



República Federativa do Brasil
Ministério do Desenvolvimento, Indústria
e do Comércio Exterior
Instituto Nacional da Propriedade Industrial

(21) PI 1104979-0 A2



* B R P I 1 1 0 4 9 7 9 A 2 *

(22) Data de Depósito: 04/11/2011
(43) Data da Publicação: 17/09/2013
(RPI 2228)

(51) Int.Cl.:
C12Q 1/68
C12N 15/10

(54) Título: MÉTODO E KIT PARA IDENTIFICAÇÃO GENÉTICA HUMANA POR MEIO DE POLIMORFISMOS DO DNA MITOCONDRIAL PARA APLICAÇÃO EM POPULAÇÕES MISCIGENADAS

(73) Titular(es): UNIVERSIDADE ESTADUAL PAULISTA "JÚLIO DE MESQUITA FILHO"

(72) Inventor(es): GREICIANE GABURRO PANETO, HEIDI PFEIFFER, REGINA MARIA BARRETTO CICARELLI, STEPHAN KÖHNEMANN

(57) Resumo: MÉTODO E KIT PARA IDENTIFICAÇÃO GENÉTICA HUMANA POR MEIO DE POLIMORFISMOS DO DNA MITOCONDRIAL PARA APLICAÇÃO EM POPULAÇÕES MISCIGENADAS. A presente invenção refere-se a um método e kit para identificação genética humana por meio da análise de polimorfismos específicos do DNA mitocondrial para aplicação em populações miscigenadas como a brasileira, por exemplo. A técnica desenvolvida, além de permitir a identificação do indivíduo, permite também classificá-lo em halogrupos do DNA mitocondrial possibilitando a identificação da origem ancestral materna do indivíduo testado. A referida invenção pode ser aplicada na área de genética forense, pela polícia científica ou por laboratórios particulares, tendo como principais beneficiados populações miscigenadas que não dispõem de técnicas específicas para sua identificação e classificação genética.

**MÉTODO E KIT PARA IDENTIFICAÇÃO GENÉTICA HUMANA POR MEIO DE
POLIMORFISMOS DO DNA MITOCONDRIAL PARA APLICAÇÃO EM
POPULAÇÕES MISCIGENADAS**

CAMPO DE APLICAÇÃO

5 A presente invenção refere-se a um método e kit para
identificação genética humana por meio da análise de
polimorfismos específicos do DNA mitocondrial para
aplicação em populações miscigenadas como a brasileira, por
exemplo. A técnica desenvolvida, além de permitir a
10 identificação do indivíduo, permite também classificá-lo em
haplogrupos do DNA mitocondrial possibilitando a
verificação da origem ancestral materna do indivíduo
testado. A referida invenção pode ser aplicada na área de
genética forense, pela polícia científica ou por
15 laboratórios particulares, tendo como principais
beneficiados populações miscigenadas que não dispõem de
técnicas específicas para sua identificação e classificação
genética.

ESTADO DA TÉCNICA

20 A identificação humana por meio do DNA constitui um
dos produtos mais revolucionários da Genética Moderna. Em
menos de 20 anos, tornou-se uma ferramenta indispensável na
área forense, sendo aceita rotineiramente em processos
judiciais em todo o mundo. Sua importância não está apenas
25 na possibilidade de provar a culpabilidade de uma pessoa ou
inocentá-la, mas sim em estabelecer uma conexão irrefutável
entre a pessoa e sua presença na cena do crime. O perfil
genético de um indivíduo, comumente utilizado na
identificação humana, está baseado na combinação de
30 diversos marcadores que são herdados de seus progenitores.

Esses marcadores referem-se, geralmente, a diferenças nas sequências de DNA entre os indivíduos (polimorfismos). Em alguns casos, entretanto, a análise do DNA nuclear não pode ser aplicada. Isso ocorre quando o DNA da amostra
5 apresenta-se muito degradado, ou os indivíduos em questão apresentam vínculos genéticos distantes, ou ainda em situações onde o material biológico não apresenta o DNA nuclear, como em fragmentos de cabelo, por exemplo, (SALAS; LAREU; CARRACEDO, 2001). Nestes casos, a única alternativa,
10 ou pelo menos a alternativa de maior sucesso, é a análise do DNA mitocondrial.

Entretanto, a falta de kits e metodologias para análise de DNA em identificação humana aplicáveis na população brasileira foi o motivo principal que auxiliou o
15 desenvolvimento desta nova tecnologia. A utilização do DNA mitocondrial na identificação humana já está bem documentada e já foi utilizada na identificação de soldados mortos na Guerra do Vietnam e na identificação de vítimas da tragédia de 11 de setembro de 2001 no *World Trade*
20 *Center*, EUA. Entretanto, a tecnologia do DNA mitocondrial pelo sequenciamento da região hipervariável utilizada nos casos acima não se aplica em todas as situações e por isso uma nova tecnologia denominada minisequenciamento do DNA ou SNaPshot, também utilizando o DNA mitocondrial, foi
25 desenvolvida.

A Patente US 5,925,520 intitulada "Multiplex sequencing method using primers of different lengths to detect polymorphisms" descreve a metodologia de SNaPshot utilizada no desenvolvimento da presente invenção. A
30 tecnologia de SNaPshot permite analisar diversos SNPs

(polimorfismos de um único nucleotídeo) distribuídos pelo genoma em uma única reação em "multiplex". Entretanto, a escolha dos SNPs é específica para cada população.

O pedido de Patente US 2009/0233292 A1 intitulado
5 "Method and kit for identification of genetic polymorphisms" descreve uma metodologia e kit semelhantes ao da invenção, com objetivo principal de classificar amostras em haplogrupos "globais". As principais diferenças entre este pedido de patente e a presente invenção estão na
10 escolha dos haplogrupos que serão avaliados, bem como os SNPs escolhidos, e conseqüentemente os primers utilizados. O pedido de patente acima utiliza 16 SNPs enquanto que a metodologia ora descrita engloba 42 SNPs, sendo por isso muito mais ampla e detalhada.

15 Diversos kits também já foram relatados na literatura, especialmente para populações européias, mas nenhum dos kits pode ser utilizado plenamente na população brasileira devido à grande miscigenação da mesma, isto é, presença de europeus, africanos e nativo-americanos,
20 principalmente. A presente tecnologia preenche esta lacuna uma vez que permite a utilização da técnica na população brasileira e em outras populações miscigenadas uma vez que não existem kits para genotipagem de SNPs do DNA mitocondrial que permitam a classificação em haplogrupos em
25 nossa população. Além disso, a metodologia foi modificada para permitir a geração de resultados mesmo com pequenas quantidades de DNA para que o mesmo possa ser utilizado em amostras degradadas e de difícil análise. Este kit poderá ser utilizado tanto na identificação humana quanto no
30 estudo de linhagens matrilineas de populações brasileiras.

Uma das possíveis desvantagens da presente tecnologia é a dependência de outra tecnologia já patenteada (vide patente "SNaPshot" citada acima) para sua execução. Além disso, atualmente, o custo para sua execução ainda é relativamente alto, o que tende a diminuir com o difundir da tecnologia.

