell BG, et al. Sequence and organization of the human mitochondrial genome. *Nature.* 1981;290:457-65.

31. Andrews R, et al. Reanalysis and revision of Cambridge reference sequence for human mitochondrial DNA. *Nat Genet.* 1999;23:147.
32. Huang D, Gi C, Yi S, et al. Typing of 24 mtDNA SNPs in Chinese Population Using SNaPshot Minisequencing. *Med Sci.* 2010;30:291-98.
33. Bini C, et al. Different informativeness of the three hypervariable mitochondrial DNA regions in the population of Bologna (Italy). *Forensic Sci Int.* 2003;135:48-52.
34. Vanecek T, Vorel F, Sip M. Mitochondrial DNA D-loop hypervariable regions: Czech population data. *Int J Legal Med.* 2004;118:14-8.
35. Zhang YJ, et al. Haplotype diversity in mitochondrial DNA hypervariable region I, II and III in northeast China Han. *Forensic Sci Int.* 2005;149:267-9.
36. Holland MM, Fisher DL, Mitchell LG. Mitochondrial DNA sequence analysis of human skeletal remains: identification of remains from the Vietnam war. *J Forensic Sci.* 1993;38:542-53.
37. Farah SB. *DNA segredos e mistérios.* 2 ed. São Paulo: Sarvier; 2007.
38. Wilson AC, et al. Mitochondrial DNA and two perspectives on evolutionary genetics. *Biol J Linn.* 1985;6:375-400.
39. Scientific Working Group on DNA Analysis Methods (Swgdam). Guidelines for Mitochondrial DNA (mtDNA) Nucleotide Sequence Interpretation. *Forensic Science Communications.* 2003. 5 p.
40. Bandelt HJ, Macaulay V, Richards M. *Human mitochondrial DNA and the evolution of Homo sapiens.* 1st ed. Heidelberg: Springer; 2006. p.153.
41. Singer-Sam J, Tanguay R, Riggs AD. Use of chelex to improve the PCR signal from a small number of cells. *Amplifications: A Forum for PCR Users.* 1989;3:11.