



Universidade do Estado do Rio de Janeiro
Centro Biomédico
Instituto de Biologia Roberto Alcântara Gomes

Fernanda Saloum de Neves Manta

**Marcadores Inserção/Deleção (Indel): estudo de ancestralidade e
identificação humana na população brasileira**

Rio de Janeiro

2013

Fernanda Saloum de Neves Manta

Marcadores Inserção/Deleção (Indel): estudo de ancestralidade e identificação humana na população brasileira



Tese apresentada, como requisito parcial para obtenção do título de Doutor, ao Programa de Pós-graduação em Biociências, da Universidade do Estado do Rio de Janeiro.

Orientador: Prof. Dr. Elizeu Fagundes de Carvalho

Coorientadora: Prof.^a Dra. Leonor Gusmão

Rio de Janeiro

2013

CATALOGAÇÃO NA FONTE
UERJ/REDE SIRIUS/BIBLIOTECA CB-A

M291 Manta, Fernanda Saloum de Neves.
Marcadores Inserção/Deleção (Indel): estudo de ancestralidade e identificação humana na população brasileira / Fernanda Saloum de Neves Manta.
126 f.

Orientador: Elizeu Fagundes de Carvalho
Coorientadora: Leonor Gusmão
Tese (Doutorado) – Universidade do Estado do Rio de Janeiro, Instituto de Biologia Roberto Alcântara Gomes. Pós-graduação em Biociências.

1. Polimorfismo (Genética) - Teses. 2. Genética forense - Teses. 3. Antropologia forense – Teses. 4. Brasil – População – Teses. 5. Hereditariedade - Teses. 6. Genética da população humana - Teses. 7. Mutação INDEL. I. Carvalho, Elizeu Fagundes de. II. Gusmão, Leonor. III. Universidade do Estado do Rio de Janeiro. Instituto de Biologia Roberto Alcântara Gomes. IV. Título.

CDU 591.151

Autorizo, apenas para fins acadêmicos e científicos, a reprodução total ou parcial desta tese, desde que citada a fonte.

Assinatura

Data

Fernanda Saloum de Neves Manta

Marcadores Inserção/Deleção (Indel): estudo de ancestralidade e identificação humana na população brasileira

Tese apresentada, como requisito parcial para obtenção do título de Doutor, ao Programa de Pós-graduação em Biociências, da Universidade do Estado do Rio de Janeiro.

Aprovada em: 04 de julho de 2013.

Coorientadora: Prof^a. Dra. Leonor Gusmão
Instituto de Biologia Roberto Alcântara Gomes – UERJ

Banca Examinadora:

Prof. Dr. Elizeu Fagundes de Carvalho (Orientador)
Instituto de Biologia Roberto Alcântara Gomes – UERJ

Prof. Dra. Cíntia Barros Santos-Rebouças
Instituto de Biologia Roberto Alcântara Gomes – UERJ

Prof. Dr. Milton Ozório Moraes
Fundação Oswaldo Cruz

Prof. Dr. Rinaldo Wellerson Pereira
Universidade Católica de Brasília

Rio de Janeiro

2013

DEDICATÓRIA

Aos meus filhos, Milena e Enzo, que foram minha fonte de inspiração e força. Obrigada por vocês existirem e me deixarem mais forte.

Aos meus pais, Dalva e Carlos, agradeço pelo que sou hoje e por todas as minhas conquistas. Sem a base de caráter, fé e educação que vocês me deram, nunca teria chegado até aqui.

Ao meu marido, Rafael, pelo amor, compreensão, credibilidade e acima de tudo pelo companheirismo.

AGRADECIMENTOS

Ao meu orientador, Dr. Elizeu, por me receber na sua equipe há nove anos e assim, me permitir conhecer e fazer parte do tão sonhado “mundo da genética forense” e principalmente por todas as oportunidades dadas.

À minha coorientadora, Leonor Gusmão, por todos os ensinamentos passados, com tanta dedicação e carinho. Muito obrigada por tudo e mais um pouco!!!

Ao meu colaborador e amigo, Rui Pereira, por tudo que me ensinou, por todos os momentos exaustivos, estressantes, porém necessários ... E como não vivemos apenas de trabalho, agradeço também pelos momentos alegres, repletos de companheirismo e confiança.

À Dra. Dayse Aparecida da Silva, que foi quem me ensinou “os primeiros passos” nessa área e que acompanhou minha trajetória desde o começo.

Aos meus ex-colegas de trabalho, mas atuais e sempre amigos, Suellen e Alexandre, por poder contar com vocês em todos os momentos desde os felizes e alegres aos mais difíceis. Sempre me dizendo palavras amigas, porém sinceras. Adoro vocês!

À Andrea, por toda parceria e gargalhadas que tivemos nesses 3 anos de doutoramento.

À Anna, Jú e Rose por estarem sempre por perto, dividindo sala, conversas e principalmente me dando todo o apoio físico e moral necessário nesse final.

Ao Rômulo e Alexandre, por serem meu apoio na parte prática, me ajudando em algumas tipagens e análises necessárias no decorrer dessa tese.

Aos amigos do LDD, Cris, Paty, João, Esther, Monique, Saulo, Holly, Januzzi, Marcinha, Henrique, Cíntia, Rosangela, Silvia, Celso, Agda, Jaque, Flávio e Marcos pela amizade cultivada em todos esses anos. Cada um de vocês tem uma importância singular nesses anos de LDD. Obrigada por tudo!

Aos novos companheiros doutorandos, porém não menos especiais, Chuck, Rodrigo e Kleyton, pelas conversas e momentos de terapia, afinal todos nós estamos no mesmo barco. Ao Chuck, em especial, agradeço por todo apoio dado nesse meu último ano de doutoramento, principalmente com o photoshop.

Aos meus amigos de vida, especialmente, à Sabrina, Viviane, Vivian, Suellen, Guilherme e Fábio, obrigada por toda amizade, carinho e paciência que vocês tiveram comigo nesses últimos anos. Desculpe a ausência e todas as vezes que vocês falaram comigo mas que eu não ouvia por que embora meu corpo estivesse presente, minha cabeça não estava. Sem vocês tudo teria sido mais difícil com certeza. Prometo que em breve voltarei ao normal. Amo vocês!

Aos meus filhotes, por aceitarem minha ausência em alguns momentos, tornando esse processo um pouco menos dolorido. Amo vocês mais que tudo!

Ao meu marido, minha mãe, sogra, sogro, cunhada e cunhado, por terem ajudado em todos os momentos nos quais precisei estar ausente da vida familiar, principalmente, nos 45 dias que estive em Portugal. Obrigada por cuidarem dos meus filhos com tanto carinho. O fato de ter vocês, com certeza me fez ter mais força de ir à busca do meu crescimento profissional.

A todas as pessoas que conheci no IPATIMUP, em especial, Dr. António Amorim, Cíntia Alves, Verónica Gomes, Miguel Gonçalves, Ricardo Araújo, Sofia Quental e Luis Alvarez, pela ótima recepção, atenção e carinho que recebi no período que passei no Porto.

Por fim, á todos que direta ou indiretamente colaboraram para este trabalho, o meu muito obrigada.

Talvez não tenha conseguido fazer o melhor, mas lutei para que o melhor fosse
feito.

Não sou o que deveria ser, mas Graças a Deus, não sou o que era antes.

Marthin Luther King

RESUMO

MANTA, Fernanda Saloum de Neves. *Marcadores de inserção/deleção (Indel): estudo de ancestralidade e identificação humana na população brasileira*. 2013. 126f. Tese (Doutorado em Biociências) - Instituto de Biologia Roberto Alcântara Gomes, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2013.

Os polimorfismos denominados Indels são variações de comprimento geradas por inserção ou deleção de um ou mais nucleotídeos em uma sequência de DNA. Estes marcadores genéticos vêm apresentando um grande potencial para fins forenses e populacionais por combinar características dos marcadores SNPs, tais como a capacidade de analisar fragmentos curtos (menores que 250pb) e baixas taxas de mutação, com a facilidade da genotipagem dos STR em uma única PCR, seguida de detecção dos fragmentos amplificados por eletroforese. Com o objetivo de avaliar a eficiência dos Indels em aplicações forenses e esclarecer os detalhes da formação de diferentes populações brasileiras através de dados genéticos, amostras populacionais de diferentes estados brasileiros foram genotipadas através de dois sistemas multiplex. O primeiro (indelplex-HID) foi otimizado para fins de Identificação Humana (HID) e inclui um grupo de 38 marcadores Indels selecionados por apresentarem altos valores de diversidade genética dentro das principais populações continentais. Já o segundo (46-AI-indels), foi selecionado para estudos de ancestralidade e é composto por um conjunto de 46 marcadores informativos de ancestralidade (AIMs). Nesse último caso, ao contrário do anterior, o sistema multiplex inclui marcadores com alta divergência nas frequências alélicas entre populações continentais. Na primeira etapa, o multiplex HID foi aplicado em uma amostra populacional do Rio de Janeiro e em uma amostra populacional dos índios Terena. Um banco de dados de frequências alélicas foi construído para essas duas amostras populacionais. Os valores das frequências alélicas foram utilizados nas comparações estatísticas e parâmetros de vínculos genéticos e forenses foram calculados. O Poder de Discriminação acumulado na população do Rio de Janeiro para os 38 *loci* testados foi de 0,99999999999999990 e na população dos índios Terena de 0,999999999999997, validando o uso desse sistema numa população heterogênea como a brasileira. A eficiência do indelplex-HID também mostrou-se elevada nas amostras de casos forenses comprometidas, apresentando melhor resultados que marcadores STR em termos de número de *loci* genotipados e de qualidade de amplificação. Na segunda etapa, o multiplex 46-AI-indels foi aplicado com objetivo de avaliar a ancestralidade em amostras de diferentes estados do Brasil por permitir a identificação de diferenças entre frequências alélicas de grupos populacionais separados geograficamente. A maioria das populações analisadas apresentou elevada herança européia. As populações do Rio de Janeiro, Pernambuco, Mato Grosso do Sul, Amazonas, Alagoas, Minas Gerais e São Paulo apresentaram cerca de 50% de ancestralidade européia, enquanto que nas populações que formam o sul do país e o Espírito Santo este percentual girou em torno de 70%. De uma maneira geral, as contribuições ameríndias e africanas variaram um pouco de acordo com a região. As amostras de Santa Isabel do Rio Negro e dos índios Terena (amostras indicadas como ameríndio-descendentes) de fato mostraram majoritariamente ancestralidade ameríndia (>70%). Os resultados obtidos indicaram que os dados gerados a partir da tipagem dos AIMs estão em estreita concordância com os registros históricos e com outros estudos genéticos acerca da formação da população brasileira e os *loci* do sistema HID evidenciaram que os são altamente informativos, constituindo uma ferramenta importante em estudos de identificação humana e de relações de parentesco.

Palavras-chave: Polimorfismo genético. Indels. População brasileira. Ancestralidade. Identificação humana.

ABSTRACT

Indels are length polymorphisms created by the insertion or deletion of one or more nucleotides in a DNA sequence. This type of genetic marker is potentially very useful for forensic and population genetic applications because it combines some desirable SNP features, such as the possibility of being analyzed in small fragments (less than 250bp), and low mutation rate, and in the same way as for the STRs it is easily genotyped in a single PCR followed by capillary electrophoresis detection of the amplified fragments. In order to evaluate the efficiency of Indels in forensic applications, and clarify some details on the formation of different Brazilian populations through genetic data, population samples from different Brazilian States were genotyped through two Indel multiplex systems. The first (Indelplex-HID) has been optimized for Human Identification (HID), and includes a group of 38 Indel markers selected by presenting similarly high values of genetic diversity within the main continental populations. The second (46-AI-indels) was selected for studies of ancestry, and it is composed by a set of 46 ancestry informative markers (AIMs). The latter, unlike the first one, includes markers with high divergence in allele frequencies among populations. In a first stage, the multiplex HID was used to study a population sample of Rio de Janeiro and a population sample of Terena Amerindians. Allele frequency databases were built for these two population samples. Allele frequency values were used in statistical comparisons, and to calculate relevant kinship and identification parameters. The cumulative discrimination power for the 38 loci in the population from Rio de Janeiro was 0.9999999999999990, and for the Terena Indians it was 0.999999999999997, demonstrating the utility of this system in a heterogeneous population like the Brazilian. The effectiveness of the indelplex-HID was also revealed to be high in low template samples that have been analyzed, which were previously selected from forensic cases, showing better results than STR markers in terms of number of genotyped loci and amplification quality. In the second stage, the multiplex 46-AI-indels was used to assess the ancestry makeup of different States of Brazil, since it allows the identification of differences among allele frequencies in geographically separated population groups. Most of the populations analyzed showed high European heritage. The populations from Rio de Janeiro, Pernambuco, Mato Grosso do Sul, Manaus, Alagoas, Minas Gerais and São Paulo presented about 50% European ancestry, while in Espírito Santo and south of Brazil the frequency of European ancestry was about 70%. In General, the Amerindian and African contributions varied somewhat according to the region. Santa Isabel do Rio Negro and Terena (indicated as Ameridians-descendants) in fact showed mostly Amerindian ancestry (>70%). The results of AIMs typing are in close agreement with the historical records and other genetic studies on the formation of Brazilian population. The loci of the HID system were highly informative, and constitute an important tool in human identification and kinship studies.

Keywords: Genetic Polymorphism. Indels. Brazilian population. Ancestry. Human Identification

LISTA DE FIGURAS

Figura 1 –	Comparação da análise dos diferentes tipos de marcadores genéticos.....	22
Figura 2 –	Detecção dos Indels por PCR.....	23
Figura 3 –	Herança de recombinação e não recombinação de segmentos do genoma ao longo das gerações.....	27
Figura 4 –	As migrações históricas no mundo.....	34

LISTA DE TABELA

Tabela 1 – Amostras biológicas utilizadas nas diferentes etapas do estudo.....	45
Tabela 2 – HID-Indels utilizados neste estudo, incluindo a localização de cada marcador no genoma, os alelos relatados e os tamanhos esperados dos <i>amplicons</i>	49
Tabela 3 – AIM-Indels utilizados neste estudo, incluindo a localização de cada marcador no genoma e os alelos relatados.....	51

LISTA DE ABREVIATURAS E SIGLAS

AIMs	Marcadores informativos de ancestralidade (do inglês, <i>Ancestry Informative Markers</i>)
dbSNP	Banco de dados de marcadores SNPs (do inglês, <i>Single Nucleotide Polymorphism Database</i>)
DNA	Ácido desoxirribonucléico (do inglês, <i>DeoxyriboNucleic Acid</i>)
Indels	Polimorfismos de inserção / deleção (do inglês, <i>Insertions and deletions polymorphisms</i>)
FUNAI	Fundação Nacional do Índio
HWE	Equilíbrio de Hardy-Weinberg (do inglês, <i>Hardy-Weinberg Equilibrium</i>)
HID	Identificação Humana (do inglês, <i>Human Identification</i>)
IBGE	Intituto Brasileiro de Geografia e Pesquisa
Kb	Quilo base
LDD	Laboratório de Diagnósticos por DNA da UERJ
MDS	Escalonamento Multidimensional (do inglês, <i>Multidimensional scaling</i>)
MP	Probabilidade de Coincidência (do inglês, <i>Matching Probability</i>)
mtDNA	DNA mitocondrial (do inglês, <i>Mitochondrial DNA</i>)
NRY	Região não recombinante do cromossomo Y (do inglês, <i>Non-recombining Region of the Y chromosome</i>)
pb	Pares de base
PD	Poder de Discriminação (do inglês, <i>Power of discrimination</i>)
PE	Poder de Exclusão (do inglês, <i>Power of exclusion</i>)
PCR	Reação em cadeia da polimerase (do inglês, <i>Polymerase Chain Reaction</i>)
PNAD	Pesquisa Nacional por Amostra Domiciliar
RNA	Ácido ribonucléico (do inglês, <i>Ribonucleic Acid</i>)
VNTR	Repetições em tandem de números variáveis (do inglês, <i>Variable Number of Tandem Repeat</i>)
SNP	Polimorfismo de base única (do inglês, <i>Single Nucleotide Polymorphism</i>)
STR	Repetições curtas em tandem (do inglês, <i>Short Tandem Repeat</i>)
UERJ	Universidade do Estado do Rio de Janeiro

SUMÁRIO

	INTRODUÇÃO	16
1	OBJETIVOS	41
1.1	Objetivo geral	41
1.2	Objetivos específicos	41
2	MATERIAL E MÉTODOS	43
2.1	Amostras biológicas	43
2.1.1	<u>Detalhamento das amostras populacionais</u>	44
2.2	Extração de DNA	46
2.2.1	<u>Extração por Salting-out (Miller, 1988)</u>	46
2.2.2	<u>Extração por Chelex (Walsh et al., 1991)</u>	46
2.2.3	<u>Extração por método semi-automatizado. Maxwell 16® (Promega Corporation)</u>	47
2.3	Genotipagem	47
2.3.1	<u>HID-Indelplex (Pereira et al., 2009)</u>	48
2.3.2	<u>46-AIM Indels (Pereira et al., 2011)</u>	50
2.4	Detecções dos alelos	52
2.5	Análise estatística	52
3	RESULTADOS	54
3.1	Artigo científico I: Indel markers: genetic diversity of 38 polymorphisms in Brazilian populations and application in a paternity investigation with post mortem material	55
3.2	Artigo científico II: Analysis of Genetic Ancestry in the Admixed Brazilian Population from Rio de Janeiro using 46 Autosomal Ancestry-Informative Indel Markers	60
3.3	Artigo científico III: Inferring the ancestry information to different Brazilian states using AIM-Indels autosomal	66
4	DISCUSSÃO	89
5	CONCLUSÕES	93
	REFERÊNCIAS	95
	ANEXO A - Termo de Consentimento Livre e Esclarecido	102

ANEXO B - Aprovação do Comitê de Ética em Pesquisa	103
ANEXO C - Lista de soluções e tampões utilizados.....	104
ANEXO D - Material suplementar dos artigos publicados e submetido.....	105
ANEXO E - Comprovante de submissão do Artigo III à PLOS ONE ...	126

INTRODUÇÃO

O estudo do DNA ocupa, nos dias atuais, um lugar de destaque na comunidade científica. O extenso conhecimento adquirido sobre a estrutura e função da molécula e os avanços das técnicas de Biologia Molecular contribuíram para a utilização do DNA nas mais diferentes áreas das Ciências Biológicas, Medicina e Antropologia. Além disso, a temática do estudo do DNA foi inserida não somente na comunidade científica, mas também em assuntos relacionados ao cotidiano da sociedade.

ESTRUTURA DO DNA E O GENOMA HUMANO

O genoma humano corresponde ao total de informação genética contida nas células humanas, armazenado na molécula de DNA (ácido desoxirribonucleico). O DNA organiza-se em dois componentes: um nuclear, complexo, que contém a quase totalidade da informação genética, e um mitocondrial, mais simples, onde ocorrem poucos genes, porém vitais, herdado exclusivamente por herança materna.

O genoma nuclear diploide humano está distribuído em unidades denominadas cromossomos, constituídos por 22 pares de autossomos e um par de cromossomos sexuais. Nas mulheres, o par sexual é formado por dois cromossomos X enquanto que nos homens é formado por um cromossomo X e um cromossomo Y.

Até os anos 80, a análise da variabilidade genética humana estava limitada a aproximadamente 3% do genoma humano, por meio do estudo de polimorfismos proteicos, tendo destaque o grupo sanguíneo ABO e os marcadores genéticos do Complexo de Histocompatibilidade (sistema HLA), ou chamados de “marcadores clássicos” (Cavalli-Sforza *et al.*, 1994).

Após este período, o extraordinário desenvolvimento tecnológico da Genética Molecular possibilitou o estudo de diversos tipos de variações genéticas (“polimorfismos”) através da análise da estrutura da molécula de ácido desoxirribonucleico (DNA) (Pena & Birchal, 2006). Em 2001, a publicação da

sequência completa do genoma humano foi o marco que possibilitou um melhor conhecimento acerca do conteúdo e arranjo da sequência do DNA (Lander *et al.*, 2001).

O DNA está presente em todas as células nucleadas e possui características particulares, nas quais se fundamentam a identificação genética de indivíduos. Cada indivíduo apresenta no seu genoma um arranjo único e exclusivo das unidades que compõem o DNA. Tal arranjo das seqüências nucleotídicas confere a cada ser vivo uma identidade também única e exclusiva, o que permite que a “leitura” do DNA, ou trechos dele, possibilite identificar cada ser humano a partir de qualquer vestígio biológico, tais como sangue, saliva, pelos, pele, entre outros (Butler, 2009).