A presente tecnologia apresenta a possibilidade de genotipagem de um grande número de SNPs em uma única reação permitindo a classificação do indivíduo em haplogrupos do DNA mitocondrial e/ou identificação humana, aplicado ao contexto brasileiro de população miscigenada.

A mesma mostra-se útil, principalmente no cenário brasileiro, para ser utilizada na identificação humana nos casos onde o DNA encontra-se degradado ou em pequenas concentrações, ou ainda na presença de fios de cabelo sem raiz. Assim, poderá ser aplicado pela polícia científica ou laboratórios particulares na identificação de indivíduos em desastres em massa, como queda de aviões, ou como metodologia de *screening*, ou ainda como metodologia de determinação da origem materna do indivíduo.

Esta tecnologia poderá ser aplicada na confecção de um kit voltado para a identificação humana e/ou classificação em haplogrupos. Este kit poderá ser produzido por empresas especializadas na área e utilizado tanto em laboratórios públicos (polícia científica) como em laboratórios particulares.

BREVE DESCRIÇÃO DOS DESENHOS

A figura 1 é um esquema da localização dos 42 polimorfismos do DNA mitocondrial utilizados na reação de SNaPshot para classificação em haplogrupos.

A figura 2 é um diagrama de blocos indicando todas as etapas realizadas na análise das amostras.

A figura 3 é um eletroferograma, mostrando o perfil de análise de uma amostra para os 42 SNPs.

5 A figura 4 é uma árvore filogenética englobando haplogrupos europeus, africanos e nativo-americanos gerada por este painel de 42 SNPs.

A figura 5 é um gráfico do tipo pizza da classificação das amostras segundo a origem dos haplogrupos.
10

DESCRIÇÃO DETALHADA DA INVENÇÃO

DEFINIÇÕES

O termo "DNA mitocondrial" refere-se ao DNA localizado dentro de organelas celulares denominadas mitocôndrias.
15

O termo "SNP" é uma sigla para *Single Nucleotide Polymorphism* (em português - Polimorfismo de um Único Nucleotídeo) e acontece quando um nucleotídeo do DNA é substituído por outro.

20 O termo "nucleotídeo" refere-se às bases nitrogenadas do DNA podendo ser uma das quatro possíveis: adenina, timina, citosina ou guanina.

O termo "polimorfismo" refere-se a variações que ocorrem, com certa frequência, na sequência do DNA.

25 O termo "PCR" é uma sigla para *Polymerase Chain Reaction* (em português - Reação em Cadeia da Polimerase) e refere-se a uma metodologia utilizada para amplificar exponencialmente um fragmento de DNA.

O termo "multiplex" significa que mais de uma reação química ocorre simultaneamente.
30

O termo "primer" refere-se às sequências sintéticas de DNA construídas para serem utilizadas em diferentes metodologias como PCR, SNaPshot, etc.

O termo "SNaPshot" refere-se a uma metodologia também conhecida como minisequenciamento na qual é possível identificar a mudança de um nucleotídeo do DNA para outro.

O termo "haplogrupo" refere-se a um grupo de haplótipos do DNA mitocondrial (conjunto de polimorfismos) que apresentam um ancestral comum.

O termo "amostra" refere-se ao material biológico selecionado para a análise.

O termo "forense" é um adjetivo relacionado ao que se utiliza no foro ou nos tribunais e denomina, na maioria das vezes, o uso da ciência e tecnologia para a reconstrução e obtenção de provas de crimes.

METODOLOGIA

Seleção dos SNPs (polimorfismos)

Para o desenvolvimento desta tecnologia quarenta polimorfismos (39 SNPs e 1 deleção), presentes na região codificadora do DNA mitocondrial, foram selecionados de forma a permitirem a classificação das amostras em haplogrupos europeus, africanos e nativo-americanos. Outros dois SNPs presentes na região não-codificadora (SNP 16519 e 16362) foram escolhidos para o estudo do aumento do poder discriminatório dos mesmos em nossa população. Isto foi feito utilizando uma reação SNaPshot em *multiplex*, composta apenas por SNPs que permitiam classificação em haplogrupos europeus (KÖHNEMANN et al., 2009), a qual foram adicionados SNPs para a classificação em haplogrupos africanos (L0, L1, L2, L3, L4) e nativo-americanos/asiáticos (A, B, C, D)

baseados nos haplogrupos já descritos na literatura para a população brasileira (ALVES-SILVA et al., 2000). Os SNPs escolhidos foram descritos por PhyloTree.org (VAN OVEN; KAYSER, 2009) conforme mostrado na figura 1.

5 **Reação de PCR em multiplex**

Os *primers* para a reação de PCR em *multiplex* foram desenhados utilizando o programa PerlPrimer (MARSHALL, 2004) e a possibilidade de interação entre os mesmos foi descartada utilizando o programa Autodimer (VALLONE; 10 BUTLER, 2004). Trinta e três pares de *primers* foram utilizados para amplificar 42 SNPs. O tamanho dos amplicons variou de 55 a 196 pares de base para permitir a utilização da metodologia para análise de amostras degradadas. A concentração final dos *primers* utilizada na reação e outras 15 informações estão dispostas na Tabela 1.

Tabela 1 - Informações sobre primers da reação de PCR em multiplex.

Posição do SNP	Primer	Seqüência do primer (5'-3')	Tamanho do amplicon (pb)	Concentração final (μ M)
G709A e G769A	mtG709A-FWD mtG709A-REV	AAGATTACACATGCAAGCATCC CTAGGCTAAGCGTTTGGAGCTG	112	0,178
G1719A; A1811G e A1736G	mtG1719 + 1811A-FWD mtG1719 + 1811A-REV	CTTGACCGCTCTGAGCTAAA GCAGAAGGTATAGGGTTAGTCC	196	0,178
T2416C	mtT2416C-FWD mtT2416C-R	CAGCCAAATATCTACAATCAACC GTAAGATTTGCCGAGTTCCT	114	0,178
A2706G e C2772T	mtA2706G + C2772T-FWD mtA2706G + C2772T-REV	ACTTTTAACCAGTGAATTTGACCT GCCCAACCGAAATTTTA	157	0,128
G3915A	mtG3915A-FWD mtG3915A-REV	TAGCAGAGACCAACCGAACC GGCGTATTCGATGTTGAAGC	80	0,134
G4580A	mtG4580A-FWD mtG4580A-REV	CGCTAAGCTCGCACTGATTT TTTTGGTTAGAACTGGAATAAAGC	130	0,125
C4883T	mtC4883T-FWD mtC4883T-R	CCTTTCACTTCTGAGTCCCA GCTTACGTYTAGTGAGGGAG	128	0,036
T5004C	mtT5004C-FWD mtT5004C-REV	TCCATCATAGCAGGCAGTTG TGGTTATGTTAGGGTTGTACGG	124	0,107

T5442C	mtT5442C-FWD mtT5442C-R	AATCTACTCCACCTCAATCAC GTAGGAGTAGCGTGGTRAGG	125	0,178
T6365C e T6392C	mtT6365C + T6392C-FWD mtT6365C + T6392C-REV	TGGAGCCTCCGTAGACCTAA TGGCAGGGGGTTTTATATTG	110	0,045
R-T6776C	mtT6776C-FWD mtT6776C-REV	TGGCTTCCTAGGGTTTTATCG CGGAGGTGAAATATGCTCGT	83	0,032
C7028T e A7055G	mtC7028T-FWD mtC7028T-REV	AGCAAATCATCACTAGACATCG GATGGCAAATACAGCTCCTATTGA	94	0,071
C7256T	mtC7256T-FWD mtC7256T-R	TCTTCCCACAACACITTTCTC GAGAAATGAATGAAGCCCTACAG	109	0,089
A8251G e 8281del19bp	mtG8251A-FWD mtDEL8281-9-R	CATGCCCATCGTCCCTAGAAAT TGCTAAGTTAGCTTTACAGTGG	113	0,089
T8404C	mtT8404C-FWD mtT8404C-REV	CAGTGAAATGCCCCAACTAAA TTTGGTGAGGGAGGTAGGTG	126	0,107
G9055A e A9066G	mtG9055A + A9066G-FWD mtG9055A + A9066G-REV	CGCCTAACCGCTAACATTACT TGATAAGTGTAGAGGGAGGTTAATG	100	0,045
C10400T	mtC10400T-FWD mtC10400T-REV	GCCCTAAGTCTGGCCTATGA TGAGTCGAAATCAATTCGTTTTG	90	0,178

T10873C	mtT10873C-FWD mtT10873C-REV	AATCAACACAACCCACCACA GGGGAACAGCTAAATAGGTTG	90	0,178
A11251G	mtA11251G-FWD mtA11251G-REV	CGCAGGCACATACTTCCTATTCC AGTGAGCCTAGGGTGTGGAG	90	0,032
A11812G	mtA11812G-FWD mtA11812G-REV	TCCTCTCAAGGACTTCAAACCT GGCTTGCTAGAAAGTCATCAAAAA	63	0,071
G12372A	mtG12372A-FWD mtG12372A-REV	AAAAATTTTGGTGCAACTCCA AACGAGGGTGGTAAGGATGG	107	0,089
C12705T	mtC12705T-FWD mtC12705T-REV	CCCAACATTAATCAGTTCTTCAA GAACAGTTGGAATAGGTTGTTAGC	91	0,107
C12858T	mtC12858T-FWD mtC12858T-REV	ATGATACGCCCGAGCAGA TGTGGTCTCATGAGTTGGA	126	0,045
A13263G	mtA13263G-FWD mtA13263G-R	CCCCTAAATCCAAACTCTAACAC GCCGATTGTAACCTATTATGAGTCC	145	0,045
G13708A	mtG13708A-FWD mtG13708A-REV	CGGTTCCCCACCCCTTACTA GGGGGAAATGTTGTTAG	119	0,196
T14470A	mtT14470A-FWD mtT14470A-REV	CCCCATGCCTCAGGATACTC TTTAGGGGAAATGATGGTTG	126	0,268

C14766T e T14798C	mtC14766T-FWD mtC14798C-REV	CAACTACAAGAACACCAATG ATCATGGGAGATGTTGGAT	82	0,125
A15607G	mtA15607G-FWD mtA15607G-REV	TCTCCGATCCGTCCTAAC GATAGTAATAGGGCAAGGAC	55	0,02
C15904T	mtC15904T-FWD mtC15904T-REV	TACTCAAAATGGGCCCTGTCCT TGATTTGTCCTTGGAAAAAGGT	86	0,089
T16362C	mtT16362C-FWD mtT16362C-R	TAGCACATTACAGTCAAATCCC GGATATTGATTCACGGAGGA	151	0,071
T16519C	mtT16519C-FWD mtT16519C-REV	CCATAACACTTGGGGGTAGC TTATTTAAGGGGAACGTGTGG	99	0,089

A reação de PCR *multiplex* foi realizada utilizando-se:

- 1,25 µl de MgCl₂ (50 mM);
- 1,25 µl de dNTPs (2,5 mM);
- 5 - 1,25 µl de *PCR Gold buffer* (10x) (Applied Biosystems by Life Technologies);
- 0,2 µl de soro albumina bovina (20 mg/ml);
- 0,2 µl de *AmpliTaq Gold* (Applied Biosystems by Life Technologies);
- 10 - 100 pg de DNA;
- os *primers* da tabela 1; e
- água ultrapura (Milli-Q) autoclavada qsp. 12,5 µl.

Os fragmentos foram amplificados em termociclador *GeneAmp PCR System 9700* (Perkin Elmer) conforme a ciclagem:

15 Pré-incubação a 95°C por 10 min, seguida de três ciclos de 30 seg a 95°C, 45 seg a 55°C e 30 seg a 72°C. Após esta etapa, 19 ciclos de 30 seg a 95°C, 45 seg a 55°C com aumento de temperatura em 0,2°C por ciclo, 30 seg a 72°C; seguida de 11 ciclos de 30 seg a 95°C, 45 seg a 60°C, 30

20 seg a 72°C; e finalmente 7 min a 72°C. Os produtos da reação de PCR foram purificados através de reação enzimática adicionando-se 2,5 µl de *Exo/SAP-IT* (USB) e incubando-se por 120 min a 37°C, seguida de desnaturação da enzima a 75°C por 15 min.

25 **Reação de SNaPshot**

Os *primers* para a reação de SNaPshot foram desenhados manualmente para se anelarem a um nucleotídeo antes da posição do SNP que se deseja genotipar. Uma cauda nucleotídica, descrita por Sanchez et al. (2005), com

30 diferentes números de bases foi adicionada à região 5` de

cada *primer*. Esta cauda, que não se anelava ao produto de PCR previamente amplificado, foi adicionada para permitir a separação dos fragmentos por tamanho em eletroforese capilar. A temperatura de anelamento de cada um dos *primers* 5 foi checada com o programa Perlprimer e a formação de dímeros e *hairpins* foi descartada a partir da análise dos mesmos pelo programa Autodimer. Os *primers* utilizados e outras informações podem ser visualizados na Tabela 2.

Tabela 2 - Informações sobre primers da reação de SNaPshot em multiplex.