O DNA é composto por regiões gênicas que comportam genes com funções determinadas, nas quais se inserem codificações para a produção de proteínas e ácidos ribonucleicos, e regiões extragênicas caracterizadas por sequências de nucleotídeos que não são traduzidas em proteínas. As regiões extragênicas são subdivididas em sequências únicas ou em baixo número de cópias e em sequências repetitivas, sequências repetidas *em tandem* ou repetições agrupadas e as repetições dispersas. Segundo Lander e colaboradores (2001), estas zonas de DNA repetitivo constituem cerca de 50% do genoma humano.

POLIMORFISMOS DE DNA

Novos arranjos nas sequências de nucleotídeos do material genético caracterizam-se como mutações. Quando uma mutação é verificada no genoma de uma população com frequência igual ou superior a 1 %, esta região é considerada sítio polimórfico (Jobling *et al.*, 2004).

A palavra polimorfismo tem sua origem no grego antigo, onde “poli” significa “muitos” e “morphos” significa “formas”, portanto, polimorfismo exprime a idéia de “muitas formas”. Na biologia molecular as regiões polimórficas são caracterizadas pela ocorrência simultânea, no mesmo *locus*, de duas ou mais formas diferentes ou alelos.

As seqüências polimórficas são tipicamente referidas como marcadores de DNA, sendo o local que apresenta um dado marcador genético denominado de *locus*. Em humanos, indivíduos diploídes, verificam-se, portanto, dois alelos para um dado *locus* genético, um herdado por via materna e outro paterna. Exceção ocorre na região não homóloga de cromossomos sexuais (X e Y). Em relação a um dado *locus*, quando os dois alelos diferem entre si, o indivíduo é dito heterozigoto. Por outro lado, sendo os alelos idênticos, é dito homozigoto. Em termos populacionais, verifica-se um menor ou maior número de alelos para um dado *locus*, o que se relaciona com o grau de polimorfismo do *locus*.

Teoricamente, qualquer região polimórfica do DNA pode ser utilizada como um marcador genético para fins de identificação humana, sendo as seqüências repetitivas as de maior interesse por apresentarem um elevado grau de variabilidade interindividual (Jobling *et al.*, 2004).

A variação genética existente na seqüência do DNA humano compreende os polimorfismos de tamanho e os de seqüência. Os polimorfismos de tamanho podem ser gerados pela inserção ou deleção de um ou mais nucleotídeos na seqüência do DNA. Este grupo inclui um grande número de variantes genéticas como as inserções e deleções de pequenos ou grandes fragmentos de DNA, os elementos repetitivos, como os minissatélites e microsatélites, e outras inserções/deleções de seqüências aleatórias de DNA (Lander *et al.*, 2001). Os polimorfismos de seqüência surgem pela substituição, em um dado local do genoma, de um ou mais nucleotídeos da seqüência do DNA. Os polimorfismos de seqüência mais comuns são as substituições de uma só base, também conhecidos como polimorfismos de nucleotídeo único.

MINISSATÉLITES - *Variable Number of Tandem Repeat (VNTR)*

Os minissatélites, ou *loci* VNTR (do inglês, *Variable Number of Tandem Repeat*), são polimorfismos de tamanho que consistem em repetições em *tandem* de unidades básicas de 10 a 60 pares de base (pb), com tamanhos que variam de centenas a milhares de bases (Kb). Os minissatélites são encontrados em

abundância no genoma e foram os primeiros polimorfismos do DNA a serem utilizados em estudos de identificação humana (Jeffreys *et al.*, 1985a, b; Kobilinski *et al.*, 2005).

Os VNTR são marcadores altamente informativos devido ao seu elevado grau de variabilidade. Entretanto, algumas condições da técnica de análise limitam o uso desta metodologia em larga escala. Condições como a quantidade mínima de DNA amostral e o extenso comprimento dos fragmentos de DNA necessários para a análise encontram-se entre os principais fatores que contribuíram para que atualmente a sua aplicação em Genética Forense seja residual, uma vez que o estudo de microssatélites se apresenta mais vantajoso.

MICROSSATÉLITES - *Short Tandem Repeats* (STR)

Os microssatélites, ou *loci* STR (do inglês, *Short Tandem Repeats*) apresentam repetições com unidades básicas de 2 a 6 pb, sendo o polimorfismo baseado no número de repetições.

Os STRs possuem características que os tornaram marcadores de escolha em diversas temáticas científicas, tais como: ampla distribuição no genoma humano; elevado grau de polimorfismo de tamanho; taxa de mutação na ordem de 10^{-3} e facilidade de análise face aos tamanhos reduzidos quando comparados aos dos minissatélites, o que possibilita a sua amplificação pela reação em cadeia da polimerase, do inglês, *Polymerase Chain Reaction* – PCR (Charlesworth *et al.*, 1994; Butler 2005a).

As análises de parentesco distante e de ancestralidade através da utilização de marcadores STR são dificultadas devido à elevada taxa de mutação deste marcador, responsável pelo aumento do número de variações alélicas existentes em um *locus*. Apesar de não se apresentarem como marcadores de escolha nos casos citados acima, estes marcadores são ferramentas úteis em análise genética de identificação individual ou em estudos que buscam avaliar eventos populacionais mais recentes,

uma vez que possuem uma diversidade alélica que, em conjunto, permitem a criação de perfis genéticos individuais.

Atualmente, os microssatélites são os marcadores genéticos mais utilizados em estudos genéticos populacionais. Entretanto, a utilização de novos marcadores que pertencem à classe de polimorfismos de DNA de evolução lenta, constituídos de fragmentos de DNA que podem ter sido substituídos, acrescentados ou eliminados de uma sequência do genoma vem crescendo.

POLIMORFISMO DE BASE ÚNICA - *Single Nucleotide Polymorphisms* (SNP)

Eventos mutacionais como inserções, deleções e substituições de bases ocorrem com frequência muito baixa, na ordem de 10^{-9} , e geram uma pequena variabilidade genética, já que, pela sua natureza, originam apenas dois estados alélicos, ancestral e mutado, designados de polimorfismos bialélicos (Butler, 2005a; Butler *et al.*, 2007).

Dentre os polimorfismos de sequência presentes na molécula de DNA humano, as substituições de um único nucleotídeo (SNPs), são as mais frequentes (Lander, 2001; International HapMap Consortium 2003; 2005). Estes polimorfismos são originados por mutações pontuais, que produzem diferenças entre sequências de cromossomos em um único par de bases (Budowle e Daal, 2008).

Mais de dezessete milhões de SNPs já foram mapeados, sendo as respectivas informações depositadas em bases de dados (dbSNP- <http://www.ncbi.nlm.nih.gov/SNP>). Além dos sistemas multiplexes de SNPs desenvolvidos para fins de identificação humana (Sanchez *et al.*, 2006), com ampla aplicação em estudos relativos às migrações de populações através dos tempos, os *loci* SNP vêm sendo utilizados como marcadores de susceptibilidade a diversas doenças e/ou seus respectivos graus de severidade e respostas diferenciais a drogas (Melton, 2003; Phillips *et al.*, 2004). Também estão descritos na literatura estudos que empregam SNPs para prever a origem étnica e as características físicas do indivíduo (Butler, 2005a).

Devido ao pequeno tamanho dos SNPs, estes marcadores apresentam um aumento da capacidade de genotipar eficientemente DNA degradado ou em pequena quantidade. A implementação do uso dos SNPs em laboratórios forenses e seu emprego como uma ferramenta complementar em investigações de parentesco, ainda é limitada devido, principalmente, ao alto custo e à complexidade técnica envolvida na tipagem de tais marcadores (Sanchez *et al.*, 2006; Butler, 2007; Budowle & Daal, 2008).

POLIMORFISMOS DE INSERÇÃO E DELEÇÃO - *Insertion Deletion Polymorphisms* - Indels

Os polimorfismos denominados Indels são variações de comprimento geradas por inserção ou deleção de nucleotídeo(s) em uma sequência de DNA, cujos tamanhos podem variar de uma a milhões de bases até grandes pedaços do cromossomo. Não obstante, o maior grupo é constituído por Indels, cujas diferenças entre os alelos residem em apenas alguns nucleotídeos. Depois dos SNPs, os Indels são as variações mais abundantes no genoma, compreendendo cerca de 20% de todos os polimorfismos existentes no DNA humano (Weber *et al.*, 2002; Mills *et al.*, 2006).

Os Indels localizados no interior de regiões codificadoras de um gene podem causar alteração na fase de leitura ou adição/subtração de algum códon, o que, conseqüentemente, irá alterar o produto protéico. Além disso, os Indels em regiões não traduzidas podem afetar a estrutura e estabilidade da molécula de DNA ou comprometer a expressão de RNAs regulatórios não codificantes. Evidências genéticas sugerem que os Indels constituem uma fonte considerável de “defeitos genéticos”, ou seja, uma força produtora de mudanças evolutivas significantes (Britten *et al.*, 2003).

Em 2002, Weber e colaboradores identificaram e caracterizaram 2000 polimorfismos bialélicos de inserção-deleção no genoma humano e indicaram a aplicação dos Indels em estudos genéticos, devido a sua abundância e à facilidade de análise. A partir deste momento, os Indels passaram a constituir ferramentas

poderosas em estudos acerca de migrações populacionais, fluxo genético e ancestralidade, o que possibilitou o início da construção do banco de dados *online* dos polimorfismos de inserção/deleção (<http://www.marshfieldclinic.org/mgs/>). Este banco inclui informações como: o tipo de polimorfismo (bialélico ou multialélico), diferenças de comprimento de alelos, desenho de *primers*, posição cromossômica, assim como, frequências alélicas nas maiores populações mundiais. Os Indels até então caracterizados foram também incluídos no banco de dados - dbSNP, permitindo assim uma maior facilidade no acesso à informações. Desde então, o número de estudos utilizando os Indels com diferentes propósitos cresceu (Rosenberg *et al.* 2005; Bastos-Rodrigues *et al.*, 2006; Pereira *et al.*, 2009; Santos *et al.*, 2010; Pena *et al.*, 2011; Vullo *et al.*, 2011; Francez *et al.*, 2012; Pereira *et al.*, 2012).

Os Indels também passaram a chamar a atenção da comunidade forense por apresentarem um elevado potencial como marcadores genéticos nesta área. Eles combinam características dos marcadores SNPs, tais como, a capacidade de analisar fragmentos curtos (menores que 250 pb) e baixas taxas de mutação (10^{-9}), com a facilidade da genotipagem dos STR, ou seja, necessita-se de uma única PCR seguida da detecção dos fragmentos amplificados por eletroforese. Com isso, o número de reações químicas necessárias para uma análise é menor, o que facilita os processos e diminui os custos (Pereira & Gusmão, 2011) (Figura 1).

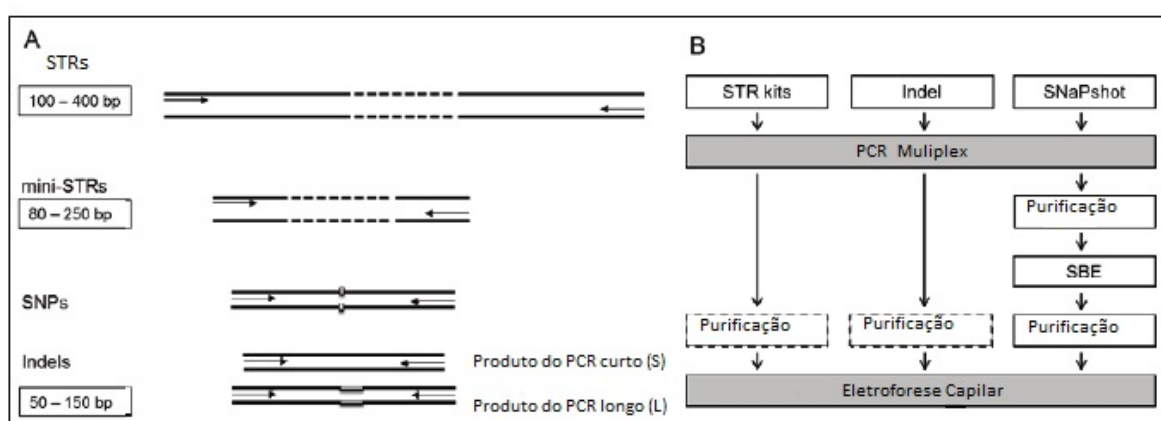


Figura 1- Comparação da análise dos diferentes tipos de marcadores genéticos.

(A) Indels, como os SNPs, são passíveis de análises em *amplicons* muito curtos comparados aos STRs e até mesmo em relação aos Mini-STR; (B) Indels, como STRs, são facilmente tipados com uma simples PCR seguida de uma eletroforese capilar (fonte: adaptada de Pereira & Gusmão, 2011).

Os dois alelos dos marcadores Indels bialélicos podem ser classificados como curto ou longo (Butler, 2005a). Os mini-Indels bialélicos, cujas diferenças entre os alelos residem em apenas alguns nucleotídeos, além de possuírem a capacidade de análise através de uma única PCR, assim como os STR, se tornaram marcadores genéticos mais práticos e fáceis de analisar (Figura 2), devido ao tamanho reduzido e à necessidade de verificar apenas a presença ou ausência de uma inserção/deleção.

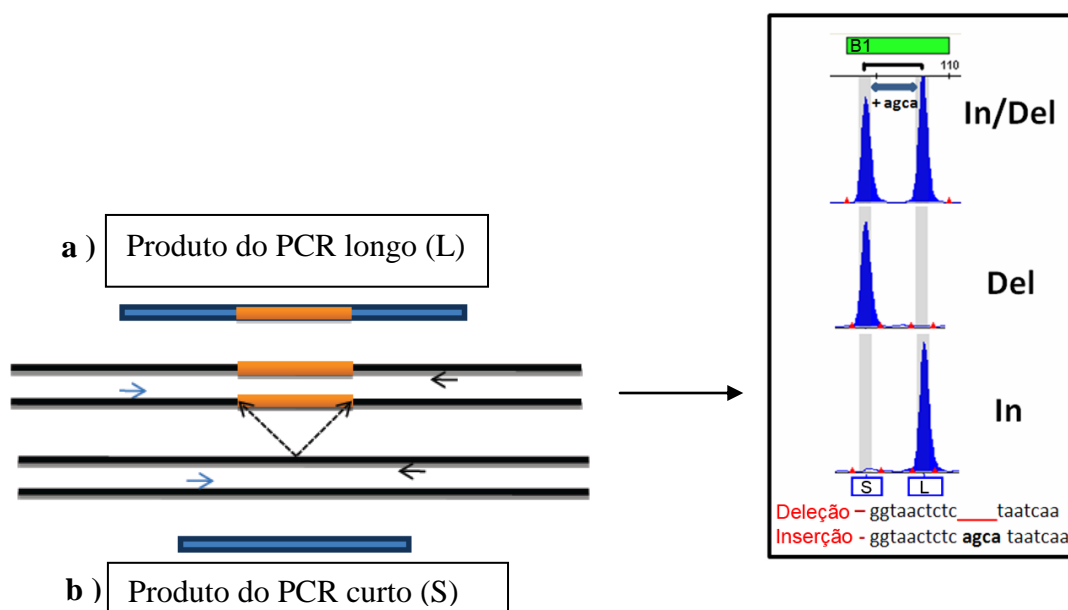


Figura 2. Detecção dos Indels por PCR.

Os *primers* da reação estão representados pelas setas e se encontram nas regiões flangeadoras da sequência alvo. **a)** Se o segmento de DNA estiver presente (caixa laranja), o produto do PCR correspondente ao alelo longo será amplificado. **b)** Se o segmento de DNA estiver ausente, o produto de PCR correspondente ao alelo curto que será amplificado (Fonte: adaptada de Pereira & Gusmão, 2011).

O emprego de marcadores genéticos de evolução lenta, como os Indels, em estudos populacionais diminui a probabilidade de mutação recorrente, mas gera uma menor diversidade de alelos, pois cada marcador apresenta apenas dois alelos (Pereira & Gusmão, 2011). Entretanto, estes marcadores são ferramentas úteis para estudos de Genética de Populações, já que os alelos mais recentes apresentam uma distribuição geográfica restrita, enquanto que os alelos que surgiram há mais tempo já

foram espalhados por meio dos movimentos migratórios e apresentam-se em frequências elevadas na região próxima ao seu surgimento.

Vale ressaltar, que esse tipo de marcador, além da utilização em identificação humana ainda é de interesse relevante em estudos de ancestralidade por permitir a identificação de mínimas diferenças entre frequências alélicas de grupos populacionais separados geograficamente (Santos *et al.*, 2010; Pereira *et al.*, 2012).

CARACTERÍSTICAS DE UM MARCADOR GENÉTICO

Os diferentes tipos de polimorfismos de DNA podem ser classificados de acordo com a natureza molecular e sua localização no genoma. A qualidade de um marcador genético para estudos de genética de populações depende de três características principais: o seu tempo de evolução, taxa de mutação e taxa de recombinação (Shaffner, 2004).

A idade, ou tempo de evolução, de um *locus* corresponde à diferença temporal entre o momento da pesquisa e o do ancestral comum mais antigo, ou seja, define o período no qual suas variações genéticas ocorreram, delimitando assim o período histórico que pode ser investigado utilizando determinado *locus*. Estudos envolvendo populações humanas utilizam marcadores com diferentes escalas temporais para avaliar eventos históricos recentes, a expansão do *Homo sapiens* para fora da África e o surgimento do homem moderno (Cavalli-Sforza *et al.*, 1994; Campbell e Tishkoff, 2008).

A taxa de mutação muito baixa, como muito alta, pode transformar um marcador em não informativo. De acordo com a sua taxa de mutação, os marcadores polimórficos são classificados como de evolução lenta, caracterizados por uma baixa taxa de mutação (10^{-7} a 10^{-9} por geração), ou de evolução rápida, com uma elevada taxa de mutação, na ordem de 10^{-3} (Santos & Tyler-Smith, 1996). No primeiro caso, a taxa combinada com o pouco tempo de evolução, não permitirá que haja variação genética significativa no marcador. Entretanto, se a taxa for muito alta, assim como, o tempo de evolução, mutações recorrentes poderão ter ocorrido em muitos

nucleotídeos, o que tornará obscuro o processo de compreensão da relação entre os alelos.

Na prática, as substituições simples de nucleotídeo apresentam taxas baixas o suficiente para tornar negligenciáveis a ocorrência de mutações recorrentes. A exceção está no DNA mitocondrial (mtDNA), que evolui sob taxas maiores que o resto do genoma. Os microssatélites, por outro lado, possuem taxas muito mais altas que tornam a ocorrência de mutações recorrentes problemáticas após poucas dezenas de milhares de anos (Pakendorf & Stoneking, 2005). As mutações recorrentes são caracterizadas quando uma alteração na sequência de DNA ocorre mais de uma vez de forma independente, sem necessariamente ter a mesma origem filogenética. Portanto, este evento mutacional pode proporcionar situações em que dois indivíduos apresentem o mesmo alelo para determinado *locus*, sem compartilhar qualquer ascendente próximo em comum (Schaffner, 2004).

Por último, a taxa de recombinação controla o tamanho de determinada região cromossômica que compartilha uma história genealógica única. Na ausência de recombinação, todo o cromossomo passa a pertencer a um único ramo da árvore filogeográfica. Por outro lado, a recombinação divide os cromossomos autossômicos e o cromossomo X em regiões a cada geração, dificultando a reconstrução filogeográfica destes, tendo em vista que diferentes regiões cromossômicas representarão diferentes genealogias. Portanto, estudos histórico-evolutivos que envolvam análises de cromossomos que sofrem recombinação devem restringir-se a marcadores presentes em *loci* com alto desequilíbrio de ligação (Schwartz & Vissing, 2003; Pakendorf & Stoneking, 2005).

A mutação e a recombinação são eventos biológicos naturais, enquanto que o tempo de evolução de um *locus* é baseado na aliança entre tamanho populacional e a deriva genética. Populações maiores tendem a ter diferentes formas alélicas, em *loci* neutros, mais antigas. Isto ocorre, pelo efeito da extinção de alelos, ocasionado pela flutuação aleatória da frequência destes ao longo de gerações, ser dificultado. A taxa de recombinação e o tempo de evolução são determinantes do tamanho de desequilíbrio de ligação que é observado entre pares de *loci*, uma vez que a recombinação, ao longo do tempo, tende a quebrar este desequilíbrio genético (Santos & Tyler-Smith, 1996; Schwartz & Vissing, 2003; Schaffner, 2004).

Nesse sentido, os diferentes tipos de marcadores genéticos apresentam características específicas que definem a sua utilidade em campos distintos da pesquisa genética.

O conjunto de marcadores genéticos mais utilizados atualmente (STRs, SNPs e Indels) é encontrado tanto nos autossomos como nos cromossomos sexuais (Butler, 2005a).