Nome do primer	SNP	Sequência do primer	Orientação	Conc. final (μM)
R-G709A-SNP	G709A	TTTAGAGGGTGAACCTCACTGGAA	R	0,023
SP2-R-G769A	G769A	CGTCGTGAAAAGTCTGACAAAACTGACTAAACTAGGTAGGTGCCACGTCGTGAAAAGTCTGACAAAAGCGTTTGG AGCTGCATTG	R	0,06
G1719A-SNP	G1719A	AACTGACTAAACTAGGTGCCACGTCGTGAAAAGTCTGACAAAACGTCGTGAAAAGTCTGACAAACCCTTA CTACCAGACAACCTTA	F	0,08
SP-R-A1736G	A1736G	CAATTTCAATTTTCTATCGCCTATACTTAT	R	0,05
A1811G-SNP	A1811G	AACTGACTAAACTAGGTGCCACGTCGTGAAAAGTCTGACAAAGCAAGGAAAAGATGAAAAATTATA	F	0,1
SP3-T2416C	T2416C	TGCCACGTCGTGAAAAGTCTGACAAAACTGACTAAACTAGGTGCCACGTCGTGAAAAGTCTGACAAACAAC CAACAAGTCATTATTACCC	F	0,1
R-A2706G-SNP	A2706G	GCCACGTCGTGAAAAGTCTGACAAAGTCTTCTCGTCTTGCTGTGT	R	0,1
R-C2772T-SNP	C2772T	GTCTGACAAGGTTTGGTAGTTTAGGACCTGT	R	0,075
R-G3010A-SNP	G3010A	CGTGAAAAGTCTGACAAAACCTGACTAAACTAGGTGCCACGTCGTGAAAAGTCTGACAACTCGATGTTGGA TCAGGACATCCC	F	0,024
R-G3915A-SNP	G3915A	GTATTTCGATGTTGAAGCCTGAGACTAGTTCGGACTC	R	0,06
G4580A-SNP	G4580A	ACAAAACTGACTAAACTAGGTGCCACGTCGTGAAAAGTCTGACAAATTACCTGAGTAGGCCCTAGAAATAA ACAT	F	0,025
SP2-C4883T	C4883T	GTGAAAAGTCTGACAAAACATGACAAAAACTAGCCCC	F	0,03
T5004C-SNP	T5004C	CGTGAAAAGTCTGACAAACCAACCCCA GCTACGCCAAAATC	F	0,035
SP2-T5442C	T5442C	TAAACTAGGTGCCACGTCGTGAAAAGTCTGACAAAACAYACAAAACCCACCCCA	F	0,034
T6365C-SNP	T6365C	ACGTCGTGAAAAGTCTGACAAAATCTTCTCCTACACCTAGCAGG	F	0,1

R-T14470A-SNP	T14470A	TCTGACAAAACTGACTAAACTAGGTGCCACGTCGTGAAAAGTCTGACAAAAATTTATTTAGGGGGAATGA TGGTTGTCTTTGG	R	0,15
C14766T-SNP	C14766T	GAAAAGTCTGACAAAAAAGTACTAAACTAGGTGCCACGTCGTGAAAAGTCTGACAAAAAACAATGACCCCAAT ACGCAAAA	F	0,04
T14798C-SNP	T14798C	CGTAAAAGTCTGACAAAAAAGTACTAAACTAGGTGCCACGTCGTGAAAAGTCTGACAAAAAACAATGACCCCAAT TTAAATTAACCACTCA	F	0,28
A15607G-SNP	A15607	AGTCTGACAACTCCGATCCGTCCTTACAA	F	0,018
R-C15904T-SNP	C15904T	ACGTCGTGAAAAGTCTGACAAACCGGTTTACAAGACTGGTGTATTA	R	0,007
SP2-T16362C	T16362C	CCACGTCGTGAAAAGTCTGACAAAAAAGTACTAAACTAGGTGCCACGTCGTGAAAAGTCTGACAAAAATTACA GTCAAAATCCCTTCTYG	F	0,07
R-T16519C-SNP	T16519C	CGTAAAAGTCTGACAAATGGGGCTATTTAGGCTTTATG	R	0,03

A reação de SNaPshot foi realizada utilizando-se:

- 2,5 µl *SNaPshot Mastermix*;
- 0,5 µl *PCR Gold buffer*;
- 0,5 µl do *mix* dos *primers* da tabela 2;
- 5 - 1,0 µl do produto de PCR purificado; e
- água ultrapura (Milli-Q) autoclavada qsp. 10 µl.

A reação de minisequenciamento foi realizada em termociclador PTC-100 (MJ Research) utilizando-se a ciclagem abaixo: 25 ciclos de 10 seg a 96°C, 5 seg a 55°C,
10 e 30 seg a 60°C, e finalmente 4°C por um tempo infinito.

A purificação dos produtos da reação foi realizada por tratamento enzimático adicionando-se 1,25 µl de SAP (fosfatase alcalina de camarão) e incubação a 37°C por 75 min, seguida de desnaturação da enzima por 15 min a 75°C.

15 Os fragmentos foram separados por eletroforese em *ABI 3100-Avant Genetic Analyzer*, conforme recomendações do fabricante. O resultado foi analisado utilizando o programa GeneScan (v 3.7, Applied Biosystems by Life Technologies) e Genotyper (v 2.5, Applied Biosystems by Life Technologies).

20 Sensibilidade

A validação da sensibilidade da técnica foi feita utilizando uma série de diluições de um DNA padrão (DNA 9947A - Promega): 10, 1, 0,1, 0,01 e 0,001 ng. Foram obtidos perfis completos com a análise de no mínimo 0,01
25 ng, o que demonstra a grande sensibilidade da tecnologia.

Classificação em haplogrupos

Para os resultados do SNaPshot, a classificação em haplogrupos foi baseada em Phylotree.org.

Amostras

30 Para a validação da técnica, foram analisadas 160

amostras de sangue (depositadas em papel de filtro comum) de indivíduos da Grande São Paulo. Para a extração do DNA das amostras de sangue, dois discos de 3 mm de papel de filtro foram cortados utilizando *Harris Micro Punche e Cutting Mat* e o DNA foi extraído utilizando a resina *Chelex* (SINGER-SAM et al., 1989). A quantificação das amostras foi realizada com o kit *Quantifier Human DNA Quantification* (Applied Biosystems by Life Technologies) por PCR em tempo real.

10 **Resultados e discussão**

Foram analisados 42 SNPs do DNA mitocondrial em uma única reação de PCR e SNaPshot a fim de permitir a classificação em haplogrupos e a identificação dos indivíduos dentro da população em estudo.

15 O fator mais importante observado antes da construção deste painel foi a seleção criteriosa de *primers* que não apresentem interações entre si e que sejam capazes de amplificar fragmentos menores que 200 pb, para a reação de PCR e SNaPshot para SNPs que permitam a classificação de diferentes haplogrupos para a nossa população. Até o momento, poucas reações em *multiplex* com um número tão grande de SNPs haviam sido relatadas e, até o momento, nenhum *multiplex* com 42 SNPs em uma única reação de PCR e SNaPshot foram publicados.

25 **Aplicação na classificação dos indivíduos em haplogrupos**

Quarenta polimorfismos (39 SNPs e 1 deleção) da região codificadora do DNA mitocondrial e dois SNPs da região não-codificadora (também conhecida como região hipervariável - SNPs 16519 e 16362) foram estrategicamente

30

escolhidos baseado na sua habilidade de classificar os haplogrupos presentes em nossa população e a fim de estudar o aumento do poder discriminatório do painel. A figura 3 mostra o resultado obtido da análise da amostra de um indivíduo de nossa população, onde nos quadrados abaixo dos picos os números seguidos de letras indicam a posição do SNP e a variação do nucleotídeo, números sozinhos indicam tamanho do fragmento após corrida eletroforética. Já o eixo-y indica a intensidade de fluorescência RFU (*relative fluorescence units*).