MARCADORES GENÉTICOS AUTOSSÔMICOS

Os autossomos representam cerca de 93% do genoma humano total (*International Human Genome Sequencing Consortium, 2004*) e sua informação genética é transmitida de pais para filhos em um modo típico de herança mendeliana, com igual contribuição materna e paterna.

A maior proporção da variabilidade genética observada nos cromossomos autossômicos, quando comparados aos sexuais e ao DNA mitocondrial (mtDNA), está relacionada ao fenômeno de recombinação que ocorre entre os cromossomos homólogos a cada geração (Joblin *et al.*, 2004) (Figura 3). Portanto, os marcadores autossômicos constituem a principal fonte de marcadores genéticos para serem usados em diversas aplicações, tanto na área da Biologia Forense quanto na da Genética de Populações, uma vez que as histórias de ambos os sexos são igualmente representadas.

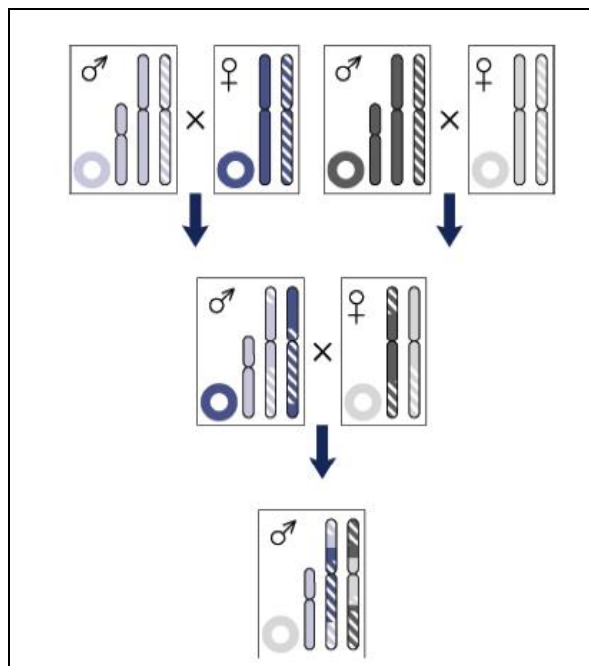


Figura 3. Herança de recombinação e não recombinação de segmentos do genoma ao longo das gerações. O cromossomo Y dos filhos (símbolo de cromossomo pequeno) e mtDNA (circular) descendem cada um apenas de uma linhagem dos avós, representando a herança patrilinear e matrilinear, respectivamente. Em contraste, os autossomos (símbolo de cromossomo comprido) descendem de todos os avós, gerando segmentos cromossômicos que sofreram recombinações em todas as gerações (Fonte: Adaptada de Jobling *et al.*, 2004).

MARCADORES DE LINHAGEM (mtDNA E CROMOSSOMO Y)

Marcadores localizados na região não recombinante do cromossomo Y (NRY) e no DNA mitocondrial (mtDNA) são herdados de forma uniparental, sendo por estes motivos chamados de marcadores de linhagem. Estes são transmitidos sem sofrer alteração, a não ser por mutação, possibilitando que linhagens maternas sejam traçadas por meio da análise de marcadores do mtDNA e linhagens paternas pela análise de marcadores da NRY (Butler, 2005a).

O cromossomo Y está presente somente em indivíduos do sexo masculino e é transmitido por via paterna a todos os descendentes masculinos. No âmbito forense, a caracterização de marcadores do cromossomo Y é particularmente útil em casos de agressão sexual praticada por homens, uma vez que permite a amplificação diferencial do material genético do agressor e, conseqüentemente, a

obtenção do seu perfil genético; e em testes de paternidade em que o descendente é do sexo masculino. No entanto, o modo de transmissão do cromossomo Y implica em limitações do seu uso no contexto forense em casos de concordância entre dois perfis genéticos, uma vez que a mesma informação genética é partilhada por todos os indivíduos masculinos numa linhagem paterna. Deste modo, a caracterização de marcadores do cromossomo Y pode ser altamente esclarecedora em casos de exclusão, mas tem um valor probatório limitado em casos de concordância de perfis genéticos entre amostras.

Devido às peculiaridades de sua herança uniparental e ausência quase que total de recombinação, o cromossomo Y possui excelentes marcadores para a inferência de filogenias paternas. Os haplótipos, que são conjunto de polimorfismos verificados em um indivíduo, estão relacionados por descendência com um ancestral comum, constituindo assim diferentes haplogrupos (conjunto de haplótipos intimamente relacionados por apresentarem ancestral comum próximo) (Jobling *et al.*, 1997; Jobling & Tyler-Smith, 2003).

O estudo da sequência de DNA mitocondrial (mtDNA) apresenta uma grande vantagem em casos forenses. Em muitos casos, o uso do mtDNA é o único modo de obter resultados positivos em amostras altamente degradadas ou com reduzido número de células. Em uma célula, existe apenas uma molécula de DNA nuclear, enquanto que do mtDNA existe um grande número de cópias. Desta maneira, a probabilidade de amplificar cópias não degradadas da molécula de mtDNA em amostras forenses é maior. No entanto, os estudos do mtDNA também apresentam limitações em uso forense, devido ao modo de transmissão por linha materna a todos os seus descendentes (Jobling & Gill, 2004; Kayser & de Kniff, 2011).

Assim como no caso do cromossomo Y, estudos direcionados à determinação de haplogrupos de mtDNA vem sendo muito utilizados na inferência de ancestralidade biogeográfica de indivíduos, já que os haplogrupos definidos por grupos de sequências (haplótipos) estão relacionados a uma origem ou região geográfica comuns (Pakendorf & Stoneking, 2005).

MARCADORES INFORMATIVOS DE ANCESTRALIDADE (AIM)

A maior parte da variabilidade genética entre os humanos é interindividual (80-90%) e apenas 10-20% da variação total é devido às diferenças entre populações (Cavalli-Sforza *et al.*, 1994). Além disso, também já foi amplamente observado que a maioria das populações possui alelos em comum para qualquer *locus*, e alguns alelos que são mais freqüentes em uma população também podem ser encontrados em altas freqüências em outras populações. Esse alto grau de similaridade genética reflete a recente divergência da espécie humana em grupos continentais (Shriver *et al.*, 1997). Existem, contudo, alguns marcadores de DNA que são específicos ou que apresentam freqüências diferenciadas entre populações definidas geográfica e/ou etnicamente (Cavalli-Sforza *et al.*, 1994; Santos *et al.*, 2010; Pereira *et al.*, 2012).

Os marcadores denominados de informativos de ancestralidade (*Ancestry Informative Markers* – AIMS) são aqueles que apresentam alto diferencial entre suas freqüências alélicas em populações ancestrais diferentes ou geograficamente distantes. Estes marcadores são especialmente úteis para inferir a origem ancestral de um dado indivíduo e estimar as proporções de distintas origens ancestrais de populações miscigenadas (Shriver *et al.*, 1997). Conseqüentemente, os AIMS apresentam um interesse geral em estudos de populações e Genética Clínica, bem como em Genética Forense, onde estes em alguns casos possibilitam a correlação de características físicas e ancestrais de um dado indivíduo, podendo ser úteis em investigações criminais (Phillips *et al.*, 2007).

UTILIZAÇÃO DE MARCADORES GENÉTICOS EM ESTUDOS DE IDENTIFICAÇÃO HUMANA NA PRÁTICA FORENSE

Atualmente, os estudos genéticos são freqüentemente requeridos pelo sistema de justiça, tendo-se tornado prática corrente no auxílio da resolução de questões legais em situações cada vez mais complexas. Neste contexto, a

designação de “genética forense” surge pela aplicação de conhecimentos científicos da área da genética com o objetivo de obter informações no âmbito judicial. No campo forense, o DNA assume assim um papel preponderante em estudos de identificação, especialmente em investigações de relações de parentesco, mais freqüentemente em testes de paternidade, e em criminalística biológica, por exemplo, na identificação do doador de vestígios biológicos encontrados no local de um crime.

Desde a década de 1980, os polimorfismos de DNA passaram a ser utilizados em estudos de identificação genética, substituindo progressivamente os grupos sanguíneos e proteínas, resultando em um enorme avanço científico e tecnológico na área da genética forense (Carracedo & Sánchez-Diz, 2005).

As técnicas de genotipagem de marcadores STR tornaram-se usuais em laboratórios de Genética Forense, permitindo a uniformização de procedimentos e a comparação de resultados entre laboratórios. A facilidade de análise pela utilização da técnica de PCR, a possibilidade de obter a genotipagem de amostras biológicas mesmo a partir de pouca quantidade de DNA, foi fundamental para o STR tornar-se o marcador preferencial para estudos de identificação humana (Buttler, 2005a).

A técnica de reação em cadeia da polimerase (PCR) é baseada na síntese enzimática "*in vitro*" de milhões de cópias de um único segmento de DNA orientada por oligonucleotídeos, usualmente referidos como *primers* (iniciadores), e catalisada pela enzima DNA polimerase de *Thermos aquaticus*, conhecida por *Taq* DNA Polimerase. Durante o processo de PCR, verifica-se o pareamento de um par de *primers*, seguido pela extensão da cadeia polinucleotídica através da ação da DNA polimerase, os quais delimitam a sequência alvo da amplificação na fita de DNA (Butler, 2003). A PCR ganhou rápida aceitação em toda comunidade científica como uma ferramenta biotecnológica para a realização de estudos de Genética de Populações, diagnósticos moleculares, mapeamento do genoma e na identificação de indivíduos.

Uma vez que a técnica tornou-se bastante acessível e o desenvolvimento de sistemas multiplexes de amplificação tornou possível a genotipagem simultânea de um grande número de marcadores a partir de uma única PCR, laboratórios em universidades e centros de pesquisa pelo mundo implementaram a realização de

estudos populacionais, estabelecendo-se banco de dados de frequências de alelos, oferecendo suporte estatístico na área da Genética Forense (Butler, 2005b).

A análise dos STRs tornou-se amplamente utilizada na Genética Humana, devido ao seu alto poder de individualização e praticidade (Szibor *et al.*, 2005). A capacidade informativa dos marcadores genéticos está necessariamente ligada ao grau de polimorfismo. Os STRs, até o momento, são os marcadores mais usados no campo forense por serem multialélicos e, portanto, altamente polimórficos, sendo bastante discriminativos quando na identificação de indivíduos. Porém, na análise de amostras difíceis, com DNA altamente degradado ou em pequena quantidade, é praticamente impossível obter sucesso na amplificação de *amplicons* longos. Deste modo, cientistas forenses concentraram-se na descoberta de novos marcadores e no desenvolvimento de novas metodologias (Kayser & de Knijff, 2011).

Inicialmente os SNPs representavam os marcadores de escolha para resolver o problema da análise de *amplicons* curtos. No entanto, a análise dos SNPs apresentou algumas desvantagens, tais como, metodologia de análise complexa, natureza bialélica que implica na necessidade de um maior número de marcadores para gerar um poder de informação igual ao dos multiplex STR e a dificuldade na distinção entre perfis em amostras com mistura (Pereira *et al.*, 2009).

Os marcadores Indels surgem em seguida, reunindo as mesmas vantagens do SNPs para aplicações forenses, porém com uma metodologia simples que é suficientemente informativa na maioria dos casos forenses e que se encontra implementada em qualquer laboratório que já utilize os STRs (Pereira & Gusmão, 2011). Vários estudos apontam para a possibilidade dos SNPs e Indels substituírem os microssatélites, podendo inclusive ser utilizados nos bancos de dados de criminosos (Weber *et al.*, 2002; Budowle, 2004; Gill *et al.*, 2004).

No tocante à aplicação dos Indels, os mesmos podem ser selecionados com propósito de identificação humana (Pereira *et al.*, 2009; Francez *et al.*, 2012; LaRue *et al.*, 2012) ou análise de ancestralidade (Bastos-Rodrigues *et al.*, 2006; Santos *et al.*, 2011; Pereira *et al.*, 2012) sendo a frequência alélica, encontrada para cada um desses marcadores nos diferentes grupos continentais, o principal critério de escolha.

Em 2009, Pereira e colaboradores desenvolveram um multiplex formado por 38 marcadores mini-Indels autossômicos bialélicos destinado a Identificação Humana e demonstraram a eficiência da análise deste em obter perfis completos, nos casos em que os STRs não proporcionaram informação suficiente, por gerarem *amplicons* mais curtos. Confirmando assim a utilidade desse tipo de marcador não somente em identificação individual como em análises de amostras de má qualidade ou quantidade limitadas.

Recentemente, sistemas multiplexes constituídos por Indels informativos de ancestralidade foram desenvolvidos e utilizados em diferentes populações (Santos, *et al.*, 2010; Pena *et al.*, 2011; Pereira *et al.*, 2012), se tornando uma nova estratégia no campo da genética forense a nível individual por permitir avaliar proporções de ancestralidade ou a origem geográfica de um dado indivíduo, dado esse que pode ser bastante útil no direcionamento do doador de vestígios biológicos encontrados no local de um crime, por exemplo.

Além da importância dos estudos de genética populacional destinados à identificação humana na prática forense, alguns marcadores têm sido utilizados em genética clínica e na avaliação da constituição genética para fins históricos da população atual.

UTILIZAÇÃO DE MARCADORES GENÉTICOS EM ESTUDOS DE GENÉTICA CLÍNICA

A genotipagem a partir de marcadores informativos de ancestralidade, capazes de gerar estimativas de componentes ancestrais, vem apresentando uma grande importância na detecção de possíveis efeitos de subestruturação populacional, assim como na correção de diferenças individuais ao calcular parâmetros específicos para estudos de associação genética (da Silva *et al.*, 2011; Pereira *et al.*, 2012).

Quando há diferenças entre grupos de indivíduos em uma população, diz-se que ela está estratificada ou subdividida. As estimativas de origem ancestral podem ajudar na correção de falsos resultados positivos devido à estratificação

populacional entre indivíduos “casos e controles” em estudos de associação genética. Esses estudos são amplamente aplicados para identificar a base genética de características quantitativas, como a suscetibilidade às doenças complexas, o que pode possibilitar a detecção de efeitos genéticos de pequenos a moderados e possíveis interações com os diversos fatores ambientais envolvidos (Tsai *et al.*, 2006; Tian *et al.*, 2008). Ademais, a utilização dos AIMS para a obtenção de estimativas de ancestralidade pode ser usada como uma ferramenta inicial, por serem econômicas em uma pré-avaliação de indivíduos relacionados a estudo de doenças genéticas (Pereira *et al.*, 2012).

Sabe-se que uma série de ensaios clínicos (risco e/ou predisposição a doenças e respostas a tratamentos) apresentam diferenças de acordo com a etnia pré-estabelecida. Para uma população heterogênea como a Brasileira, o correto conhecimento da miscigenação existente é extremamente importante para implicações clínicas, possibilitando um melhor entendimento e análise dos dados em genética clínica e medicina genômica (Kittles & Weiss, 2003).

VARIABILIDADE GENÉTICA DE POPULAÇÕES HUMANAS

A população humana tem como principal característica uma significativa diferenciação entre os indivíduos, reflexo da variabilidade genética verificada entre os seres humanos, e moldada por mutações, e outras forças evolutivas, como: a migração, seleção natural, deriva genética e recombinação gênica (Jobling *et al.*, 2004).

A atual diversidade genética da população humana contém informações sobre parâmetros populacionais do passado e de sua história de adaptação às constantes variações do ambiente. Algumas análises da variação genética contemporânea têm sido realizadas para traçar aspectos do passado dos seres humanos. Combinações de padrões moleculares devem ser utilizadas para validação das afirmações indicadas pelos resultados obtidos a partir de diferentes marcadores genéticos. Deste modo, podemos descrever a distribuição da diversidade genética atual entre as populações humanas ou ainda fazer inferências de como esta evoluiu (Nei, 1972).

O aprofundamento do conhecimento do processo evolutivo é o cerne de todas as pesquisas no ramo da Genética de Populações humana. Desta forma, os avanços científicos trouxeram a possibilidade de avaliação e estudo de uma das questões mais pertinentes do homem moderno, a de confirmar a história de sua dispersão pela Terra (Figura 4) (Cavalli- Sforza *et al.*,1994).

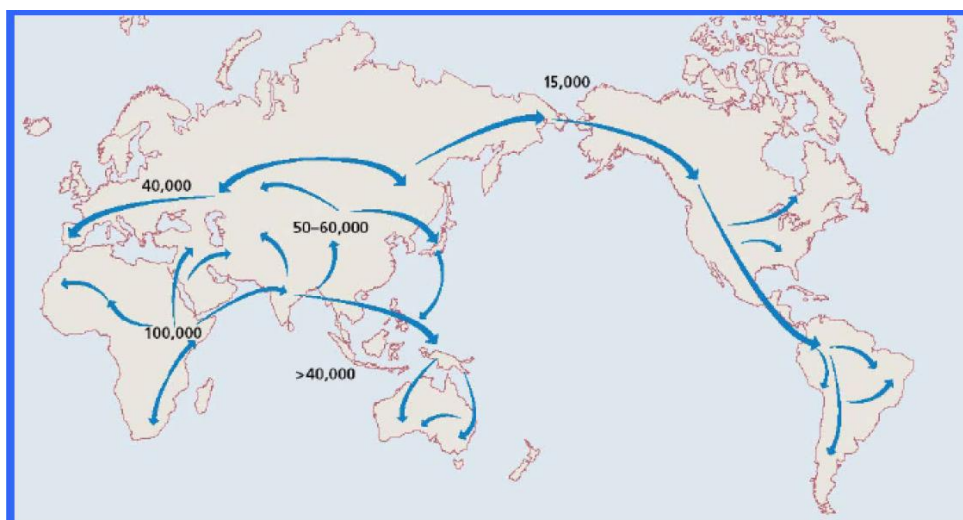


Figura 4: As migrações históricas no mundo. A diáspora do homem moderno a partir da África para Ásia, Europa e Oceania e depois para as Américas. Os números correspondem à estimativa em anos determinada principalmente por fósseis encontrados em diferentes lugares. Fonte: Cavalli- Sforza *et al.*,1994

As migrações humanas por diferentes regiões continentais produziram marcas genéticas, possíveis de serem confirmadas atualmente através da diferenciação das frequências alélicas. Estudos envolvendo populações humanas utilizam marcadores com escalas temporais de alguns milhares de anos para avaliar eventos históricos recentes, de 50.000 anos para avaliar a expansão do *Homo sapiens* para fora da África e de 100.000 a 200.000 anos para estudos sobre o surgimento do homem moderno. A partir de estudos filogeográficos e do mapeamento das frequências dos marcadores genéticos, pode ser criado um quadro que sugere por onde e quando os ancestrais do homem moderno se deslocaram (Campbell & Tishkoff, 2008)

A diversidade e a estrutura genética das populações podem ser investigadas com o objetivo de avaliar a variabilidade genética dentro e entre as populações.

Além de indicar o grau de polimorfismo dos marcadores, tais estudos possibilitam avaliar o fluxo gênico, o grau de intercruzamento, extinção e migração das populações, possibilitando, portanto, inferir fatores que podem estar relacionados com o desvio do Equilíbrio de Hardy-Weinberg.

A maioria da variação vista hoje originou-se no período de tempo no qual os humanos distribuíram-se ao longo do Velho Mundo. O aumento do tamanho populacional foi motivado pela agricultura ou outros fatores que propiciaram a fixação das sociedades. Aproximadamente, 90% do total da variação genética detectada ocorrem entre indivíduos de um mesmo continente e apenas 10% entre populações de continentes diferentes. Portanto, o padrão de compartilhamento da variação genética entre as populações tem implicações importantes para o entendimento sobre as diferenças e similaridades entre elas (Jorde & Wooding, 2004).

As populações tendem a não serem "puras" em um sentido genético. A definição de limites entre indivíduos ou populações pode ser inadequada, pois tende a formar agrupamentos de acordo com a distância geográfica que há entre elas (Nei & Roychoudhury, 1973). As populações da África são as que mais divergem, e a maior distância genética é vista entre essas e as não-africanas.

A população alvo de nosso estudo é a população brasileira, uma das mais heterogêneas do mundo.

UTILIZAÇÃO DE MARCADORES GENÉTICOS NA RECUPERAÇÃO DA HISTÓRIA DE POPULAÇÕES BRASILEIRAS

A formação da população brasileira através de marcadores genéticos vem sendo estudada há mais de duas décadas. Nesse sentido, os polimorfismos de DNA presentes nos cromossomos autossômicos, sexuais e no DNA mitocondrial, possuem destaque em estudos relacionados à estruturação gênica e à origem do brasileiro (Alves-Silva *et al.*, 2000; Góes *et al.*, 2002; Góes *et al.*, 2004; Silva *et al.*, 2006; Domingues *et al.*, 2007; Manta *et al.*, 2012).