A figura 4 mostra a árvore filogenética que foi utilizada para a classificação dos indivíduos baseada nos polimorfismos presentes em cada um. Os SNPs escolhidos foram baseados nos haplogrupos esperados para nossa população (encontrados na literatura) e os SNPs para sua análise foram baseados no site Phylotree.org. Para a classificação de uma amostra no respectivo haplogrupo, iniciamos a análise pelos SNPs 10400 e 10873 (posição Start na região central/superior da figura 4) e seguimos baseados nos alelos de cada SNP naquela amostra até o haplogrupo correspondente. Por exemplo, uma amostra classificada como haplogrupo A percorre o caminho desde a posição Start até a posição do haplogrupo A, conforme mostrado na figura 4. Para se enquadrar neste haplogrupo, tal amostra deve possuir, obrigatoriamente, os polimorfismos 10400C, 10873T e 1736G.

Já uma amostra classificada como haplogrupo W percorre o caminho desde a posição Start até a posição do haplogrupo W e para isso deve possuir, obrigatoriamente, os polimorfismos 10400C, 10873T, 709A e 8251A.

Como a classificação em haplogrupos baseada apenas na análise da região hipervariável é, em alguns casos, deficiente e pode não expressar a correta classificação em haplogrupos, a classificação obtida a partir da análise da região hipervariável em nossa amostragem (realização com a técnica de sequenciamento) foi comparada com a classificação gerada por SNaPshot. Algumas amostras que não puderam ser corretamente classificadas pelo sequenciamento, obtiveram sucesso pelo SNaPshot (principalmente na correta definição dos haplogrupos HV, H e V). Em algumas amostras, entretanto, o painel completo (42 SNPs) não foi obtido, o que, entretanto, não inviabilizou sua classificação em haplogrupo. Isso aconteceu mais comumente em amostras classificadas em haplogrupos africanos e pode ser explicada por mutações nos sítios de ligação dos *primers*. Um exemplo claro ocorreu com amostras que apresentavam mutação no nucleotídeo 16360, o que impedia o anelamento do *primer* responsável pela genotipagem do SNP 16362 na reação de SNaPshot, causando uma falha na detecção deste SNP. Este problema foi solucionado utilizando *primers* degenerados (mistura de *primers* com nucleotídeo C e T na posição 16360).

A população estudada foi, então, classificada em 46,6% de origem africana, 27,3% europeia, 25,5% nativo-americana e 0,6% de origem asiática conforme mostrado na figura 5. Os haplogrupos mais comumente encontrados foram L3 (19,4%), L2 (14,4%) e H (13,7%).

Dentre as amostras de origem africana, o haplogrupo mais comumente encontrado foi L3 (41,3%), seguido por L2 (30,7%), L1 (12,0%), L0 (8,0%), L4 (2,7%) e L* (4,0%).

Resultados similares foram descritos por outros autores (GONÇALVES et al., 2008; HÜNEMEIER et al., 2007).

Dentre as amostras de origem européia, o haplogrupo mais comumente encontrado foi H (50,0%), seguido por J (11,3%), K, U e T (9,1% cada), X (4,5%), W, HV e V (2,3% cada). A frequência encontrada para o haplogrupo predominante H foi similar àquela observada em países da Europa ocidental (PEREIRA et al., 2000; ROOSTALU et al., 2007; TURCHI et al., 2008).

Os haplogrupos de origem nativo-americanos mais comumente encontrados foram: Haplogrupos B e C (33,3% cada), seguido por A (26,2%) e D (4,8%). Duas amostras de origem asiática foram classificadas em haplogrupo M. Entretanto, uma das amostras foi corretamente classificada como haplogrupo M1a1 africano (2,4% da população africana) somente quando o sequenciamento da região hipervariável foi analisado (OLIVIERI et al., 2006). A outra amostra classificada como haplogrupo M parece representar a população asiática que recentemente imigrou para o Brasil.

Aplicação na identificação genética dos indivíduos

A metodologia descrita também pode ser aplicada na identificação genética dos indivíduos. Isso ocorre porque além de existir um grande número de haplogrupos envolvidos, o que já separa os indivíduos em grupos, alguns SNPs ainda variam dentro do mesmo haplogrupo. Com isso os grupos de indivíduos com mesmo haplogrupo podem ser subdivididos e, grande parte dos mesmos, apresenta um conjunto de polimorfismos que aparecem unicamente nele, permitindo sua identificação. Esta capacidade de identificar o indivíduo e distingui-lo de outros é medida pela diversidade

haplotípica. A diversidade haplotípica encontrada foi de 96,04% utilizando o painel com 42 SNPs em nossa amostra populacional. Isto significa que a probabilidade de selecionarmos, ao acaso, um indivíduo de nossa população e seu perfil genético não coincidir com o perfil de outro indivíduo de nossa população é de 96,04%. Os SNPs 16519 e 16362 foram adicionados à metodologia unicamente para aumentarem o poder discriminatório, visto que não são utilizados para classificação em haplogrupos e são altamente variáveis entre os indivíduos e haplogrupos.

O SNP 16519 foi altamente informativo (presença do alelo mais raro em mais de 30% das amostras) nos haplogrupos A, D, L0, L2, L3 e J. Nos haplogrupos europeus H e U, o alelo mais raro estava presente em aproximadamente 25% das amostras. Uma pequena variação foi encontrada nos haplogrupos C e L1.

O SNP 16362 não foi tão informativo quanto o SNP 16519. Entretanto, variações foram encontradas nos haplogrupos U, H, L3, L2 e C. Estes SNPs podem ser importantes para o uso forense, por exemplo, a fim de escolher o melhor SNP a ser tipado para discriminar dois indivíduos com o mesmo haplogrupo.

No que se refere à sensibilidade do painel, um bom resultado foi obtido com 10 picogramas de DNA. Esse resultado, aliado à amplificação de tamanhos pequenos gerados na PCR, permitem sugerir o uso deste painel em casos forenses, especialmente para amostras contendo pequena quantidade ou com degradação do DNA.

KIT

O kit é composto por um conjunto de *primers*

específicos para a reação de PCR (Tabela 1) e um conjunto de *primers* específicos para a reação de minisequenciamento (Tabela 2) e deve ser utilizado na metodologia descrita neste relatório.

5 **REFERÊNCIAS BIBLIOGRÁFICAS**

1) ALVES-SILVA, J.; DA SILVA SANTOS, M.; GUIMARÃES, P.E. et al. The ancestry of Brazilian mtDNA lineages. **Am. J. Hum. Genet.**, v. 67, p. 444-61, 2000.

2) GONCALVES, V.F.; CARVALHO, C.M.; BORTOLINI, M.C.
10 et al. The phylogeography of African Brazilians. **Hum. Hered.**, v. 65, p. 23-32, 2008.

3) HUNEMEIER, T.; CARVALHO, C.; MARRERO, A.R. et al. Niger-Congo speaking populations and the formation of the Brazilian gene pool: mtDNA and Y-chromosome data. **Am. J. Phys. Anthropol.**, v.133, p. 854-67, 2007.
15

4) KÖHNEMANN, S.; HOHOFF, C.; PFEIFFER, H. An economical mtDNA SNP assay detecting different mitochondrial haplogroups in identical HVR 1 samples of Caucasian ancestry. **Mitochondrion**, v. 9, p. 370-5, 2009.