Até pouco tempo, os estudos de Genética Populacional eram exclusivamente realizados através da análise de STRs e SNPs. A partir de 2006, os primeiros

estudos de ancestralidade baseados em marcadores Indels autossômicos foram tomando espaço (Bastos-Rodrigues *et al.*, 2006; Santos *et al.*, 2010; Pena *et al.*, 2011). Desde então, estudos acerca das proporções de genes herdados de cada um dos três principais povos ancestrais (europeus, africanos e ameríndios) na formação das diferentes regiões do Brasil vêm sendo realizados com o intuito de responder questões relacionadas à ancestralidade e caminhos evolucionários (Pereira & Pena, 2006)

Eventos evolutivos e relações interpopulacionais são exemplos de informações que podem ser obtidas a partir de dados de frequência alélica (Hartl e Clark, 1997). Nesta perspectiva, utilizar marcadores genéticos que, por suas características e propriedades, possibilitam caracterizar a identidade ou história evolutiva de um indivíduo ganha várias e importantes abordagens, entre elas a caracterização de populações miscigenadas.

A miscigenação avaliada através de diferentes marcadores vem mostrando um processo tendencioso envolvendo predominantemente descendentes de homens europeus e de mulheres africanas e ameríndias. Tais fatos permitem corroborar e afirmar que a população brasileira realmente é resultado de cinco séculos de mistura étnica entre populações de três continentes: a europeia, a africana e a ameríndia (Alves-Silva *et al.*, 2000; Pena *et al.*, 2000; Santos *et al.*, 2010; Pena *et al.*, 2011).

CONSTITUIÇÃO ÉTNICA DA POPULAÇÃO BRASILEIRA

Os índios brasileiros, os primeiros habitantes, são descendentes do primeiro grupo de caçadores asiáticos que chegaram às Américas, provavelmente através do estreito de Bering, durante o período glacial, há 25.000 anos ou mais, até chegarem ao Alasca. Espalharam-se pelas Américas desde o extremo norte até o extremo sul (Cunha, 1992). Esta população que originalmente saiu da Ásia, sofreu uma redução drástica no seu tamanho populacional diminuindo a variabilidade

genética existente nestes primeiros povos que migraram para o continente americano (Santos *et al.*, 1999).

Estima-se que a chegada dos ameríndios ao território conhecido hoje como Brasil tenha ocorrido por volta de 12.000 anos atrás. Não é possível uma determinação precisa do tamanho da população que habitava este território antes da chegada dos portugueses, devido à precariedade de dados históricos nesse período (Salzano e Callegari-Jacques, 1988). Alguns autores estimam que à época da chegada dos portugueses, o Brasil era habitado por cerca de 2,4 milhões de ameríndios (IBGE, 2000) e muitos afirmam que entre 1 e 10 milhões de indivíduos que habitavam a região eram, sobretudo, pertencentes aos grupos Tupi e Jê ou Tapuia (Fundação Nacional do Índio - FUNAI. www.funai.gov.br).

A partir da colonização, muitas sociedades indígenas que viviam no território dominado foram sendo exterminadas pela ação de armas, em decorrência do contágio por doenças trazidas dos países distantes ou, ainda, pela aplicação de políticas visando a “assimilação” dos índios à nova sociedade implantada, com forte influência europeia (FUNAI, 2012).

Atualmente os índios sobrevivem, não apenas biologicamente, mas também do ponto de vista das tradições culturais, segundo dados do Instituto Brasileiro de geografia e Estatística (IBGE), os quais demonstram que a população indígena apresenta hoje cerca de 220 diferentes povos que somam mais de 800 mil pessoas, que falam 180 línguas distintas. Os índios vivem nas mais diversas localidades do território brasileiro e representam, em termos demográficos, um pequeno percentual da população de mais de 190 milhões de habitantes no Brasil. Todavia, é um exemplo concreto e significativo da grande diversidade cultural existente no País (IBGE, 2010; FUNAI, 2012).

Os portugueses chegaram ao litoral brasileiro oficialmente em 22 de abril de 1500 e já trouxeram séculos de interação genética e cultural com outros povos de origens geográficas distintas. Estudos indicam que os primeiros brasileiros foram resultados do “casamento” entre homens portugueses e mulheres ameríndias (Carvalho-Silva *et al.*, 2001).

Como os portugueses já utilizavam mão de obra escrava africana supõe-se que os primeiros escravos brasileiros tenham vindo junto com a esquadra de Cabral. Os primeiros navios negreiros chegaram ao Brasil em meados do século XVI trazendo escravos para trabalhar na agricultura e na mineração. Registros históricos sugerem que entre os séculos XVI e XIX, o Brasil foi o país para o qual foi trazido maior número de escravos africanos, aportando cerca de 4 milhões de indivíduos, o equivalente a mais de um terço de todo o comércio negreiro (IBGE, 2000).

Entre 1500 e 1978, o Brasil também recebeu outros imigrantes, como os alemães, espanhóis, italianos, sírio-libaneses e japoneses também que contribuíram para a heterogeneidade da população brasileira, intensificada e continua até os dias atuais. Estima-se que 58% dos imigrantes foram europeus, 40% africanos e 2 % asiáticos (Callegari-Jacques & Salzano, 1999; Santos *et al.*, 2009). Por fim, houve também um processo de incorporação de várias outras culturas à brasileira, e, conseqüentemente suas características genotípicas e fenotípicas também foram adicionadas à nossa população.

O primeiro estudo sobre a constituição étnica da população brasileira data de 1872, ainda durante o império, e informa que ela seria composta por 2 milhões de negros, 6,5 milhões de brancos, 200 mil índios e 6 milhões de indivíduos miscigenados (Salzano & Callegari-Jacques, 1988). A distribuição de grupos étnicos baseada no auto-relato da cor de pele é heterogênea entre as regiões geopolíticas brasileiras. De acordo com a Pesquisa Nacional por Amostra Domiciliar (PNAD) realizada em 2010 pelo IBGE, cerca de 50% dos mais de 180 milhões de brasileiros deve ter indicações morfológicas de contribuição africana (autodefinidos como negros ou pardos, de acordo com a nomenclatura utilizada no levantamento).

Devido a sua grande extensão territorial, a distribuição destes grupos no território brasileiro não foi homogênea e reflete-se na composição da população atual, sendo a contribuição dos ameríndios para o *pool* gênico maior no norte, enquanto no nordeste se observa uma maior contribuição africana. De acordo com o IBGE (2010), a população branca está mais concentrada no sul do país, diminuindo à medida que avançamos para as regiões norte e nordeste. A maioria da população negra está localizada nas regiões nordeste e sudeste, apresentando

um menor percentual na região sul. A região norte é marcada pela maior presença relativa de ameríndios, porém nota-se uma ligeira queda na medida em que rumamos em direção à região sul, passando pelas regiões centro-oeste e nordeste, chegando, por fim, as regiões sudeste e sul, onde se apresentam os menores índices (IBGE, 2010).

JUSTIFICATIVA

Até pouco tempo, os estudos de genética populacional e forense eram exclusivamente realizados com o emprego de marcadores polimórficos de DNA, notadamente através da análise de STR ou SNPs em cromossomos autossômicos, sexuais masculinos e femininos e do DNA mitocondrial. Recentemente, grupos de pesquisa da área de genética forense têm juntado esforços na seleção de novos marcadores que permitissem a identificação tanto a nível individual como a nível populacional. Nesse contexto surgiram os marcadores Indels, capazes de discriminar tanto a diversidade interindividual quanto a existente entre populações de origens geográficas distintas.

Com a finalidade de contribuir com o avanço da aplicação de marcadores Indels na área da genética forense no Brasil, utilizamos na primeira etapa deste trabalho um multiplex constituído por 38 mini-Indels bialélicos, que já havia apresentado alta eficiência na discriminação de indivíduos de diferentes amostras populacionais, entre elas: Portugal, Angola, Moçambique, Macau, Taiwan (Pereira *et al.*, 2009) e Argentina (Vullo *et al.*, 2011), mas que até o momento não tinha sido analisado em nenhuma população altamente miscigenada como a brasileira.

Em relação a análises de ancestralidade, ainda são escassos os estudos que abordam a relação entre a composição genética de uma população altamente miscigenada e com características diferenciadas a nível nacional, regional e também interindividual. Nesse contexto, os polimorfismos autossômicos de inserção e deleção podem complementar, ou mesmo substituir a utilização dos marcadores de linhagem, que nos informam somente um lado da história. Para isso, utilizamos o

sistema multiplex 46-AI-Indels, constituído por 46 marcadores informativos de ancestralidade, na genotipagem de diferentes amostras populacionais brasileiras.

Vale ressaltar, que resultados obtidos a partir de uma população altamente miscigenada como a brasileira são de relevância no processo de validação e na comprovação do poder de discriminação de novos sistemas multiplexes de tipagem humana, bem como na avaliação da utilidade desses marcadores em genética de populações e na forense.

1 OBJETIVOS

1.1 Objetivo Geral

Desenvolver estudos de Genética Populacional e Forense através de regiões polimórficas do tipo Indels.

1.2 Objetivos Específicos

- Avaliar o potencial do Multiplex Indelplex-HID formado por 38 marcadores Indels autossômicos altamente informativos para aplicações forenses, destinado à identificação humana e relações de parentesco, em diferentes amostras da população Brasileira, através da (o):

- Determinação das frequências de alelos Indels em uma amostra populacional miscigenada (estado do Rio de Janeiro) e em uma população indígena (Índios Terena - Mato Grosso do Sul).

- Construção de bases de dados de frequências dos polimorfismos dos marcadores analisados, com aplicação na área de identificação humana.

- Cálculo de parâmetros estatísticos e de diversidade genética com o propósito de avaliar o poder do sistema para uso forense e em investigação de vínculos genéticos.

- Avaliar o potencial do Multiplex 46AIM-Indel formado por 46 marcadores Indels informativos de ancestralidade, em diferentes populações brasileiras, através da(o):

- Determinação das frequências alélicas obtidas para os marcadores AIM-Indels e construção de banco de dados dessas frequências para amostras populacionais de diferentes regiões brasileiras. Sudeste (Rio de Janeiro, Espírito Santo, Minas Gerais e São Paulo), Norte (Manaus-Amazonas, Santa Isabel do Rio

Negro- Amazonas), Nordeste (Alagoas, Pernambuco) Centro-oeste (Mato Grosso do Sul, índios Terena - Mato Grosso do Sul), Sul (Santa Catarina, Paraná e Rio Grande do Sul) como subsídios para estudos populacionais e forenses;

- Identificação da contribuição de ancestralidade que cada indivíduo possui de linhagens europeias, africanas e ameríndias através dos marcadores Indels autossômicos;

- Emprego de polimorfismos do tipo Indel autossômicos como marcadores de ancestralidade, comparando os resultados obtidos com aqueles observados quando da utilização de outros marcadores de DNA, como o mtDNA e os Y-SNPs para a população do Rio de Janeiro;

- Investigação de populações de diferentes estados brasileiros com a finalidade de verificar a influência das migrações recentes bem como investigar a contribuição genética dos principais grupos étnicos parentais na formação da população brasileira.

2 MATERIAL E MÉTODOS

2.1 Amostras biológicas

As amostras populacionais utilizadas no presente estudo consistiam de indivíduos não relacionados geneticamente, residentes nos estados do Rio de Janeiro, São Paulo, Minas Gerais, Espírito Santo, Rio Grande do Sul, Santa Catarina, Paraná, Mato Grosso do Sul, Pernambuco, Alagoas, Amazonas, assim como, de indivíduos da tribo Terena, localizada no Mato Grosso do Sul e de indivíduos do município de Santa Isabel do Rio Negro, localizado na Amazônia. O número total de indivíduos contidos em cada amostra populacional, utilizados nas diferentes etapas do trabalho, são apresentados na tabela 1.

Todos os indivíduos relacionados consentiram a utilização de suas amostras biológicas, de forma não identificada, para a realização de estudos científicos (ANEXO A). As amostras biológicas dos índios Terena também foram obtidas por doação voluntária, de forma livre e esclarecida, e aprovada para fins de pesquisa no Comitê de Ética em Pesquisa da Universidade Federal do Mato Grosso do Sul (CI nº 015/98), tendo sido o ingresso na área indígena permitido pelo FUNAI, segundo a autorização nº016/CGEP/99.

Este estudo foi aprovado pelo Comitê de Ética da Universidade do Estado do Rio de Janeiro (CAAE: 0067.0.228.000-09) (ANEXO B).

As amostras de sangue periférico dos indivíduos citados acima foram coletadas através do sistema Vacutainer contendo EDTA como anticoagulante e mantidas a 4° C até a extração do DNA.

Para o estudo relacionado com a genotipagem de amostras degradadas foram selecionadas amostras de indivíduos *post mortem* (fêmur e dentes) provenientes das análises de casos criminais realizados no LDD-UERJ.

2.1.1 Detalhamento das amostras populacionais

A amostragem do Rio de Janeiro, São Paulo, Minas Gerais, Rio Grande do Sul, Santa Catarina e Paraná se originaram de perícias realizadas no Laboratório de Diagnósticos por DNA (LDD) da Universidade do Estado do Rio de Janeiro (UERJ).

As amostras de Pernambuco foram coletadas no Laboratório de Identificação Humana da Universidade de Pernambuco a partir de casos de investigação de paternidade. As amostras do Espírito Santo e Manaus foram obtidas a partir de estudantes e professores da Universidade Federal do Espírito Santo e do Amazonas, respectivamente. As amostras do Mato Grosso do Sul foram obtidas a partir de estudantes da Escola Estadual João Ponce de Arruda. As amostras de Alagoas e de Santa Isabel do Rio Negro foram selecionadas a partir de projetos pré-existentes de genética clínica de grupos colaboradores da Universidade Federal de Alagoas e do Instituto de Pesquisa Fundação Oswaldo Cruz (FIOCRUZ), respectivamente.

Em função do objetivo do nosso trabalho, as amostras da população do Rio de Janeiro foram divididas em dois subgrupos: um formado apenas por indivíduos que se autodenominaram afrodescendentes e o segundo por indivíduos selecionados ao acaso no banco de dados do LDD-UERJ, representando a população geral do Rio de Janeiro.

Tabela 1 - Amostras biológicas utilizadas nas diferentes etapas do estudo.

Estudo	Procedência	Número de amostras
Identificação Humana	Rio de Janeiro - população geral	137
Identificação Humana	Rio de Janeiro- Indivíduos autodeclarados afrodescendentes	115
Identificação Humana	Índios Terena (Mato Grosso do Sul)	45
Identificação Humana	Amostras post mortem	2
Ancestralidade Genética	Manaus	42
Ancestralidade Genética	Santa Isabel do Rio Negro	100
Ancestralidade Genética	Pernambuco	133
Ancestralidade Genética	Alagoas	104
Ancestralidade Genética	Mato Grosso do Sul	84
Ancestralidade Genética	Índios Terena (Mato Grosso do Sul)	42
Ancestralidade Genética	Minas Gerais	88
Ancestralidade Genética	Espírito Santo	92
Ancestralidade Genética	São Paulo	49
Ancestralidade Genética	Rio de Janeiro - população geral	280
Ancestralidade Genética	Rio de Janeiro- Indivíduos autodeclarados afrodescendentes	113
Ancestralidade Genética	Paraná	21
Ancestralidade Genética	Santa Catarina	20
Ancestralidade Genética	Rio Grande do Sul	23

Nota: Para o estudo de Identificação Humana foi utilizado o sistema HID-Indelplex e para o estudo de ancestralidade utilizou-se o sistema 46-AI-Indels.

2.2 Extração de DNA

2.2.1 Extração por Salting-out (Miller, 1988)

À uma alíquota de 2 mL de sangue foram adicionados 10 mL de tampão de lise de hemácias (TLH), homogeneizado manualmente por inversão e centrifugado por 5 min a 1000 x g (Centrifuga Excelsa, modelo 206BL). Em seguida, o sobrenadante foi descartado. Ao precipitado de células foram adicionados mais 10 mL de tampão TLH, repetindo-se o processo inicial. Após a centrifugação e remoção do sobrenadante, o precipitado de células foi ressuspensionado em 500 µL de tampão de lise de membrana nuclear (NLB), seguido da adição de 15 µL de Dodecil Sulfato de Sódio (SDS) 10 % e 40 µL de proteinase 20 mg/mL e homogeneizado manualmente, permanecendo a 37°C durante a noite ou por 2 horas a 56°C. Após a incubação, foram adicionados 150 µL de NaCl 6 M, sendo agitado vigorosamente, por aproximadamente 30 seg. O material foi centrifugado por 10 min a 3000 x g e o sobrenadante transferido para um tubo tipo falcon. Ao sobrenadante foram adicionados 2,5 mL de etanol absoluto com agitação em movimentos circulares para precipitação dos ácidos nucleicos. O álcool foi descartado e o precipitado ressuspensionado em volumes entre 100 a 500 µL de tampão Tris-EDTA (TE), dependendo da quantidade de *pellet* de DNA obtido.

2.2.2 Extração por Chelex (Walsh et al., 1991)

Em tubos de microcentrífuga foram adicionados 1 mL de água milliQ autoclavada para cada 10 µL de amostra de sangue. Após 30 min à temperatura ambiente, o material foi centrifugado por 5 min a 1.200 x g (Centrifuga Centrimicro 242). O sobrenadante foi descartado, o precipitado ressuspensionado em 0,5 mL de água milliQ autoclavada, sendo deixado novamente por 30 min à temperatura ambiente. O material foi novamente centrifugado por 5 min a 13000 x g e ao

precipitado foram adicionados 300 µl de solução de Chelex-100 a 5%, permanecendo então em banho-maria a 56°C por 20 min e posteriormente a 100°C por 8 min. Após a segunda incubação, foi acrescentado 1 mL de TE sendo centrifugado por 3 min à 13000 x g, para posterior armazenamento a 4°C .

2.2.3 Extração por método semi-automatizado. Maxwell 16® (Promega Corporation)

Esta metodologia de extração de DNA foi utilizada somente para as amostras de restos mortais (dentes e ossos). Após a limpeza do material com água deionizada autoclavada, o material foi centrifugado por 10 minutos a 4000 rpm descartando-se o sobrenadante. Em seguida foram transferidos 100 mg do pó de osso ou fragmentos de dentes, para um tubo cônico de 15 ml juntamente com 944 µl de solução *bone incubation* (promega) e 56 µl de proteinase K (18 mg/ml). Posteriormente a solução foi homogeneizada vigorosamente durante 10 segundos e deixada em repouso a 56°C por 16 horas. Após a incubação, a solução foi centrifugada a 8000 rpm por 10 min, em seguida, transferiu-se 1 mL do sobrenadante para um novo tubo cônico de 15 mL, onde foram adicionados 700 µl de tampão de lise e 7 µl de DTT. Por último, foram transferidos 750µL da solução para o sistema Maxwell 16®, segundo protocolo específico do aparelho (Promega Corporation).

2.3 Genotipagem

A escolha dos marcadores Indels autossômicos bialélicos utilizados neste trabalho relaciona-se com sua elevada frequência no genoma humano e por apresentarem distintas vantagens em estudos populacionais e forenses, tais como a baixa taxa de mutação, apresentação de *amplicons* curtos (com 50 pares de base ou menos), bem como menor risco de geração de falsas interpretações, observadas frequentemente em sistemas multialélicos.

2.3.1 HID-Indelplex (Pereira *et al.*, 2009).

Para fins forense e de vínculos genéticos foi utilizado um multiplex formado por 38 marcadores Indels autossômicos, não codificantes, bialélicos, que apresentaram elevada diversidade, frequência alélica mínima ≥ 0.25 e heterozigosidade ≥ 0.40 em todos os grupos populacionais que foram selecionados com a finalidade de permitir a maior discriminação possível entre indivíduos de toda e qualquer população.

Para a amplificação por PCR dos *loci* alvo reunidos no sistema HID-Indelplex, foi utilizado o QIAGEN® Multiplex PCR Kit, com tampão QIAGEN® multiplex PCR master mix 1x, 1 μ L do *primer* mix e de 1ng DNA genômico em um volume final de 10 μ L. As condições termocíclicas utilizadas nas reações de amplificação foram: incubação inicial de 95°C por 15 min; 10 ciclos de 94°C por 30 seg, 60°C por 90 seg e 72°C por 60 seg; 20 ciclos de 94°C por 30 seg, 58°C por 90 seg e 72°C por 60 seg; com uma extensão final de 72°C por 60 min. As reações foram procedidas no termociclador Veriti (*Life Technologies*).

Os marcadores amplificados neste sistema multiplex, as sequências alélicas, as respectivas localizações no genoma e o tamanho esperado do *amplicon* para os alelos curtos e longos são mostrados na tabela 2.

Tabela 2: HID-Indels utilizados neste estudo incluindo a localização de cada marcador no genoma, os alelos relatados e os tamanhos esperados dos *amplicons*.

Código rs	Cromossomo	Posição (bp)	Código e posição (bp)	Alelos	Tamanho esperado do <i>Amplicon</i> (S-L)
rs3047269	1	161077452	NT_004487.18 pos. 13301183	-/CTGA	126-130
rs2307579	1	245878706	NT_004836.17 pos. 12569872	-/ATG	104-107
rs16624	2	234681130	NT_005120.15 pos. 949145	-/GT	65-67
rs2308242	3	8591709	NT_022517.17 pos. 8556709	-/CT	106-108
rs2308026	4	119404855	NT_016354.18 pos. 43733552	-/CA	83-85
rs2307526	5	5178112	NT_006576.15 pos. 5115112	-/ACAC	95-99
rs1160956	5	65414216	NT_006713.14 pos. 15972818	-/AGA	128-131
rs1610871	5	171020572	NT_023133.12 pos. 15897552	-/TAGG	61-65
rs2307710	6	47929222	NT_007592.14 pos. 38679494	-/AGGA	92-96
rs2307839	6	117200251	NT_025741.14 pos. 21262987	-/GA	152-154
rs2308137	6	149655891	NT_025741.14 pos. 53718627	-/GA	61-63
rs2307978	7	83121850	NT_007933.14 pos. 8518190	-/GA	156-158
rs35769550	8	76681235	NT_008183.18 pos. 28372034	-/TGAC	89-93
rs5895447	8	138489776	NT_008046.15 pos. 51638773	-/CA	128-130
rs16402	9	38396788	NT_008413.17 pos. 38396788	-/TTAT	150-154
rs2067294	9	70504241	NT_023935.17 pos. 478953	-/CTT	80-83
rs2307580	9	104626014	NT_008470.18 pos. 12907398	-/AATT	120-124
rs140809	10	6027167	NT_077569.2 pos. 350057	-/CAA	115-118
rs1160886	10	54112392	NT_008583.16 pos. 2993541	-/ACT	75-78
rs10688868	11	258180	NT_035113.6 pos. 208180	-/CT	81-83
rs34811743	11	30134266	NT_009237.17 pos. 28964931	-/TG	108-110
rs33972805	11	125794082	NT_033899.7 pos. 29851288	-/CT	135-137
rs1610919	12	14801263	NT_009714.16 pos. 7668970	-/AT	142-144
rs2067238	12	113772931	NT_009775.16 pos. 5858057	-/GCT	71-74
rs2308171	13	43778155	NT_024524.13 pos. 25860155	-/TCTG	135-139
rs2308189	14	28106508	NT_026437.11 pos. 10036508	-/AACTA	119-124
rs2308020	15	51268809	NT_010194.16 pos. 24272074	-/TT	127-129
rs2067208	16	83139788	NT_010498.15 pos. 38196486	-/GCCAG	93-98
rs3051300	17	10076666	NT_010718.15 pos. 9733290	-/GTAT	63-67
rs3080855	18	21507205	NT_010966.13 pos. 4742309	-/AATT	133-137
rs34511541	18	34677042	NT_010966.13 pos. 17912146	-/CTCTT	143-148
rs36040336	19	1353662	NT_011255.14 pos. 1342662	-/AT	65-67
rs2307689	19	48896180	NT_011109.15 pos. 16472558	-/TTC	74-77
rs33917182	20	11643625	NT_011387.8 pos. 11635625	-/CA	142-144
rs34541393	20	30165066	NT_028392.5 pos. 897497	-/AACT	57-61
rs35605984	21	14556736	NT_011512.10 pos. 1296736	-/TAAAG	151-156
rs10629077	21	30294208	NT_011512.10 pos. 17034208	-/AT	74-76
rs2307700	22	25120901	NT_011520.11 pos. 6181470	-/TCAC	101-105

Legenda- Código rs: Listagem dos marcadores de acordo com dbSNP – 129; S=*short* = alelos curtos e L=*long* =alelo longos.

2.3.2 46-AIM Indels (Pereira *et al.*, 2011)

Com o propósito de inferir a origem ancestral de cada indivíduo analisado nesse estudo e estimar suas proporções a nível populacional para diferentes amostras da população brasileira foi utilizado o sistema multiplex AIMs composto por 46 marcadores cujas frequências alélicas apresentam diferenças significativas entre europeus, africanos, nativos americanos e asiáticos, o que os tornam altamente informativos para cada um dos quatro grupos populacionais.

O sistema 46-AIM-Indels permite medir eficientemente proporções de mistura populacional de quatro origens continentais diferentes (África, Europa, Ásia Oriental e América) em uma só reação (Pereira *et al.*, 2012).

Para a PCR utilizando o multiplex AIM-Indels, usamos o QIAGEN Multiplex PCR kit (Qiagen). A mistura de reação foi composta por master mix 1X, 1µL do *primer* mix e 1-2ng de DNA genômico em uma reação de 10µL de volume final. As condições termocíclicas foram: etapa inicial de 95°C por 15 min; 30 ciclos de 94°C por 30 seg, 60°C por 90 seg, e 72°C por 45 seg; e 72°C por 60 min de extensão final. As PCRs foram desenvolvidas no termociclador Veriti (*Life Technologies*).

A listagem dos marcadores amplificados neste multiplex, a descrição dos alelos, bem como, as localizações físicas são mostradas na tabela 3.

Tabela 3: AIM-Indels utilizados neste estudo, incluindo a localização de cada marcador no genoma e os alelos relatados.

Código rs	MID	Cromossomo	Posição(bp)	Alelos descritos no dbSNP
rs2307666	MID-1470	11	64729920	-/GTTAC
rs1610863	MID-777	16	6551830	-/GAA
rs16635	MID-196	6	99789775	-/CAT
rs1610965	MID-881	5	79746093	-/ACTT
rs35451359	MID-3122	18	45110983	-/ATCT
rs140837	MID-548	6	3708909	-/CT
rs1160893	MID-659	2	224794577	-/CT
rs2308203	MID-2011	2	109401291	-/CTAGA
rs33974167	MID-2929	8	87813725	-/TA
rs1160852	MID-593	6	137345857	-/TT
rs1610884	MID-798	5	56122323	-/GGAAA
rs2067280	MID-1193	5	89818959	-/AT
rs2308067	MID-1871	7	127291541	-/TT
rs4183	MID-17	3	3192524	-/TAAC
rs3054057	MID-2538	15	86010538	-/AACA
rs2307840	MID-1644	1	36099090	-/GT
rs60612424	MID-3854	6	84017514	-/TCTA
rs3033053	MID-2275	14	42554496	-/TCAGCAG
rs16384	MID-94	22	42045009	-/AAC
rs34611875	MID-3072	18	67623917	-/GCCCCCA
rs1610859	MID-772	5	128317275	-/TAG
rs3045215	MID-2313	1	234740917	-/ATTATAACT
rs25621	MID-397	6	139858158	-/TTCT
rs2307832	MID-1636	1	55590789	-/AA
rs16343	MID-51	4	17635560	-/TTTAT
rs3031979	MID-2431	8	73501951	-/ATTG
rs34122827	MID-2264	13	63778778	-/AAGT
rs133052	MID-2256	22	41042364	-/CAT
rs6490	MID-128	12	108127168	-/ATT
rs4181	MID-15	2	42577803	-/AAATACACAC
rs3030826	MID-2241	6	67176774	-/GTCCAATA
rs140708	MID-419	6	170720016	-/AATGGCA
rs1611026	MID-943	5	82545545	-/TGAT
rs16438	MID-159	20	25278470	-/CCCCA
rs2308161	MID-2005	10	69800909	-/AACAAT
rs16687	MID-250	7	83887882	-/CA
rs2307998	MID-1802	5	7814345	-/GGA
rs2307803	MID-1607	3	108981031	-/TG
rs2307930	MID-1734	6	84476378	-/CCAT
rs25630	MID-406	6	14734341	-/AG
rs2307582	MID-1386	1	247768775	-/AAACTATTCATTTTTACCCCT
rs2307922	MID-1726	1	39896964	-/CAAGAACTATAAT/CACTATCTATTAT
rs11267926	MID-3626	15	45526069	-/AATATAATTCTCCA
rs25584	MID-360	12	112145217	-/AA
rs2307799	MID-1603	5	70828427	-/TTGT
rs34541393	MID-2719	20	30701405	-/AACT

Legenda- Código rs: Listagem dos marcadores de acordo com dbSNP -132; MID: Nomenclatura dos marcadores de acordo com o banco de dados Marshfield Diallelic Insertion/Deletion Polymorphisms.

2.4 Detecções dos alelos

Os produtos amplificados foram submetidos à eletroforese capilar pelo sistema de análise molecular ABI3130 e ABI3500 (*Life Technologies*). A análise dos produtos amplificados e a nomeação dos alelos foram feitas com o software GeneMapper ID v3.2 ou GeneMapper ID v4.1, respectivamente (*Life Technologies*).

2.5 Análise estatística

Os cálculos das frequências alélicas observadas para os 38 marcadores de identificação e para os 46 marcadores de ancestralidade, assim como, as análises de diversidade genética (Heterozigosidade esperada - H_e), distância genética (F_{ST}), e o teste de Equilíbrio de Hardy–Weinberg foram realizadas através do *software* Arlequin versão 3.5 (Excoffier and Lischer 2010). A partir das frequências alélicas encontradas foram construídos bancos de dados de identificação e de ancestralidade.

Com o propósito de estimar a diferenciação genética entre as populações estudadas neste trabalho, os dados obtidos para as frequências dos marcadores informativos de ancestralidade foram comparados com os das seguintes populações: africanas, européias e ameríndias (Pereira *et al.*, 2012). O mesmo foi feito com os dados obtidos para as frequências dos marcadores Indels destinados à identificação humana. As amostras populacionais dos índios Terena, da população geral do Rio de Janeiro e dos indivíduos autodenominados afrodescendentes do Rio de Janeiro foram comparadas com as Africanas, Européias e Asiáticas (Pereira *et al.*, 2009).

Para obter uma representação gráfica das afinidades entre as distâncias (F_{ST}) das várias populações comparadas, foi realizada a análise do escalonamento multidimensional (MDS) das distâncias genéticas (F_{ST}) entre as diversas populações, utilizando o programa STATISTICA versão 8.0 (StatSoft, 2007; www.statsoft.com).

Para os cálculos estatísticos de interesse forense foram utilizados os bancos de dados de frequências alélicas, obtido a partir da genotipagem com o sistema

Indelplex HID, das amostras populacionais do Rio de Janeiro e dos índios Terena. Foram calculados através do programa Powerstats v1.2 (Promega Corporation), os seguintes parâmetros: Poder de Discriminação (PD), Poder de Exclusão (PE) e Probabilidade de Coincidência (MP, do inglês, *Matching Probability*).

As estimativas de ancestralidade individual e global foram realizadas no STRUCTURE v2.3.3 (Pritchard *et al.*, 2000), usando como amostras de referência as populações ancestrais africanas, européias e ameríndias do painel de diversidade HGDP-CEPH (H952) (Pereira *et al.*, 2012). Estes dados foram gerados apenas para as amostras genotipadas com os marcadores informativos de ancestralidade.

3 RESULTADOS

Optou-se em apresentar esta tese de doutoramento sob a forma de coletânea de artigos publicados e/ou submetidos à revistas científicas. Os resultados referentes aos objetivos dessa tese estão apresentados na forma de três artigos. As figuras e tabelas suplementares estão posicionadas nos ANEXOS.

Os três artigos tratam da análise da viabilidade de se associar polimorfismos do tipo Indels em estudos forenses e de genética populacional, a partir de diferentes amostras da população brasileira.

O primeiro artigo avaliou o multiplex destinado a Identificação Humana em diferentes tipos de amostras biológicas e em análises de vínculo genético. O segundo e o terceiro artigo foram realizados com base nos dados obtidos a partir da genotipagem de indivíduos de diferentes regiões brasileiras com marcadores Indels informativos de ancestralidade.

O primeiro artigo, publicado na *Forensic Science International: Genetics* em 2012, intitulado “Indel markers: Genetic diversity of 38 polymorphisms in Brazilian populations and application in a paternity investigation with post mortem material”, apresenta os resultados da investigação dos Indels-HID em três amostras da população brasileira: duas do estado do Rio de Janeiro (incluindo uma amostra de indivíduos autodeclarados afrodescendentes) e uma população ameríndia da tribo Terena do Mato Grosso do Sul.

O segundo artigo, publicado na revista *Annals of Human Biology* em 2013, denominado “Analysis of Genetic Ancestry in the Admixed Brazilian Population from Rio de Janeiro using 46 Autosomal Ancestry-Informative Indel Markers” apresenta os resultados das proporções de mistura interétnica encontrada em duas amostras populacionais do estado do Rio de Janeiro utilizando marcadores autossômicos informativos de ancestralidade.

Por fim, no terceiro artigo intitulado “Revisiting the genetic ancestry of Brazilians using autosomal AIM-Indels” submetido á revista *PLOS one* (ver ANEXO E), estão apresentados os resultados de ancestralidade genética de 798 indivíduos residentes em doze populações brasileiras, que representam as cinco macrorregiões geográficas do Brasil, genotipados pelo sistema 46 AI-Indels.

3.1 Artigo I:

Titulo do Artigo: **Indel markers: genetic diversity of 38 polymorphisms in Brazilian populations and application in a paternity investigation with post mortem material**

Publicação: Forensic Science International: Genetics. 2012; 6 (5): 658-661.
<http://dx.doi.org/10.1016/j.fsigen.2011.12.008>.

RESUMO:

Objetivando avaliar a utilidade de 38 Indels autossômicos bialélicos não-codificantes em identificação genética e em análises de parentesco, foram estudadas três amostras da população brasileira: duas do estado do Rio de Janeiro (incluindo uma amostra de indivíduos autodeclarados afrodescendentes) e uma população ameríndia da tribo Terena do Mato Grosso do Sul. Baseado nas frequências alélicas observadas, foram calculados os parâmetros de relevância forense. O poder de discriminação acumulado dos 38 Indels foi elevado em todos os grupos estudados ($PD \geq 0.999999999999997$), embora um pouco menor em nativos americanos. A análise de distância genética mostrou diferenças significativas entre as frequências alélicas da população do Rio de Janeiro e aquelas anteriormente relatadas para europeus, africanos e asiáticos, explicados por sua posição genética intermediária entre europeus e africanos. Como esperado, a amostra dos Terena foi significativamente diferente de todas as outras populações analisadas: brasileiros da população geral do Rio de Janeiro, autodeclarados afrodescendentes, europeus, africanos e asiáticos do leste. Finalmente, a eficiência do multiplex 38-Indelplex foi testada em material *post-mortem*, com resultados positivos, enfatizando o uso de marcadores bialélicos de *amplicons* curtos como uma ferramenta adicional para análise de STR, quando as moléculas de DNA estão degradadas.



Contents lists available at SciVerse ScienceDirect

Forensic Science International: Genetics

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

Short communication

Indel markers: Genetic diversity of 38 polymorphisms in Brazilian populations and application in a paternity investigation with post mortem material

Fernanda Manta^a, Alexandre Caiafa^a, Rui Pereira^b, Dayse Silva^a, António Amorim^{b,c}, Elizeu F. Carvalho^{a,*}, Leonor Gusmão^b^a DNA Diagnostic Laboratory – Institute of Biology, State University of Rio de Janeiro, Rio de Janeiro, Brazil^b IPATIMUP – Institute of Pathology and Molecular Immunology of the University of Porto, Porto, Portugal^c Faculty of Sciences of the University of Porto, Porto, Portugal

ARTICLE INFO

Article history:

Received 12 September 2011

Received in revised form 28 December 2011

Accepted 30 December 2011

Keywords:

Indel

Human identification

Post-mortem material

Brazil

Rio de Janeiro

Terena

ABSTRACT

Aiming to evaluate the usefulness of 38 non-coding bi-allelic autosomal indels in genetic identification and kinship testing, three Brazilian population samples were studied: two from Rio de Janeiro (including a sample of individuals with self-declared African ancestry) and one Native American population of Terena from Mato Grosso do Sul. Based on the observed allele frequencies, parameters of forensic relevance were calculated. The combined power of discrimination of the 38 indels was high in all studied groups ($PD \geq 0.999999999999997$), although slightly lower in Native Americans. Genetic distance analysis showed significant differences between the allele frequencies in the Rio de Janeiro population and those previously reported for Europeans, Africans and Asians explained by its intermediate position between Europeans and Africans. As expected, the Terena sample was significantly different from all the other populations: Brazilians from Rio de Janeiro general population and with self-declared African ancestry, Europeans, Africans and East Asians. Finally, the performance of the 38-indel multiplex assay was tested in post-mortem material with positive results, supporting the use of short amplicon bi-allelic markers as an additional tool to STR analysis when DNA molecules are degraded.

© 2012 Published by Elsevier Ireland Ltd.

1. Introduction

In human genetic identification, it is important to select highly polymorphic markers to reach an elevated discrimination power, even for a restricted number of loci. Nevertheless, very high levels of polymorphism can be associated with high mutation rates, which are a limitation in kinship analysis. Due to their high levels of variation, Short Tandem Repeats (STRs) are the most frequently used DNA markers in human identification. The universal forensic use of STRs, for almost two decades, has allowed the establishment of well structured national databases, simple methodologies, and to evaluate relevant statistical parameters. Despite the above advantages, most STRs present relatively large amplicon sizes, between 150 and 450 base pairs (bp), which can negatively influence the analysis of challenging forensic samples [1]. Limitations of sample quantity and/or degradation can also create problems in forensic casework, where DNA samples are usually far from pristine. In cases where the DNA is exposed to the

environment, degradation can occur due to bacterial, biochemical, or oxidative processes [2]. It has been demonstrated that smaller amplicons are much more likely to be amplified in samples containing degraded DNA [3].

In order to overcome the limitations of STRs outlined above, slow mutating bi-allelic markers analysed in shorter amplicons, including Single Nucleotide Polymorphisms (SNPs) and insertion/deletion polymorphisms (indels), have been recently introduced to the forensic identification field [4–8]. The lower mutation rate SNPs and indels present when compared to STRs contributes to their usefulness in paternity cases showing just a few observed incompatibilities [9–11]. Moreover, they have proved to be consistently better than STRs with conventional amplicon lengths for the analysis of degraded DNA samples [4,10,12].

In recent years, indels have received growing attention. After an extensive survey of ~2000 indels by Weber et al. [13], other studies aimed to adopt indels as a tool in human genetic studies. In 2006, Mills et al. [14] systematically identified indel polymorphisms using a new computational strategy, resulting in an initial insertion/deletion map containing more than 415,000 unique polymorphisms, with an average density of one indel per 7.2 kb. Recent follow-up studies have reported around 2 million small insertions/deletions ranging from 1 bp to 10,000 bp in length in the genomes of 79 diverse humans [15]. These small indels, involving

* Corresponding author at: Laboratório de Diagnóstico por DNA, São Francisco Xavier, 524 Maracanã CEP: 21550-013 Rio de Janeiro, RJ, Brazil.
Tel.: +55 21 23342183; fax: +55 21 23340594.

E-mail address: elizeutcf@hotmail.com (E.F. Carvalho).

apparently random DNA sequences, represent ~41% of all indels, and show a wide range of allele length variation, with the majority under 100 bp [14,15].

In 2009, Pereira et al. [5] described a new multiplex for human identification combining 38 small non-coding biallelic autosomal indels into a single multiplex. This multiplex benefits from using a single PCR where the products are then detected directly by standard Capillary Electrophoresis (CE) analysers. Furthermore, amplicon length was designed to be shorter than a maximum of 160 bp. The same study also demonstrated the feasibility of obtaining complete profiles using 0.3 ng of DNA, while successful genotyping of degraded samples was possible in cases where standard STR typing had partially failed [5].

The aim of the present work was to evaluate the usefulness of the 38-indel multiplex [5] applied to Brazilian forensic casework. We have studied three populations from Brazil, a sample of Native Americans from a Terena tribe living in Mato Grosso do Sul, plus two samples from individuals living in Rio de Janeiro; one selected from the general population and the other including self-declared African admixed individuals. These samples were typed for the 38 autosomal indels, and forensic informativeness parameters were calculated. We also tested the 38-indel multiplex with samples from human remains in order to evaluate the use of indels as a complementary tool in kinship analysis.