20 5) MARSHALL O.J. PerlPrimer: cross-platform, graphical primer design for standard, bisulphite and real-time PCR. **Bioinformatics**, v. 20, p. 2471-2472, 2004.

6) OLIVIERI, A.; ACHILLI, A. PALA, M. et al. The mtDNA legacy of the Levantine early Upper Palaeolithic in
25 Africa. **Science**, v. 314, p. 1767-70, 2006.

7) PEREIRA, L.; PRATA, M.J.; AMORIM, A. Diversity of mtDNA lineages in Portugal: not a genetic edge of European variation. **Ann. Hum. Genet.**, v. 64, p. 491-506, 2000.

8) ROOSTALU, U.; KUTUEV, I.; LOOGVALI, E.L. et al.
30 Origin and expansion of haplogroup H, the dominant human

mitochondrial DNA lineage in West Eurasia: the Near Eastern and Caucasian perspective. **Mol. Biol. Evol.**, v. 24, p. 436-48, 2007.

5 9) SALAS, A.; LAREU, M. V.; CARRACEDO, A.
Heteroplasmy in mtDNA and the weight of evidence in forensic mtDNA analysis: a case report. **Int. J. Legal Med.**, v. 114, p. 186-190, 2001.

10 10) SANCHEZ, J.J.; BORSTING, C.; MORLING, N. Typing of Y chromosome SNPs with multiplex PCR methods. **Methods Mol. Biol.**, v. 297, p. 209-228, 2005.

11) 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.

15 12) TURCHI, C.; BUSCEMI, L.; PREVIDERE, C. et al. Italian mitochondrial DNA database: results of a collaborative exercise and proficiency testing. **Int. J. Legal Med.**, v. 122, p. 199-204, 2008.

20 13) VALLONE, P.M.; BUTLER, J.M. AutoDimer: a screening tool for primer-dimer and hairpin structures. **Biotechniques**, v. 37, p. 226-231, 2004.

14) VAN OVEN, M.; KAYSER, M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. **Hum. Mutat.**, v. 30, p. e386-394, 2009.

REIVINDICAÇÕES

1. Método para identificação genética humana por meio de polimorfismos do DNA mitocondrial para aplicação em populações miscigenadas **caracterizado** por compreender as etapas de:

(a) extração do DNA da amostra;

(b) seleção criteriosa de primers que não apresentem interações entre si e que sejam capazes de amplificar fragmentos menores que 200 pb, para a reação de PCR e SNaPshot para SNPs que permitam a classificação de diferentes haplogrupos para a nossa população;

(c) amplificação dos fragmentos de DNA pela técnica de PCR em multiplex;

(d) purificação dos produtos obtidos na etapa (c) através de reação enzimática;

(e) realização do minisequenciamento dos fragmentos de DNA pela técnica do SNaPshot em multiplex;

(f) purificação dos produtos obtidos na etapa (e) por tratamento enzimático;

(g) separação dos fragmentos da etapa (f) por eletroforese;

(h) classificação do haplogrupo.

2. Método, de acordo com a reivindicação 1, **caracterizado** pelo fato dos primers utilizados na técnica de PCR em multiplex serem (mtG709A-FWD/mtG709A-REV); (mtG1719 + 1811A-FWD/mtG1719 + 1811A-REV); (mtT2416C-FWD/mtT2416C-R); (mtA2706G + C2772T-FWD/mtA2706G + C2772T-REV); (mtG3010A-FWD/mtG3010A-REV); (mtG3915A-FWD/mtG3915A-REV); (mtG4580A-FWD/mtG4580A-REV); (mtC4883T-FWD/mtC4883T-R); (mtT5004C-FWD/mtT5004C-REV); (mtT5442C-FWD/mtT5442C-R);

(mtT6365C + T6392C-FWD/mtT6365C + T6392C-REV); (mtT6776C-FWD/mtT6776C-REV); (mtC7028T-FWD/mtC7028T-REV); (mtC7256T-FWD/mtC7256T-R); (mtG8251A-FWD/mtDEL8281-9-R); (mtT8404C-FWD/mtT8404C-REV); (mtG8697A-FWD/mtG8697A-REV); (mtG9055A +
 5 A9066G-FWD/mtG9055A + A9066G-REV); (mtC10400T-FWD/mtC10400T-REV); (mtT10873C-FWD/mtT10873C-REV); (mtA11251G-FWD/mtA11251G-REV); (mtA11812G-FWD/mtA11812G-REV); (mtG12372A-FWD/mtG12372A-REV); (mtC12705T-FWD/mtC12705T-REV); (mtC12858T-FWD/mtC12858T-REV);
 10 (mtA13263G-FWD/mtA13263G-R); (mtG13708A-FWD/mtG13708A-REV); (mtT14470A-FWD/mtT14470A-REV); (mtC14766T-FWD/mtC14798C-REV); (mtA15607G-FWD/mtA15607G-REV); (mtC15904T-FWD/mtC15904T-REV); (mtT16362C-FWD/mtT16362C-R); (mtT16519C-FWD/mtT16519C-REV).

15 3. Método, de acordo com a reivindicação 1, **caracterizado** pelo fato dos primers utilizados na técnica SNaPshot em multiplex serem R-G709A-SNP; SP2-R-G769A; G1719A-SNP; SP-R-A1736G; A1811G-SNP; SP3-T2416C; R-A2706G-SNP; R-C2772T-SNP; R-G3010A-SNP; R-G3915A-SNP; G4580A-SNP;
 20 SP2-C4883T; T5004C-SNP; SP2-T5442C; T6365C-SNP; T6392C-SNP; T6776C-SNP; C7028T-SNP; SP2-R-A7055G; SP2-C7256T; SP-A8251G new; SP2-DEL8281-9; T8404C-SNP; R-G8697A-SNP; G9055A-SNP; R-A9066G-SNP; C10400T-SNP; T10873C-SNP; A11251G-SNP; A11812G-SNP; G12372A-SNP; R-C12705T-SNP; C12858T-SNP; SP-
 25 A13263G; G13708A-SNP; R-T14470A-SNP; C14766T-SNP; T14798C-SNP; A15607G-SNP; R-C15904T-SNP; SP2-T16362C; R-T16519C-SNP.

30 4. Método, de acordo com a reivindicação 1, 2 e 3, **caracterizado** pelo fato de analisar 42 SNPs do DNA mitocondrial em uma única reação de PCR e SNaPshot.

5. Kit para executar o método conforme descrito na reivindicação 1 **caracterizado** por compreender um conjunto de 33 primers específicos para a reação de PCR em multiplex e 42 primers específicos para a reação de SNaPshot em multiplex.