2. Materials and methods

2.1. DNA samples

A random sample of 137 unrelated individuals was selected from paternity cases, in Rio de Janeiro, South East Brazil. Additionally, two samples were selected to represent Brazilian groups with different ancestry contributions: 45 unrelated Terenas (Native Americans living in the state of Mato Grosso do Sul, Central West Brazil) and 115 unrelated individuals living in Rio de Janeiro and self-declared as having African ancestry.

Samples were collected with informed voluntary donation for research purposes approved by the Ethics Committee of the Federal University of Mato Grosso do Sul (CI n° 015/98) and the State University of Rio de Janeiro (CAAE: 0067.0.228.000-09). Genomic DNA was extracted from peripheral blood samples using a salting out method.

Although Rio de Janeiro is an admixed population where three different ancestries of European, African and Native American origins contribute to high population variability, no population substructure has been described. For this reason, we did not take into account ancestry information when select a sample to represent a forensic reference population of the people currently living in Rio de Janeiro. As it is known that a diverse pattern of admixture exists in Brazil, we included a sample from Rio de Janeiro with a high African ancestry to infer the expected diversity of Afro-descendant groups for the indel loci typed and to evaluate the impact of substructure on the results of identification when using a single Brazilian database.

In order to extend the worldwide coverage of population data for the 38-indel multiplex and to evaluate the discrimination power of this set of markers in Native American populations, the Terena sample was studied and compared to the European, African and East Asian population groups originally reported [5].

Finally, to evaluate the performance of indels in paternity cases involving post-mortem samples, we selected a mother-child-alleged father trio where the father had deceased and no amplification was achieved with standard STRs for his sample (AmpFISTR® Identifier kit used; Applied Biosystems, CA, USA). Blood samples were obtained from the mother and child and DNA was extracted using a standard Chelex method. DNA was extracted

from teeth and bones of the alleged father using the DNA IQ™ Casework Sample Kit for Maxwell® 16 (Promega Corporation, WI, USA), following manufacturer's instructions. AmpFISTR® MiniFiler (Applied Biosystems) typing was then used to compare to indel genotyping performance.

2.2. Indel genotyping

The 38 indels were genotyped using the PCR multiplex protocol originally described [5]. Amplification products were separated by electrophoresis in a 3130 Genetic Analyzer (Applied Biosystems) and fragment sizes and allele calls were determined automatically using GeneMapper v3.2 ID software (Applied Biosystems).

2.3. Data analysis

Allele frequencies, Hardy–Weinberg equilibrium analysis and estimation of population pairwise genetic distances (F_{ST}) were computed using the Arlequin v3.0 software [16]. Gene diversity was calculated as the expected heterozygosity for diploid data, using the Nei's formula, as implemented in Arlequin. Statistical parameters of forensic interest: power of discrimination (PD), power of exclusion (PE) and matching probability (MP) were calculated using PowerStats v1.2 (Promega Corporation). Genetic distances were visualized in two-dimensional plots using the multidimensional scaling (MDS) method included in data analysis software STATISTICA v8.0 (StatSoft, Inc. 2007; www.statsoft.com).

3. Results and discussion

3.1. Forensic utility of the 38 indels

The allele frequencies of the 38 indels in the three Brazilian populations are outlined in Table 1. As suggested by previous allele frequency estimates of the three continental population groups [5], a high genetic diversity was observed in the sample of Rio de Janeiro and in individuals with self-declared African ancestry, with all 38 indel markers exhibiting high heterozygosities in both cases ($0.283 < GD < 0.499$; Table 1).

In the Terena population sample, a lower mean diversity value was found compared to the Rio de Janeiro samples as well as to those previously reported in Africans, Europeans and East Asians [5]. Although this result can be expected for Amerindian populations, more subject to genetic drift, the reduced genetic diversity in such populations may also reflect the selection criteria originally used, focusing on highly polymorphic indels in those three major population groups [5].

Forensic parameter estimates comprised a combined matching probability value of 1.4×10^{-15} (PD = 0.9999999999999999) for Rio de Janeiro population; 3.2×10^{-15} for Rio de Janeiro Afro-descendant population (PD = 0.999999999999997); and 2.7×10^{-13} for the Terena population (PD = 0.9999999999997). The combined probability of exclusion in trios (PE) was 99.83%, 99.88% and 99.26%, respectively.

In summary, the 38-indel multiplex was observed to be highly informative in the Rio de Janeiro population, showing an increased efficiency in comparison to values for Europeans and Africans [5]. The high values of mean average diversity in both samples from Rio de Janeiro are in line with values that can be expected for populations with African and European admixture. Moreover, these values did not differ significantly between the two Rio de Janeiro samples, although they present differing average admixture component proportions. This result can be explained by the criteria used in the original indel selection with markers chosen to show high diversities both in Africans and Europeans. Therefore, similarly high discrimination power can be expected for other

Table 1
Allele frequencies, gene diversity, power of discrimination and power of exclusion in Rio de Janeiro general population (RJ); and self-declared Afro-descendants (AFD); and in Terena from Mato Grosso do Sul, Brazil.

Loci code	rs number	Allele frequencies (Short allele)			Gene diversity			Power of discrimination			Power of exclusion		
		RJ	AFD	Terena	RJ	AFD	Terena	RJ	AFD	Terena	RJ	AFD	Terena
B1	rs34541393	0.449	0.426	0.567	0.494	0.488	0.490	0.593	0.592	0.631	0.232	0.224	0.160
B2	rs16624	0.609	0.465	0.222	0.475	0.497	0.345	0.619	0.606	0.512	0.155	0.216	0.089
B3	rs2307689	0.383	0.322	0.333	0.472	0.436	0.444	0.615	0.588	0.610	0.155	0.137	0.089
B4	rs35769550	0.361	0.252	0.867	0.461	0.376	0.231	0.592	0.542	0.379	0.178	0.094	0.036
B5	rs2307700	0.405	0.335	0.333	0.481	0.445	0.444	0.612	0.610	0.610	0.178	0.099	0.068
B6	rs140809	0.361	0.339	0.122	0.461	0.447	0.214	0.614	0.608	0.369	0.134	0.114	0.043
B7	rs3047269	0.515	0.566	0.511	0.499	0.490	0.499	0.615	0.586	0.648	0.204	0.238	0.143
B8	rs33972805	0.522	0.443	0.622	0.498	0.493	0.469	0.647	0.639	0.630	0.144	0.149	0.114
B9	rs33917182	0.646	0.557	0.856	0.457	0.493	0.247	0.594	0.632	0.366	0.166	0.162	0.019
B10	rs16402	0.248	0.313	0.289	0.372	0.429	0.410	0.537	0.576	0.541	0.102	0.149	0.178
G1	rs1610871	0.522	0.548	0.411	0.498	0.495	0.483	0.587	0.599	0.461	0.248	0.224	0.411
G2	rs2067238	0.507	0.404	0.411	0.499	0.481	0.483	0.630	0.555	0.624	0.178	0.271	0.160
G3	rs2067294	0.266	0.243	0.267	0.390	0.368	0.390	0.553	0.528	0.539	0.106	0.114	0.143
G4	rs2307710	0.350	0.413	0.067	0.454	0.484	0.124	0.582	0.633	0.231	0.184	0.143	0.014
G5	rs2308242	0.299	0.287	0.148	0.419	0.408	0.251	0.553	0.565	0.394	0.172	0.125	0.031
G6	rs2307580	0.456	0.361	0.689	0.495	0.460	0.428	0.633	0.596	0.559	0.166	0.169	0.178
G7	rs1160956	0.668	0.617	0.533	0.443	0.472	0.497	0.592	0.595	0.604	0.144	0.191	0.218
G8	rs34511541	0.442	0.426	0.611	0.492	0.488	0.474	0.591	0.569	0.595	0.232	0.261	0.197
G9	rs2307978	0.315	0.343	0.289	0.431	0.450	0.410	0.597	0.559	0.577	0.093	0.216	0.089
Y1	rs3051300	0.401	0.294	0.133	0.480	0.414	0.231	0.635	0.562	0.391	0.129	0.145	0.051
Y2	rs10629077	0.274	0.265	0.156	0.397	0.389	0.262	0.562	0.548	0.423	0.098	0.119	0.051
Y3	rs10688868	0.292	0.243	0.444	0.413	0.368	0.493	0.553	0.531	0.624	0.160	0.104	0.178
Y4	rs2067208	0.223	0.171	0.211	0.345	0.283	0.332	0.513	0.441	0.488	0.076	0.046	0.128
Y5	rs2307579	0.423	0.491	0.089	0.487	0.499	0.162	0.615	0.623	0.269	0.184	0.191	0.014
Y6	rs2308020	0.715	0.687	0.578	0.407	0.429	0.487	0.572	0.576	0.648	0.094	0.149	0.114
Y7	rs3080855	0.296	0.324	0.300	0.416	0.437	0.419	0.581	0.592	0.575	0.098	0.130	0.128
Y8	rs1610919	0.577	0.583	0.633	0.487	0.485	0.464	0.608	0.509	0.584	0.197	0.346	0.197
Y9	rs2307839	0.212	0.230	0.200	0.333	0.354	0.319	0.500	0.513	0.486	0.073	0.119	0.068
R1	rs2308137	0.427	0.478	0.267	0.488	0.498	0.390	0.620	0.638	0.557	0.178	0.162	0.089
R2	rs36040336	0.617	0.574	0.522	0.472	0.488	0.498	0.615	0.649	0.663	0.155	0.114	0.101
R3	rs1160886	0.332	0.317	0.567	0.443	0.433	0.490	0.580	0.595	0.631	0.166	0.109	0.160
R4	rs2308026	0.314	0.222	0.389	0.430	0.345	0.474	0.589	0.511	0.595	0.119	0.090	0.197
R5	rs2307526	0.347	0.339	0.122	0.452	0.447	0.214	0.590	0.580	0.369	0.166	0.176	0.043
R6	rs34811743	0.635	0.596	0.811	0.463	0.481	0.306	0.609	0.643	0.470	0.149	0.109	0.101
R7	rs2308189	0.467	0.535	0.656	0.497	0.497	0.451	0.631	0.656	0.571	0.172	0.119	0.197
R8	rs5895447	0.296	0.296	0.278	0.416	0.416	0.401	0.559	0.555	0.568	0.155	0.162	0.078
R9	rs2308171	0.296	0.426	0.211	0.416	0.488	0.332	0.569	0.602	0.497	0.134	0.207	0.101
R10	rs35605984	0.500	0.548	0.756	0.499	0.494	0.369	0.638	0.620	0.535	0.163	0.188	0.089
Mean					0.451	0.446	0.380						

Brazilian populations, even with contrasting proportions of African and European ancestries.

In Native Americans, the results of this study show reduced diversity but still present a high discrimination power and a fair power of exclusion, providing a suitable short amplicon marker set to complement STRs applied to a fourth population group, characterised for the first time for these 38 indels.

3.2. Population analysis

All analysed loci met the Hardy–Weinberg equilibrium in both Rio de Janeiro and Terena populations, after applying the Bonferroni correction for multiple tests (exact p values ≥ 0.066 and 0.004 , respectively). Therefore, we do not expect significant levels of substructure in these two Brazilian populations. On the other hand, it is not expected a large frequency of null alleles in Native Americans, similarly to previous findings in other continental groups [5].

Pairwise genetic distance analysis was performed between the studied samples and those representatives of three continental population groups: Europe, Africa and East Asia (Portugal; Angola and Mozambique; Macau and Taiwan, respectively). In the comparison of both Rio de Janeiro samples the observed genetic distance was low, but showing a significant p value between them ($F_{ST} = 0.0054$; $p \leq 10^{-5}$). For the remaining pairwise comparisons, high values and significant genetic distances were observed in all

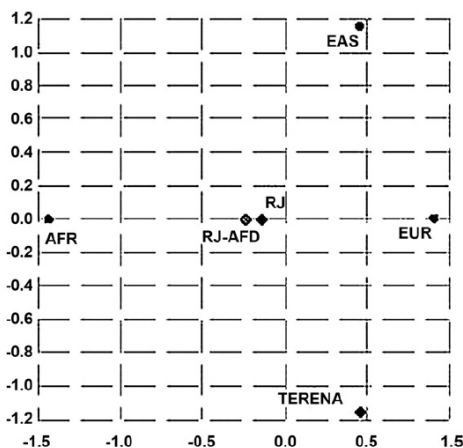
cases (Fig. 1). The MDS plot of Fig. 1, indicates the two samples from Rio de Janeiro placed between the European and African reference samples, with the Afro-descendants plotted closer to the Africans. In comparison the Native Americans are more distantly positioned from the other three groups.

3.3. Application of indels to a challenging paternity investigation

Conventional STR analysis of the degraded DNA from the alleged father showed most loci failing amplification (Supplementary data Fig. S1) and no analysis were possible from the data. In comparison, the AmpFSTR[®]Minifiler (Applied Biosystems) genotyped 6 of 8 STR loci (Supplementary data Fig. S2) giving a paternity index (PI) of 2307 (Table 2) and a near-complete profile was obtained for the 38-indel multiplex. Some evidence of allele imbalance and allelic dropout was observed (Supplementary data Figs. S3 and S4) so considering just the concordant genotypes between replicated analyses, 26 out of 38 indel loci provided a PI of

Table 2
Paternity indices (PI) and probabilities based on the results obtained with STRs, indels, and with both type of markers.

	STRs (Minifiler)	INDELS (HID Indel-plex)	STRs+INDELS
PI	2306.55	2147.43	4953147.84
W	99.957%	99.953%	99.999980%



	AFR	EUR	EAS	TERENA	RJ	RJ-AFD
AFR	-					
EUR	0.108	-				
EAS	0.122	0.091	-			
TERENA	0.145	0.092	0.095	-		
RJ	0.057	0.011	0.069	0.064	-	
RJ-AFD	0.026	0.028	0.071	0.080	0.005	-

Fig. 1. F_{ST} values (for $p \leq 10^{-5}$) and MDS plot of the pairwise genetic distances between three Brazilian samples [Rio de Janeiro general population (RJ); self-declared Afro-descendants (RJ-AFD); Terena from Mato Grosso do Sul (TERENA)] and 3 reference samples from Africa (AFR), Europe (EUR) and East Asia (EAS).

2147. Joining the information from indels and miniaturised STRs provided a considerably enhanced PI of almost 5 millions allowing a sufficiently informative PI to be achieved when individual short-amplicon markers set analysis would have given less definitive likelihoods. These results further indicate the suitability of indels for typing degraded forensic samples, while offering simplicity of analysis and reduced expenditure compared to commercial STR kits.

4. Conclusions

The 38 autosomal indels further characterised in this study were observed to be suitable for general forensic purposes in admixed populations, such as the Brazilian samples of Rio de Janeiro we analysed. The allele frequency estimates for Terena Native Americans provide reference data to complement those of the original study of the 38 indels. Indels provide an efficient way to amplify degraded DNA, especially in situations where conventional STRs can be inefficient, or may give uninformative paternity indices alone.

This paper follows the guidelines for publication of population data requested by the journal [17].

Acknowledgements

Financial support was granted by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and DNA Program – State University and Justice Court of Rio de Janeiro, Brazil. The Institute of Pathology and Molecular Immunology of the University of Porto (IPATIMUP) is an Associate Laboratory of the Portuguese Ministry of Education and Science and is partially supported by FCT.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.fsigen.2011.12.008.

References

- [1] P. Gill, Role of short tandem repeat DNA in forensic casework in the UK – past, present, and future perspectives, *Biotechniques* 32 (2002) 366–372.
- [2] W. Bär, A. Kratzer, M. Machler, W. Schmid, Postmortem stability of DNA, *Forensic Sci. Int.* 39 (1988) 59–70.
- [3] J.M. Butler, Y. Shen, B.R. McCord, The development of reduced size STR amplicons as tools for analysis of degraded DNA, *J. Forensic Sci.* 48 (2003) 1054–1064.
- [4] J.J. Sanchez, C. Phillips, C. Børsting, K. Balogh, M. Bogus, M. Fondevila, C.D. Harrison, E. Musgrave-Brown, A. Salas, D. Syndercombe-Court, P.M. Schneider, A. Carracedo, N. Morling, A multiplex assay with 52 single nucleotide polymorphisms for human identification, *Electrophoresis* 27 (2006) 1713–1724.
- [5] 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.
- [6] J.R. Pimenta, S.D. Pena, Efficient human paternity testing with a panel of 40 short insertion-deletion polymorphisms, *Gen. Mol. Res.* 9 (2010) 601–607.
- [7] N.S. Freitas, R.L. Resque, E.M. Ribeiro-Rodrigues, J.F. Guerreiro, N.P. Santos, A.K. Ribeiro-dos-Santos, S.E. Santos, X-linked insertion/deletion polymorphisms: forensic applications of a 33-markers panel, *Int. J. Legal Med.* 124 (2010) 589–593.
- [8] R. Pereira, V. Pereira, I. Gomes, C. Tomas, N. Morling, A. Amorim, M. Prata, A. Carracedo, L. Gusmão, A method for the analysis of 32 X chromosome insertion deletion polymorphisms in a single PCR, *Int. J. Legal Med.* (2011), doi:10.1007/s00414-011-0593-2.
- [9] C. Phillips, M. Fondevila, M. Garcia-Magarinos, A. Rodriguez, A. Salas, A. Carracedo, M.V. Lareu, Resolving relationship tests that show ambiguous STR results using autosomal SNPs as supplementary markers, *Forensic Sci. Int. Genet.* 2 (2008) 198–204.
- [10] R. Pereira, C. Phillips, C. Alves, A. Amorim, A. Carracedo, L. Gusmão, Insertion/deletion polymorphisms: a multiplex assay and forensic applications, *Forensic Sci. Int. Genet. (Suppl. Ser.)* 2 (2009) 513–515.
- [11] C. Børsting, N. Morling, Mutations and/or close relatives? Six case work examples where 49 autosomal SNPs were used as supplementary markers, *Forensic Sci. Int. Genet.* 5 (2011) 236–241.
- [12] C. Oliveira, S. Sousa, H. Pinheiro, R. Karam, R. Bordeira-Carrico, J. Senz, P. Kaurah, J. Carvalho, R. Pereira, L. Gusmão, X. Wen, M.A. Cipriano, J. Yokota, F. Carneiro, D. Huntsman, R. Seruca, Quantification of epigenetic and genetic 2nd hits in CDH1 during hereditary diffuse gastric cancer syndrome progression, *Gastroenterology* 136 (2009) 2137–2148.
- [13] J.L. Weber, D. David, J. Heil, Y. Fan, C. Zhao, G. Marth, Human diallelic insertion/deletion polymorphisms, *Am. J. Hum. Genet.* 71 (2002) 854–862.
- [14] R.E. Mills, C.T. Luttig, C.E. Larkins, A. Beauchamp, C. Tsui, W.S. Pittard, S.E. Devine, An initial map of insertion and deletion (INDEL) variation in the human genome, *Genome Res.* 16 (2006) 1182–1190.
- [15] R.E. Mills, W.S. Pittard, J.M. Mullaney, U. Farooq, T.H. Creasy, A.A. Mahurkar, D.M. Kemeza, D.S. Strassler, C.P. Ponting, C. Webber, S.E. Devine, Natural genetic variation caused by small insertions and deletions in the human genome, *Genome Res.* 21 (2011) 830–839.
- [16] L. Excoffier, G. Laval, S. Schneider, Arlequin ver. 3.0: An integrated software package for population genetics data analysis, *Evol. Bioinform.* 1 (2005) 47–50.
- [17] J.M. Carracedo, L. Butler, W. Gusmão, L. Parson, P.M. Roewer, Schneider, Publication of population data for forensic purposes, *Forensic Sci. Int. Genet.* 4 (2010) 145–147.

3.2 Artigo II:

Título do artigo: **Analysis of Genetic Ancestry in the Admixed Brazilian Population from Rio de Janeiro using 46 Autosomal Ancestry-Informative Indel Markers**

Publicação: Annals of Human Biology, 2013; 40(1): 94-8.

doi: 10.3109/03014460.2012.742138.

RESUMO:

A população brasileira é altamente heterogênea, como resultado de cinco séculos de mistura interétnica entre nativos ameríndios, colonizadores europeus e africanos que chegaram durante a escravidão. O objetivo deste estudo foi avaliar as proporções de mistura interétnica na população brasileira do Rio de Janeiro utilizando marcadores autossômicos Informativos de ancestralidade (AIMs). Os dados autossômicos também foram comparados com os resultados esperados a partir de marcadores genéticos uniparentais. Um total de 413 indivíduos foram genotipados para 46 AIM-Indels e estimativas de ancestralidade foram avaliadas utilizando amostras de HGDP-CEPH como referência ancestral. Indivíduos do Rio de Janeiro apresentaram padrões de mistura altamente diversificada. As estimativas de miscigenação global mostraram uma ancestralidade predominantemente europeia, acima de 55%, seguido por contribuições africanas e ameríndias. Um grupo separado de afrodescendentes autodeclarados, também incluído neste estudo, revelou uma maior ancestralidade africana, variando de aproximadamente 30% a 50%. A mistura interétnica do Rio de Janeiro, capturada pelos AIM-Indels autossômicos vão de acordo com registros históricos e são similares ao que era esperado a partir de informações do DNA mitocondrial e do cromossomo Y. O painel de AIM-Indel provou ser uma estratégia rápida para estimar a ancestralidade genética autossômica em níveis individuais e populacionais no Rio de Janeiro, sendo útil em genética de populações e em estudos de associações caso-controle.

SHORT REPORT

Analysis of genetic ancestry in the admixed Brazilian population from Rio de Janeiro using 46 autosomal ancestry-informative indel markers

Fernanda Saloum Neves Manta¹, Rui Pereira², Alexandre Caiafa¹, Dayse Aparecida Silva¹, Leonor Gusmão^{2,3} & Elizeu Fagundes Carvalho¹

¹DNA Diagnostic Laboratory (LDD), State University of Rio de Janeiro (UERJ), Rio de Janeiro, Brazil, ²Institute of Molecular Pathology and Immunology of the University of Porto (IPATIMUP), Porto, Portugal, and ³Medical and Human Genetics Laboratory, Federal University of Pará (UFPA), Belém, Pará, Brazil

The Brazilian population is highly heterogeneous as a result of five centuries of inter-ethnic mating between native Amerindians, European colonizers and Africans arrived during slavery. This study aimed to assess the proportions of inter-ethnic admixture in the Brazilian population of Rio de Janeiro using autosomal Ancestry-Informative Markers (AIMs). The autosomal data were also compared to the results expected from uniparental genetic markers. A total of 413 individuals were genotyped for 46 AIM-Indels and ancestry estimates were then assessed using HGDP-CEPH samples as ancestral reference. Individuals from Rio de Janeiro presented highly diverse admixture patterns. The global admixture estimates showed a predominantly European ancestry, above 55%, followed by African and Amerindian contributions. A separate self-declared Afro-descendant group also included in this study revealed an increased African ancestry, from ~30% to ~50%. The inter-ethnic admixture landscape of Rio de Janeiro captured by autosomal AIM-Indels is in agreement with historical records and similar to that expected from uniparental mtDNA and Y-chromosome information. The AIM-Indel panel proved to be a rapid strategy to estimate autosomal genetic ancestry at individual and population levels in Rio de Janeiro, which is useful in population genetics and in case-control association studies.

Keywords: Admixture, AIM, insertion deletion polymorphism, population structure

INTRODUCTION

The Brazilian population is one of the most heterogeneous in the world as a result of five centuries of inter-ethnic mating between individuals from three major contributor groups: the Amerindians, the Europeans who colonized South

America in the 1500s and the Africans who arrived through the slave trade over a span of more than 300 years (e.g. Curtain 1969; Lins et al. 2010; Pena et al. 2011). Primarily because of this settlement process, Brazilians have highly variable admixture patterns. Furthermore, the diverse regions of the country underwent different colonization processes, which to some extent shaped their genetic backgrounds, nowadays characterized by different proportions of Amerindian, European and African contribution.

The assessment of accurate admixture estimates is important in population genetic studies, particularly in the context of highly admixed populations similar to those of Brazil and most American countries. Some studies have focused on ancestry proportions in Brazilian populations using uniparental Y chromosome and mitochondrial DNA data (e.g. Alves-Silva et al. 2000; Carvalho-Silva et al. 2001; Hünemeier et al. 2007). However, studies that use autosomal ancestry-informative markers, bringing together the history of both lineages, remain scarce (e.g. Lins et al. 2010; Santos et al. 2010).

The heterogeneity of the Brazilian population poses an additional difficulty in pharmacogenetic and clinical genetic research, particularly when designing case-control association studies, in which the presence of stratification among case and control groups can produce confounding results leading to erroneous conclusions (Marchini et al. 2004). Recent studies have demonstrated only a weak correlation between skin pigmentation or self-declared ethnicity and genetic ancestry as determined by DNA markers (Parra et al. 2004; Leite et al. 2011; Lins et al. 2011). Therefore, an accurate assessment of a subject's genetic ancestry is needed to control for stratification in genetic association studies among admixed populations. In practice, Individual Ancestry Estimates (IAEs) can be used to validate admixture

Correspondence: Elizeu Fagundes Carvalho, Laboratório de Diagnóstico por DNA, São Francisco Xavier, 524, Maracanã, CEP: 21550-013, Rio de Janeiro, RJ, Brazil. Tel: (+55) 21 23342183. Fax: (+55) 21 23340594. E-mail: elizeufc@hotmail.com

(Received 2 March 2012; accepted 15 October 2012)

proportions and match admixed case-control groups or applied as co-variables in association calculations to adjust for population stratification effects (Tsai et al. 2006; Tian et al. 2008). Ancestry-Informative Markers (AIMs) presenting marked allele frequency differentials between geographically separated groups have proven to be especially useful for inferring the putative biogeographic origin of individuals and estimate the ancestry proportions of admixed individuals and populations (e.g. Shriver et al. 1997; Halder et al. 2008; Pereira et al. 2012).

In this work, the inter-ethnic admixture proportions in Rio de Janeiro, Brazil were estimated at the individual and population levels using a set of 46 autosomal ancestry-informative Indel markers and the results were compared to uniparental genetic data for the same regions.

MATERIALS AND METHODS

Samples were obtained from paternity casework from the DNA Diagnostic Laboratory (LDD). Although samples collected from paternity tests are not always representative of the general population, our laboratory performs paternity investigations by private or judicial request, covering, therefore, people of both low and high social status living in Rio de Janeiro. All donors gave their written consent for the use of the samples in an anonymous way. Furthermore, the scientific investigation proposal, together with the informed consent, had been evaluated and approved in advance by the Science Ethical Committee of the State University of Rio de Janeiro (CAAEE:0067.0.228.000-09). We randomly selected 280 unrelated individuals born in the metropolitan region of the Rio de Janeiro city (22°54'10"S, 43°12'28"W), the capital of the State of Rio de Janeiro (RJ), southeastern Brazil. We also studied a specific group of 113 unrelated, self-declared Afro-descendants (RJ-AFD) to check if detectable differentiation compared to the global population existed and test the sensitivity of the AIM panel.

Genotyping of the 46 AIM-Indels was accomplished using single multiplex PCR and capillary electrophoresis as described in Pereira et al. (2012).

Allele frequency estimations, Hardy-Weinberg exact tests and genetic distance analysis were performed with Arlequin v3.5 (Excoffier and Lischer 2010). STATISTICA software (Statsoft, Tulsa, Oklahoma) was used to obtain the pairwise F_{ST} multidimensional scaling (MDS) plot. Pairwise Linkage Disequilibrium was tested using Genepop 4.1 (Rousset 2008). Individual and global ancestry estimates were primarily assessed by STRUCTURE v2.3.3 (Pritchard et al. 2000) using the HGDP-CEPH Diversity panel (sub-set H952) as reference data for the African, European and Amerindian ancestral populations (data from Pereira et al. 2012). Analyses were performed employing the Admixture model with correlated allele frequencies and considering $K = 1$ to $K = 6$ (100 000 burnin plus 100 000 MCMC repetitions; three runs for each K) either using or not prior information on the origin of ancestral populations to assist ancestry estimations of the study populations (USEPO-PINFO option). A simulation summary is presented in

Supplementary Figure 1. Finally, we applied Admixture v1.22 (Alexander et al. 2009; Alexander and Lange 2011) to perform similar analyses (unsupervised and supervised; with default settings) to complement and confirm our previous results.

RESULTS AND DISCUSSION

In this study, a set of 46 autosomal ancestry-informative Indels was analysed in the general population and in a group of self-declared Afro-descendants from Rio de Janeiro to genetically characterize these populations and estimate the ancestry proportions of Amerindian, European and African origin. The allele frequencies observed are presented in Supplementary Table I. When compared to ancestral populations, average gene diversity was higher in both samples from Rio de Janeiro, as would be expected after five centuries of admixture between individuals of very different ancestral origins. For all studied loci, the Hardy-Weinberg exact test did not reveal significant departures from equilibrium expectations, for a significance level obtained after Bonferroni correction for multiple tests.

The pairwise LD status among the 46 AIMs was previously investigated in the ancestral populations showing no significant signs of associations (Pereira et al. 2012). In this study no significant results were observed for the self-declared Afro-descendant sample while, contrarily to all populations analysed until now, for loci in contiguous chromosomal positions, significant associations were found in the RJ population between five pairs (in chromosomes 1, 5, 6, 18 and 22; results in Supplementary Table III). Higher levels of LD can arise from recent population admixture and stratification, which is the most likely explanation for the observed results.

Figure 1 illustrates the MDS results from pairwise genetic distance analysis between all populations (see Supplementary Table II for details). The two samples from Rio de Janeiro appear to be closely related to each other in an intermediate position among the ancestral contributors. Despite being very low, the genetic distance between the two groups is statistically significant ($F_{ST} = 0.020$; $p \leq 10^{-5}$). Furthermore, the relative position of the general population sample (closer to Europeans) compared to that of the self-declared Afro-descendants (closer to Africans) clearly reflects this differentiation and highlights the sensitivity of the AIM panel, even to study populations that are relatively similar.

The majority of previous studies addressing ancestry estimates in Brazilian populations were based on uniparental markers. The main goal of this work was to evaluate admixture proportions in the Rio de Janeiro population using autosomal markers. Because autosomes undergo recombination every generation, they record the histories of both male and female lineages and it is expected that the ancestry proportions captured correspond approximately to the average values indicated by mtDNA and Y chromosome markers.

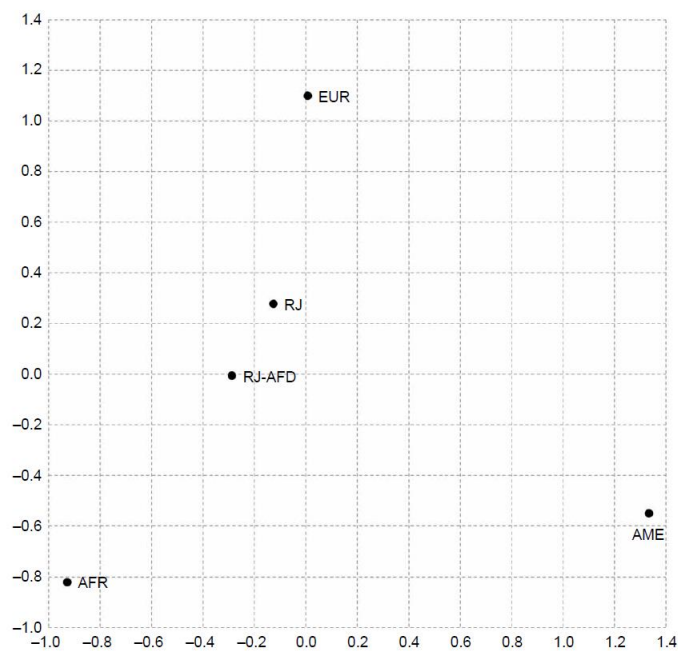


Figure 1. Multidimensional scaling plot obtained from the pairwise F_{ST} genetic distance analysis between HGDP-CEPH ancestral populations, the Rio de Janeiro general population (RJ) and self-declared Afro-descendants (RJ-AFD) using 46 AIM-Indels. (D-star: Raw stress = 0.000 000; Alienation = 0.000 000. D-hat: Raw stress = 0.000 000; Stress = 0.000 000).

Using the HGDP-CEPH diversity panel as a reference, we assessed admixture proportions in Rio de Janeiro with STRUCTURE and Admixture software in different scenarios, either providing or not prior information on the geographic location of ancestral populations. Considering the historical records of an essentially tri-hybrid contribution of Amerindians, Europeans and Africans to the formation of the present Brazilian population, in this work, we finally estimated individual and global admixture proportions assuming these three main ancestral populations (Figure 2; Table I. The individuals from Rio de

Janeiro revealed highly variable admixture patterns resulting in a global population of predominantly European ancestry (55.2% and 58.6%) with smaller African (31.1% and 30.3%) and Amerindian (13.7% and 11.0%) contributions, respectively, for STRUCTURE and Admixture. As expected, the self-declared Afro-descendant group revealed a higher African contribution (49.7% and 51.6%). The admixture proportions obtained by both programmes were similar (non-differentiation exact p -values ≤ 0.596 and ≤ 0.859 for RJ and RJ-AFD, respectively).

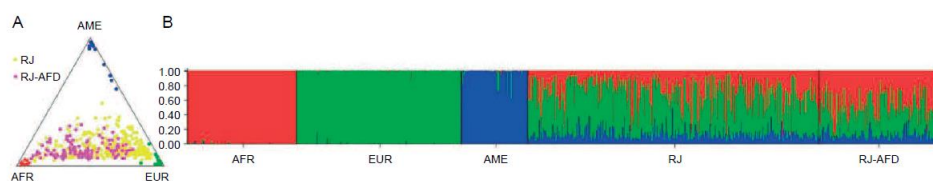


Figure 2. Ancestral membership proportions obtained by STRUCTURE for the HGDP-CEPH diversity panel reference samples and tested populations from the Rio de Janeiro general population (RJ) and self-declared Afro-descendants (RJ-AFD) using 46 AIM-Indels. (a) triangular plot; (b) bar plot. Analyses were performed using prior information on the origin of ancestral populations to assist ancestry estimations of the study populations (USEPOPINFO option) employing the Admixture model with correlated allele frequencies and considering $K = 3$ (100 000 burnin plus 100 000 MCMC repetitions). Results are presented for the replicate run with the highest $\ln P(D) = -34\,935.9$.

Table 1. Ancestral membership proportions obtained for the HGDP-CEPH diversity panel reference samples and tested populations from the Rio de Janeiro population and self-declared Afro-descendants using 46 AIM-Indels.

	STRUCTURE			Admixture		
	AFR	EUR	AME	AFR	EUR	AME
AFR	0.996	0.003	0.001	0.999 98	0.000 01	0.000 01
EUR	0.002	0.994	0.004	0.000 01	0.999 98	0.000 01
AME	0.002	0.025	0.974	0.000 01	0.000 01	0.999 98
RJ	0.311	0.552	0.137	0.303 38	0.586 42	0.110 20
RJ-AFD	0.497	0.383	0.120	0.515 81	0.390 09	0.094 10

Results are presented for admixture analyses using prior information on the origin of ancestral reference samples and $K = 3$ (STRUCTURE analysis with Admixture model and 'USEPOPINFO' option; Admixture analysis with 'supervised' option). Please note that the reported number of decimal places is different, as was output by the two programmes.

AFR, African; EUR, European; AME, Amerindian; RJ, Rio de Janeiro; RJ-AFD, Rio de Janeiro self-declared Afro-descendants.

The entrance of Europeans and Africans into the country is historically well documented and genetically supported by previous studies using uniparental markers that showed clear evidence of a biased admixture involving predominantly descendants from European men and African women (Alves-Silva et al. 2000; Carvalho-Silva et al. 2001; Hünemeier et al. 2007). In the general population from Rio de Janeiro, the vast majority of Y chromosome lineages were of European origin (88.1%), whereas the contribution of African and Amerindian native men was estimated to be as little as 8.7% and 1.6%, respectively; the ancestry of two out of 127 Y chromosomes could not be determined (Silva et al. 2006). Conversely, mtDNA lineages are mainly African (60%), with a small presence of Amerindian (26%) and European (14%) lineages (Bernardo et al., unpublished data).

Considering the above results, it is possible to conclude that the global ancestry estimates obtained with the autosomal AIM-Indels are in agreement with the results from uniparental markers. Recent studies also using autosomal AIMS have shown an essentially tri-hybrid heritage and a predominant European ancestry for most Brazilian populations, with some differences in ancestral proportions depending on the considered geographic region or on particular self-declared ethnic groups (Lins et al. 2010; Santos et al. 2010; Pena et al. 2011). For the specific case of Rio de Janeiro, the present study has characterized the highest number of markers and samples to date. Interestingly, the ancestry estimates reported for the global population in Pena et al. (2011) appear to be noticeably biased towards Europeans when compared to our results and those expected from uniparental data. However, it is worth mentioning that the set of 40 autosomal Indels used in that study were selected to present high heterozygosity in Europeans, which may contribute to such bias (Bastos-Rodrigues et al. 2006; Pena et al. 2011).

This work demonstrated the suitability of the AIM-Indel set as a rapid and inexpensive method to measure ancestry proportions in Rio de Janeiro at the individual and population levels. The inter-ethnic admixture landscape

of Rio de Janeiro captured by autosomal AIM-Indels was similar to that based on uniparental mtDNA and Y-chromosome information and supports the historical records for a predominantly European influence and only smaller African and Amerindian contributions in this region of the country.

Declaration of interest: Financial support was granted by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ) and DNA Program—State University and Justice Court of Rio de Janeiro (UERJ/TJRJ/MPRJ), Brazil. IPATIMUP is an Associate Laboratory of the Portuguese Ministry of Science, Technology and Higher Education and is partially supported by the Portuguese Foundation for Science and Technology (FCT). LG is supported by a grant from CAPES and RP by an FCT post-doctoral fellowship (SFRH/BPD/81986/2011). The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

REFERENCES

- Alexander DH, Lange K. 2011. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics* 12:246.
- Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 19: 1655–1664.
- Alves-Silva J, da Silva Santos M, Guimaraes PE, Ferreira AC, Bandelt HJ, Pena SD, Prado VF. 2000. The ancestry of Brazilian mtDNA lineages. *Am J Hum Genet* 67:444–461.
- Bastos-Rodrigues L, Pimenta JR, Pena SD. 2006. The genetic structure of human populations studied through short insertion-deletion polymorphisms. *Ann Hum Genet* 70:658–665.
- Carvalho-Silva DR, Santos FR, Rocha J, Pena SD. 2001. The phylogeography of Brazilian Y-chromosome lineages. *Am J Hum Genet* 68:281–286.
- Curtain PD. 1969. *The Atlantic Slave Trade: a Census*. Madison, WI: University of Wisconsin Press.
- Excoffier L, Lischer HEL. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* 10:564–567.
- Halder I, Shriver M, Thomas M, Fernandez JR, Frudakis T. 2008. A panel of ancestry informative markers for estimating individual biogeographical ancestry and admixture from four continents: utility and applications. *Hum Mutat* 29:648–658.
- Hünemeier T, Carvalho C, Marrero AR, Salzano FM, Pena SDJ, Bortolini MC. 2007. Niger-Congo speaking populations and the formation of the Brazilian gene pool: mtDNA and Y-chromosome data. *Am J Phys Anthropol* 133:854–867.
- Leite TKM, Fonseca RMC, França NMd, Parra EJ, Pereira RW. 2011. Genomic ancestry, self-reported "color" and quantitative measures of skin pigmentation in Brazilian admixed siblings. *PLoS One* 6: e27162.
- Lins TC, Vieira RG, Abreu BS, Gentil P, Moreno-Lima R, Oliveira RJ, Pereira RW. 2011. Genetic heterogeneity of self-reported ancestry groups in an admixed Brazilian population. *J Epidemiol* 21:240–245.
- Lins TC, Vieira RG, Abreu BS, Grattapaglia D, Pereira RW. 2010. Genetic composition of Brazilian population samples based on a set of twenty-eight ancestry informative SNPs. *Am J Hum Biol* 22: 187–192.

- Marchini J, Cardon LR, Phillips MS, Donnelly P. 2004. The effects of human population structure on large genetic association studies. *Nat Genet* 36:512–517.
- Parra EJ, Kittles RA, Shriver MD. 2004. Implications of correlations between skin color and genetic ancestry for biomedical research. *Nat Genet* 36:S54–S60.
- Pena SDJ, Di Pietro G, Fuchshuber-Moraes M, Genro JP, Hutz MH, Kehdy FdSG, Kohlrausch F, Magno LAV, Montenegro RC, Moraes MO, Moraes MEAd, Moraes MRd, Ojopi ÉB, Perini JA, Racciopi C, Ribeiro-dos-Santos ÁKC, Rios-Santos F, Romano-Silva MA, Sortica VA, Suarez-Kurtz G. 2011. The genomic ancestry of individuals from different geographical regions of Brazil is more uniform than expected. *PLoS One* 6:e17063.
- Pereira R, Phillips C, Pinto N, Santos C, Santos SEBd, Amorim A, Carracedo Á, Gusmão L. 2012. Straightforward inference of ancestry and admixture proportions through ancestry-informative insertion deletion multiplexing. *PLoS One* 7:e29684.
- Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:945–959.
- Rousset F. 2008. genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Mol Ecol Resour* 8: 103–106.
- Santos NP, Ribeiro-Rodrigues EM, Ribeiro-dos-Santos AK, Pereira R, Gusmão L, Amorim A, Guerreiro JF, Zago MA, Matte C, Hutz MH, Santos SE. 2010. Assessing individual interethnic admixture and population substructure using a 48-insertion-deletion (INDEL) ancestry-informative marker (AIM) panel. *Hum Mutat* 31:184–190.
- Shriver MD, Smith MW, Jin L, Marcini A, Akey JM, Deka R, Ferrell RE. 1997. Ethnic-affiliation estimation by use of population-specific DNA markers. *Am J Hum Genet* 60:957–964.
- 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.
- Tian C, Gregersen PK, Seldin ME. 2008. Accounting for ancestry: population substructure and genome-wide association studies. *Hum Mol Genet* 17:R143–R150.
- Tsai HJ, Kho JY, Shaikh N, Choudhry S, Naqvi M, Navarro D, Matallana H, Castro R, Lilly CM, Watson HG, Meade K, Lenoir M, Thyne S, Ziv E, Burchard EG. 2006. Admixture-matched case-control study: a practical approach for genetic association studies in admixed populations. *Hum Genet* 118:626–639.