6. Kit, de acordo com a reivindicação 5, **caracterizado** pelo fato dos 33 primers específicos para a reação de PCR em multiplex serem (mtG709A-FWD/mtG709A-REV); (mtG1719 + 1811A-FWD/mtG1719 + 1811A-REV); (mtT2416C-FWD/mtT2416C-R); (mtA2706G + C2772T-FWD/mtA2706G + C2772T-REV); (mtG3010A-FWD/mtG3010A-REV); (mtG3915A-FWD/mtG3915A-REV); (mtG4580A-FWD/mtG4580A-REV); (mtC4883T-FWD/mtC4883T-R); (mtT5004C-FWD/mtT5004C-REV); (mtT5442C-FWD/mtT5442C-R); (mtT6365C + T6392C-FWD/mtT6365C + T6392C-REV); (mtT6776C-FWD/mtT6776C-REV); (mtC7028T-FWD/mtC7028T-REV); (mtC7256T-FWD/mtC7256T-R); (mtG8251A-FWD/mtDEL8281-9-R); (mtT8404C-FWD/mtT8404C-REV); (mtG8697A-FWD/mtG8697A-REV); (mtG9055A + A9066G-FWD/mtG9055A + A9066G-REV); (mtC10400T-FWD/mtC10400T-REV); (mtT10873C-FWD/mtT10873C-REV); (mtA11251G-FWD/mtA11251G-REV); (mtA11812G-FWD/mtA11812G-REV); (mtG12372A-FWD/mtG12372A-REV); (mtC12705T-FWD/mtC12705T-REV); (mtC12858T-FWD/mtC12858T-REV); (mtA13263G-FWD/mtA13263G-R); (mtG13708A-FWD/mtG13708A-REV); (mtT14470A-FWD/mtT14470A-REV); (mtC14766T-FWD/mtC14798C-REV); (mtA15607G-FWD/mtA15607G-REV); (mtC15904T-FWD/mtC15904T-REV); (mtT16362C-FWD/mtT16362C-R); (mtT16519C-FWD/mtT16519C-REV).

7. Kit, de acordo com a reivindicação 5, **caracterizado** pelo fato dos 42 primers específicos para a reação de SNaPshot em multiplex serem R-G709A-SNP; SP2-R-

G769A; G1719A-SNP; SP-R-A1736G; A1811G-SNP; SP3-T2416C; R-
A2706G-SNP; R-C2772T-SNP; R-G3010A-SNP; R-G3915A-SNP;
G4580A-SNP; SP2-C4883T; T5004C-SNP; SP2-T5442C; T6365C-SNP;
T6392C-SNP; T6776C-SNP; C7028T-SNP; SP2-R-A7055G; SP2-
5 C7256T; SP-A8251G new; SP2-DEL8281-9; T8404C-SNP; R-G8697A-
SNP; G9055A-SNP; R-A9066G-SNP; C10400T-SNP; T10873C-SNP;
A11251G-SNP; A11812G-SNP; G12372A-SNP; R-C12705T-SNP;
C12858T-SNP; SP-A13263G; G13708A-SNP; R-T14470A-SNP;
C14766T-SNP; T14798C-SNP; A15607G-SNP; R-C15904T-SNP; SP2-
10 T16362C; R-T16519C-SNP.

FIGURA 1

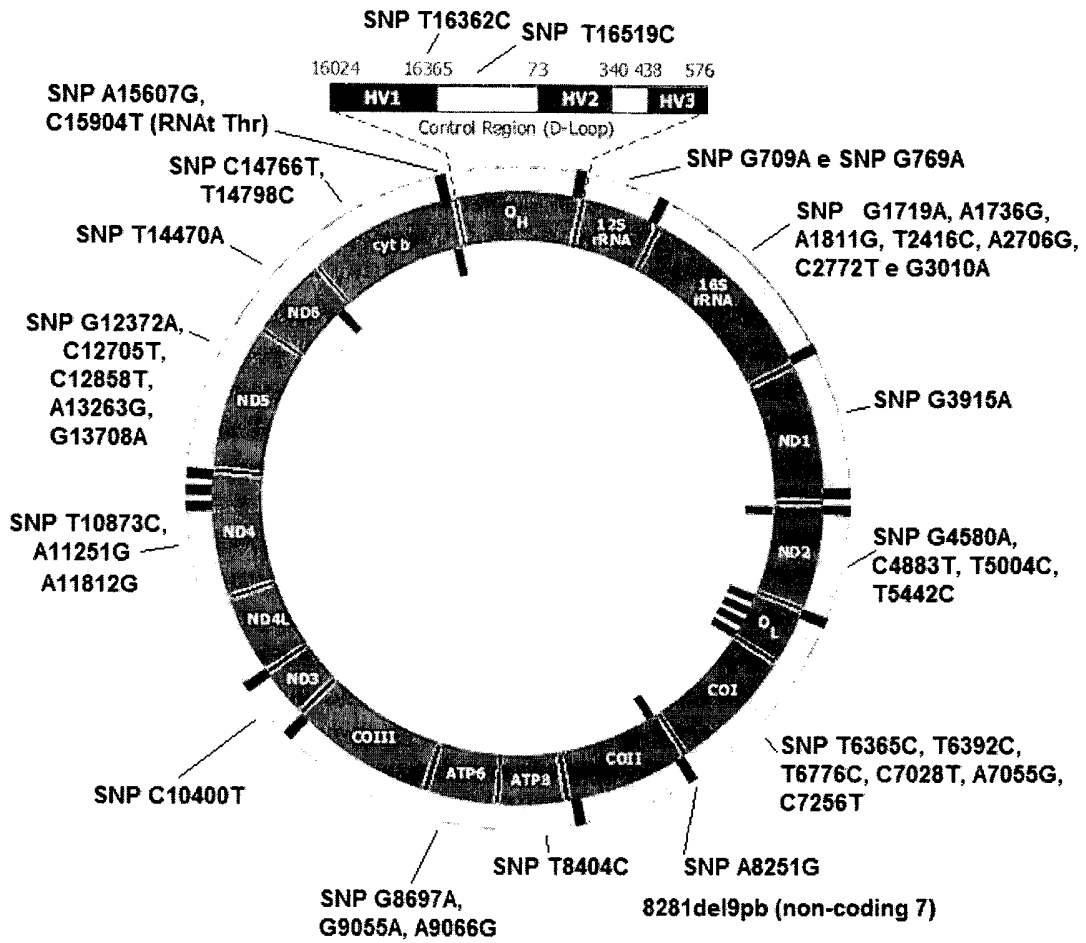


FIGURA 2

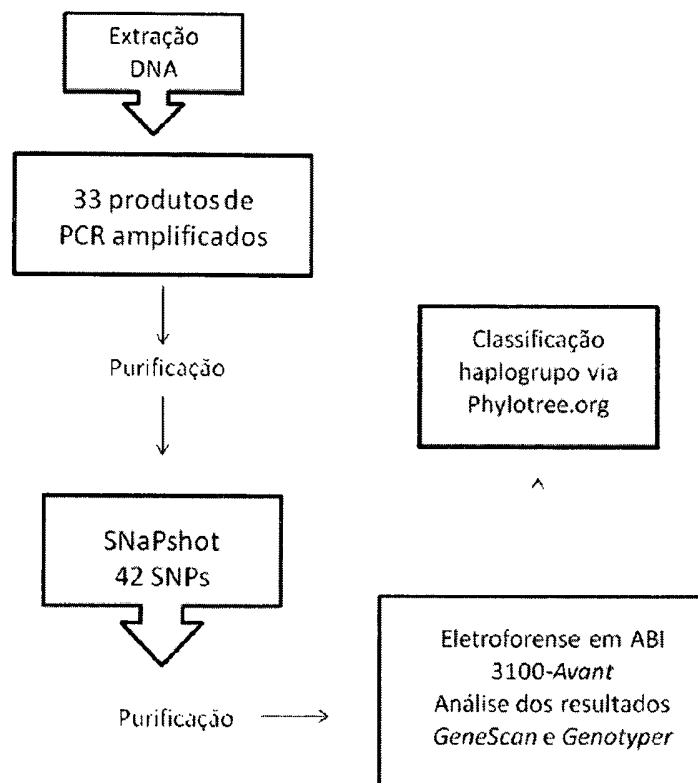


FIGURA 3

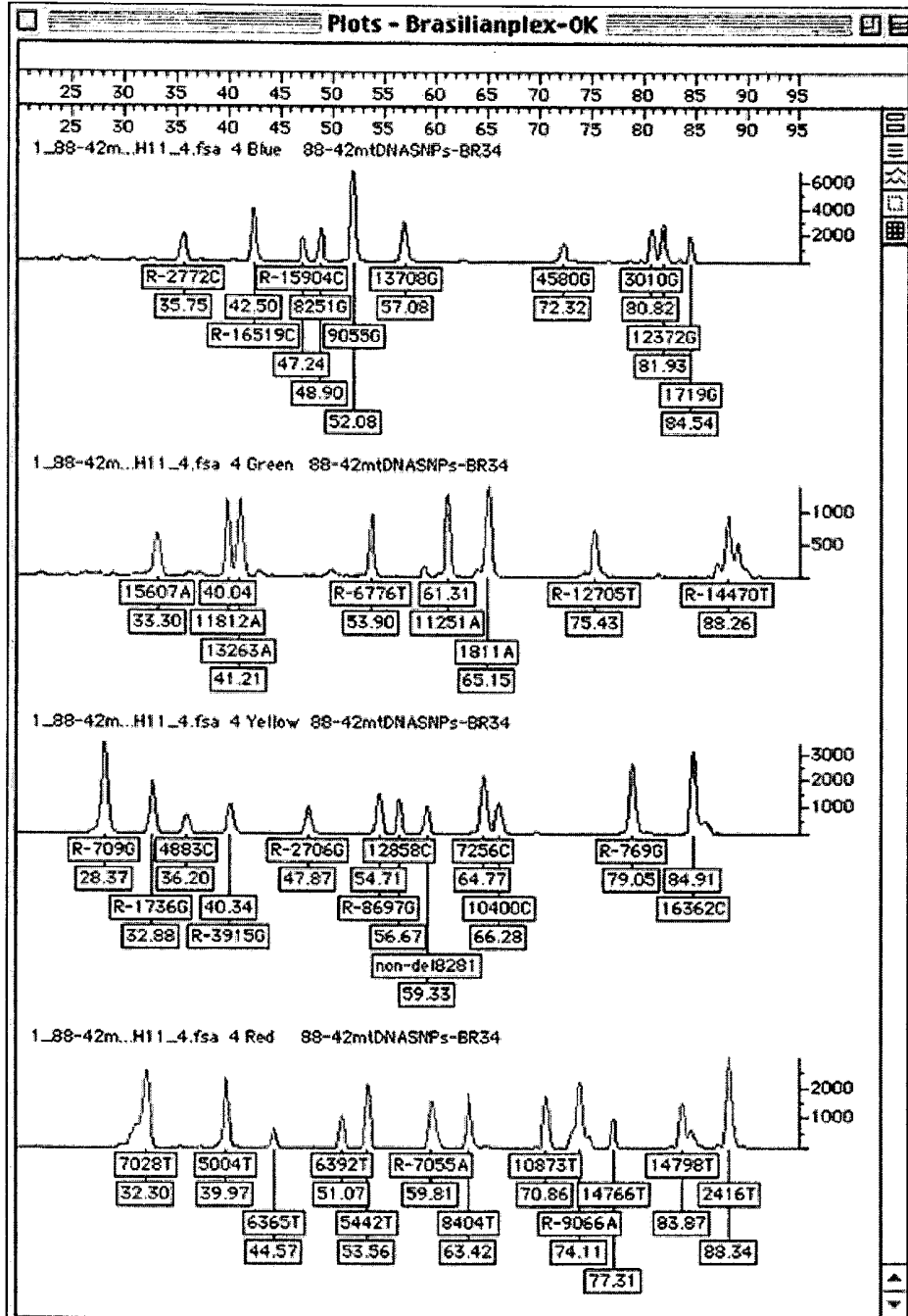
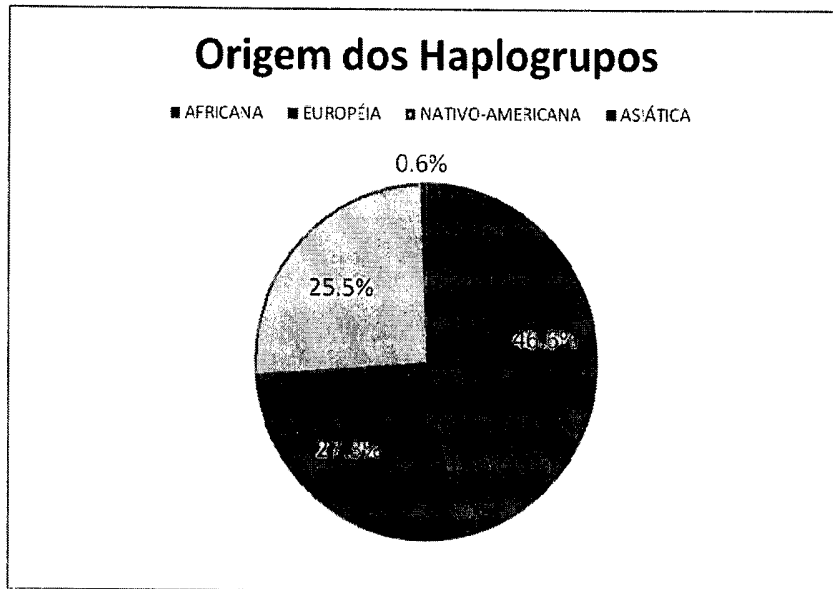


FIGURA 5



Resumo**MÉTODO E KIT PARA IDENTIFICAÇÃO GENÉTICA HUMANA POR MEIO DE
POLIMORFISMOS DO DNA MITOCONDRIAL PARA APLICAÇÃO EM
POPULAÇÕES MISCIGENADAS**

5 A presente invenção refere-se a um método e kit para
identificação genética humana por meio da análise de
polimorfismos específicos do DNA mitocondrial para
aplicação em populações miscigenadas como a brasileira, por
exemplo. A técnica desenvolvida, além de permitir a
10 identificação do indivíduo, permite também classificá-lo em
haplogrupos do DNA mitocondrial possibilitando a
identificação da origem ancestral materna do indivíduo
testado. A referida invenção pode ser aplicada na área de
genética forense, pela polícia científica ou por
15 laboratórios particulares, tendo como principais
beneficiados populações miscigenadas que não dispõem de
técnicas específicas para sua identificação e classificação
genética.