Supplementary material available online

3.3 Artigo III:

Título do artigo: **Revisiting the genetic ancestry of Brazilians using autosomal AIM-Indels**

Submetido à revista PLOS ONE (em maio de 2013)

RESUMO:

Há muitos estudos diferentes que contribuem para a imagem global da heterogeneidade étnica na população brasileira. Estes estudos usam diferentes tipos de marcadores genéticos e concentram-se na comparação de populações em diferentes níveis. Em alguns deles, cada região geográfica é tratada como uma população homogênea, enquanto que, outros estudos criam subdivisões diferentes: político (ex. juntando as populações por Estado), demográfica (ex. urbanas e rurais), étnica (ex. cultura, auto-declaração ou cor da pele). Neste estudo, realizamos uma reavaliação reforçada da ancestralidade genética de aproximadamente 1.300 brasileiros caracterizadas por 46 marcadores autossômicos informativos de ancestralidade (AIM). Além disso, 798 novos indivíduos de doze populações brasileiras, que representam as cinco macrorregiões geográficas do Brasil foram genotipados, incluindo uma comunidade indígena e uma comunidade rural da Amazônia. Seguindo um crescente gradiente norte ao Sul, a ancestralidade europeia foi a mais prevalente em todas as populações urbanas (com valores até 74 %). As populações do Norte consistiram de uma proporção significativa de ancestralidade indígena que foi cerca de duas vezes maior do que a contribuição africana. Por outro lado, no nordeste, centro-oeste e sudeste, a ancestralidade africana foi a segunda mais prevalente. A nível intrapopulacional, todas as populações urbanas eram altamente miscigenadas, e a maioria da variação nas proporções de ancestralidade foi observada entre os indivíduos dentro de cada população, em vez de entre as populações. No entanto, indivíduos com uma alta proporção de ancestralidade indígena só foram encontrados nas amostras dos Terena e de Santa Isabel. Nossos resultados nos permitiram refinar ainda mais o panorama genético dos brasileiros ao estabelecer a base para uma aplicação eficaz de um painel de AIM autossômicos em casos forenses e em estudos de associação clínica dentro da altamente miscigenada população brasileira.

Revisiting the genetic ancestry of Brazilians using autosomal AIM-Indels

Fernanda Saloum de Neves Manta¹, Rui Pereira², Romulo Vianna¹, Alfredo Rodolfo Beuttenmüller de Araújo³, Daniel Gita⁴, Dayse Aparecida da Silva¹, Eldamária de Vargas Wolfgramm⁵, Isabel da Mota Pontes⁶, José Ivan Aguiar⁷ Milton Ozório⁸, Elizeu Fagundes de Carvalho¹, Leonor Gusmão^{1,2}

¹DNA Diagnostic Laboratory (LDD), State University of Rio de Janeiro (UERJ), Rio de Janeiro, Brazil

²Institute of Molecular Pathology and Immunology of the University of Porto (IPATIMUP), Porto, Portugal

³Laboratory of Human Identification, University of Pernambuco, Brazil

⁴Federal University of Alagoas, Brazil

⁵Núcleo de Genética Humana e Molecular, Departamento de Ciências Biológicas, Centro de Ciências Humanas e Naturais, Universidade Federal do Espírito Santo, Vitória, Brazil

⁶Federal University of Amazonas, Brazil

⁷Department of Internal Medicine, Faculty of Medical Science, Federal University of Mato Grosso do Sul, Campo Grande, Brazil

⁸FIOCRUZ (Oswaldo Cruz Foundation), Rio de Janeiro, Brazil

Keywords: Brazil, Admixture, population structure, Native Americans

Corresponding Author:

Elizeu Fagundes Carvalho

Laboratório de Diagnóstico por DNA

São Francisco Xavier, 524

Maracanã

CEP: 21550-013

Rio de Janeiro, RJ, Brazil.

Tel.: (+55) 21 23342183; fax: (+55) 21 23340594.

E-mail address: elizeufc@hotmail.com

Abstract

There are many different studies that contribute to the global picture of the ethnic heterogeneity in Brazilian populations. These studies use different types of genetic markers and are focused on the comparison of populations at different levels. In some of them, each geographical region is treated as a single homogeneous population, whereas other studies create different subdivisions: political (e.g., pooling populations by State), demographic (e.g., urban and rural), or ethnic (e.g., culture, self-declaration, or skin colour). In this study, we performed an enhanced reassessment of the genetic ancestry of ~ 1,300 Brazilians characterised for 46 autosomal Ancestry-Informative Markers (AIMs). In addition, 798 newly genotyped individuals from twelve Brazilian populations representing the five geographical macro-regions of Brazil, including a Native American community and a rural Amazonian community. Following an increasing North to South gradient, European ancestry was the most prevalent in all urban populations (with values up to 74%). The populations in the North consisted of a significant proportion of Native American ancestry that was about two times higher than the African contribution. Conversely, in the Northeast, Center-West and Southeast, African ancestry was the second most prevalent. At an intrapopulation level, all urban populations were highly admixed, and most of the variation in ancestry proportions was observed between individuals within each population rather than among population. Nevertheless, individuals with a high proportion of Native American ancestry are only found in the samples from Terenas and Santa Isabel. Our results allowed us to further refine the genetic landscape of Brazilians while establishing the basis for the effective application of an autosomal AIM panel in forensic casework and clinical association studies within the highly admixed Brazilian populations.

Introduction

Despite the nature of genetic markers (e.g., blood groups, proteins or DNA sequences) or their location in the genome (e.g., mitochondria, autosomes or heterosomes), polymorphisms known as Ancestry Informative Markers (AIMs), present very high intercontinental allelic differentiation across populations (e.g., [1,2,3]).

The investigation of genetic ancestry profiles of human populations is a valuable tool to understand the dynamics of migrations and colonisation events, as well as to determine admixture patterns inside populations.

Ancestry estimates play an important role in correcting for population stratification effects in case-control genetic association studies (e.g., [4,5]), particularly in studies carried out on ethnic admixed individuals in which spurious genotype-phenotype associations may appear due to differences in the allele frequencies of parental groups that contribute differentially in case and control samples. To avoid misinterpretation of the association results, individual ancestry estimates can be considered when calculating statistics (e.g., STRAT software; <http://pritch.bsd.uchicago.edu/software/STRAT.html>), in addition to pre-selecting criteria to match controls with available cases (for a revision on this subject see [6]).

In the field of forensic genetics, having a set of markers that may provide estimates of ancestral membership proportions or help identify the source population of the donor of a certain piece of evidence can help direct the criminal investigation [7,8].

Brazil is well known for the heterogeneous distribution of three main ancestral contributions from Native Americans, Europeans and Africans. These people met and mated among themselves in distinct ways, giving rise to a highly multiethnic admixed population. The European and African colonisation of the Brazilian territory, previously occupied only by Native Americans, started on the coast and gradually reached the interior. The progression of colonisation was highly diverse in different regions, as far as European, African and Native American parentages were concerned [9]. This complex process, in a territory of almost continental dimensions, is consequently reflected in the variance in the genetic composition of the present populations (e.g., [10,11,12]).

The first attempts to evaluate the ethnic diversity of Brazilians were based on blood groups and protein markers and provided a broad overview of the ethnic heterogeneity associated with the different geographic regions of the country (e.g., [13,14]). Since the beginning of its application to population genetics, lineage markers have been used to understand the complex process of admixture and to characterise the mating patterns across the country (e.g., [15,16,17,18,19]). More recently, a series of studies were undertaken using autosomal markers representing different types of DNA variation, namely, Short Tandem Repeats (STRs), Single Nucleotide polymorphisms (SNPs) and Insertion-deletions (Indels) (e.g., [3,11,12,20]).

Apart from the variation in the number and/or type of markers that have been used to evaluate admixture processes in Brazil, consideration of the various sampling strategies that have been used until now is also important. In many studies, it is not possible to have a comprehensive view of the population gene pool because only some population subgroups are

investigated, namely, groups classified based on (a) self-reported ancestry, (b) social status, (c) skin colour (d) or other phenotypic classifications.

Considering the demography of Brazil in addition to the disparity of markers and sampling criteria that have been used in different publications, the ancestry of the Brazilians is far from fully known. Hence, additional analyses using larger random samples to cover new populations can be helpful in determining novel aspects of the genetic structure within Brazil's five macro-regions. To accomplish this objective, we performed thorough analyses involving existing data [3,21], the typing of ancestry-informative autosomal Indels (AIM-Indels) from 798 new samples, and the comparison of our results with those from other publications.

The typing of 46 AIM-Indels was performed in (i) random samples from 10 out of 27 Brazilian Federative Units within Brazil's five macro-regions, (ii) an Amerindian community, and (iii) a sample of Santa Isabel do Rio Negro (also known as simply "Santa Isabel", the term adopted henceforth) that is a small admixed Amazonian population, which is quite isolated and known to have experienced a reduced exposure to European and African influxes. To compare our results with information already available concerning the ancestry of Brazilians, the data were collected from a review of previously published studies describing different types of markers and populations.

Material and Methods

Ethics Statement

All samples involved in the study were anonymised DNA extracts previously obtained from healthy unrelated individuals. The samples were collected under written informed consent to participate in this study. This study was approved by the Ethic Committees at the State of Rio de Janeiro (CAAE:0067.0.228.000-09). The current study complies with the ethical principles of the 2000 Helsinki Declaration of the 206 World Medical Association (<http://www.uma.net/e/policy/b3.htm>). The access to the Terena community living area was approved by FUNAI second authorisation n.º016/CGEP/99.

Sample collection and DNA extraction

During this study, a total of 798 samples were collected from unrelated individuals from 12 different populations (see Figure 1 for the locations and number of samples).

The samples from Rio Grande do Sul, Santa Catarina, Paraná, São Paulo and Minas Gerais were obtained from paternity casework at the DNA Diagnostic Laboratory (LDD). The samples from Pernambuco were collected at the Laboratory of Human Identification, University of Pernambuco, from cases of paternity investigation. The samples from Espírito Santo and Manaus were obtained from students and professors at the federal universities of Espírito Santo and Amazonas, respectively. The samples from Mato Grosso do Sul were obtained from students at the State school João Ponce de Arruda. The samples from Alagoas and Santa Isabel were selected from previous research projects belonging to healthy unrelated individuals used as controls in association studies, which involved the State hospitals as well as the Federal University of Alagoas and the Research Institute FIOCRUZ (Oswaldo Cruz Foundation). The Terena samples were those previously included in Silva *et al.* [22].

DNA was extracted following salting-out or Chelex standard protocols.

Our samples represent a random selection from the users of the services mentioned above (hospital, universities or paternity testing labs), and no information about the skin colour or ethnicity of the donors was considered for sample selection purposes. Other samples that were previously typed for the same set of markers were also used in data analyses, including Belém [3] and Rio de Janeiro [21] (see Figure 1). Additionally, to perform supervised ancestry analysis estimates, we used data available for HGDP-CEPH reference samples from African, European and Native American populations [3].

Genetic markers and genotyping

A panel of 46 AIM-Indels was genotyped in a single multiplex PCR followed by capillary electrophoresis, according to the protocol described by Pereira *et al.* [3]. Dye-labelled amplified fragments were separated and detected using an ABI 3500 Genetic Analyzer (Life Technologies), and automated allele calls were obtained with GeneMapper v.4.1 (Life Technologies).

Statistical analyses

Genetic diversity parameters, including the estimation of allele frequencies, observed and expected heterozygosities, Hardy-Weinberg exact tests, F_{ST} genetic distance analysis and

Analysis of MOlecular VAriance (AMOVA), were assessed by Arlequin v3.5 [23]. A multidimensional scaling (MDS) plot of the pairwise F_{ST} matrix was represented using the software STATISTICA v7.0. (Statsoft, Tulsa, Oklahoma; <http://www.statsoft.com/>).

The apportionment of genetic ancestral contributions from the different regions of Brazil was estimated using the STRUCTURE v2.3.3 software [24]. To estimate the ancestral membership proportions in the studied populations, a supervised analysis was performed using prior information on the geographic origin of the reference samples. Considering the historical formation of Brazil's lineage, we assumed an essentially tri-hybrid contribution from Native Americans, Europeans and Africans (i.e., $K=3$) to the current genetic makeup of Brazilian populations. STRUCTURE runs consisted of 100,000 burnin steps followed by 100,000 Markov Chain Monte Carlo (MCMC) iterations. The option "*Use population Information to test for migrants*" was used with the Admixture model; allele frequencies were correlated and updated using only individuals with POPFLAG=1 (in this case, the HGDP-CEPH samples used as reference).

Results

Genetic characterisation of diversity in Brazil's populations

The genotyping results for the 798 samples from 12 Brazilian populations are listed in Table S1. Allele frequencies and expected heterozygosities were estimated for the 46 AIM-Indels and are presented in Tables S2 and S3, respectively. As expected, the urban populations from Brazil show higher genetic variability than the parental populations because they harbor the contributions from three well-differentiated continental groups. The sample from the Terena native community has a similar level of diversity compared to the reference Native American sample, which is lower than that observed in urban populations. The same reduction in diversity was observed in the rural population of Santa Isabel, which is known to have been less exposed to European or African influx than the urban populations.

No statistically significant deviations from Hardy-Weinberg equilibrium expectations were found for the 46 *loci* in the 12 studied populations. Most Fisher's exact test p -values were above 1%; lower values were only observed in 11 out of the 552 tests ($0.00010 \leq p \leq 0.00764$), but these are not significant when applying Bonferroni's correction for multiple tests (significance level of 0.00009).

Genetic distance analysis

The data obtained for the 46 AIM-Indels in the 12 analysed samples were used to calculate the F_{ST} genetic distances between all population pairs together with the published data for the same markers in other Brazilian populations from Belém [3] and Rio de Janeiro [21] and from the parental African, European and Native American reference populations (Table S4). Figure 2 depicts the MDS plot of the pairwise F_{ST} matrix from Table S4.

Pairwise genetic distance analysis shows significant differentiation between most Brazilian samples and the ancestral populations, with the exception of two populations in the South (Santa Catarina and Paran), which present a low genetic distance when compared with the Europeans. In most comparisons within each geographic region, no significant genetic distances were found between urban populations; significant genetic distances were only obtained between urban and non-urban samples, as well as between Esprito Santo and two other populations in the Southeast region.

Santa Isabel and Terena samples show the lowest genetic distances to the Native Americans. The remaining Brazilian populations are all closer to the Europeans, although the positions in the MDS plot of the Northern populations of Manaus and Belm indicate a significant Native American contribution to these populations (Figure 2). In general, genetic distances to Native Americans are lower for populations in the North and higher for populations in the South. Conversely, the lowest genetic distances to Europeans are noted in Southern populations. Finally, the genetic composition of the Northeast, Center-West and Southeast regions is very similar, with slightly lower genetic distances to the Africans when compared to populations from the North or the South. The southeastern population from Esprito Santo is an exception and appears closer to the South, showing a lower genetic distance to Europeans than the other samples from the Southeast region.

Interethnic admixture analysis

The software STRUCTURE was used to estimate the ancestry membership proportions in populations from different regions in Brazil. The ancestry analyses were based on the genotypic data generated in the present study as well as previously published data for HGDP-CEPH reference samples from Africans, Europeans and Native Americans and the Brazilian populations of Belm and Rio de Janeiro [3,21]. The ancestry estimates obtained for all populations are shown in Figure 3.

The results are consistent with the genetic distance analysis. The Native American group of Terena and the non-urban Amazonian population of Santa Isabel exhibit a Native American contribution to their genetic pool that is above 75%.

European ancestry is the most prevalent in all urban populations, achieving the highest values (above 70%) in the three populations from the South. The populations in the North consist of a significant proportion of Native American ancestry that is approximately twice as high as that of African ancestry. Conversely, in the Northeast, Center-West and Southeast, the African contribution was the second most prevalent. As previously corroborated by the results of genetic distance analysis, the sample from Espírito Santo better fits the genetic profile of the populations in the South than those in the Southeast region.

At an intrapopulation level, heterogeneity is observed in all urban populations (Figure 4), with a very wide range of variation of ancestry proportions between individuals within populations. Nevertheless, individuals with a high proportion of Native American ancestry are only found in the samples from Terenas and Santa Isabel.

Discussion

A panel of 46 AIM Indels was recently described by Pereira *et al.* [3] to show marked allele frequency differentiation among main human population groups and proved to be highly informative for inferring ancestry. This panel of ancestry-informative Indels was used to characterise and compare the genetic composition of over 1,300 individuals from 14 populations among the five geopolitical regions in which Brazil is usually subdivided; to our knowledge, this represents the most comprehensive nationwide ancestry assessment undertaken using autosomal AIMs.

In the present work, a north-south decreasing F_{ST} gradient was noticeable between the Brazilian and the European samples, which appears to be directly correlated with the increasing European membership proportions depicted in the ancestry analysis. On the other hand, the lower distances between Native Americans and the populations from the North are in agreement with the slightly higher Native American composition that was detected in those populations. Additionally, in populations from the Northeast, Center-West and Southeast, lower F_{ST} genetic distances were detected when compared with Africans together with higher African ancestry proportions. In general, the genetic distances are low among samples from large urban Brazilian populations, but in many cases, they are statistically significant (Table S4). The low

differentiation associated with a high European ancestry that is observed in large urban populations cannot be extrapolated to smaller and more isolated rural populations or to the significant number of Native or Afro-Brazilian communities throughout the nation. A good example of this is the ancestry pattern observed in the three populations located in the North region that were included in this study. In these populations, the proportion of Native American ancestry increases in the smaller and more isolated populations. Indeed, it is almost 10% higher in Manaus than in Belém, and much higher in Santa Isabel than in Belém or Manaus. Different native groups from Amazonia have contributed to the population of Santa Isabel, which is indeed highly ethnically admixed and is not a native community established by a single ethno-linguistic group as is the case for Terena in Mato Grosso. Nevertheless, these two populations harbor similar African, European and Native American contributions, which emphasises the importance of geographic isolation and cultural barriers in developing the substructure within the main geopolitical regions.

These findings are consistent with the demographic patterns depicted throughout the country ([9]; Brazilian Institute of Geography and Statistics (IBGE); <http://www.ibge.gov.br>). However, IBGE statistics are based on colour classification, which precludes a strict correlation between demography and genetics. Indeed, past investigations have shown a weak correlation between skin colour or self-declared ethnicity and the genetic ancestry of individuals obtained from the characterisation of Ancestry Informative Markers (e.g., [25,26,27,28,29]).

Demographic significance of the selected samples

According to the last IBGE population survey in 2012, Brazil has approximately 194 million inhabitants distributed among the five geographic regions (<http://www.ibge.gov.br/home/estatistica/populacao/estimativa2012>). Most populations are highly admixed, and approximately 84% inhabitants live in large urban cities. In urban areas, only 0.2% of the population is Native American. The remaining 16% lives in rural areas, where Native Americans represent almost 2% of the population. Additionally, in rural areas, there are remnants of communities that were originally formed by fugitive slaves, known as Quilombos. These Afro-Brazilian communities are not demographically very significant, although they are numerous. Of the 1,826 existing communities that are dispersed all over the country, only 190 are currently officially recognised, consisting of just 11,946 families. Based on data from the IBGE, these communities occupy 0.12% of the national territory.

To maximise the collection of genetic diversity present in each region, we have selected samples from the main urban cities, where people are more concentrated and where the development of nearby small rural communities tends to genetically contribute. Samples from a native community in Mato Grosso do Sul and from a small rural population in Amazonia were also included in our study because Native American communities in Brazil represent a non-negligible fraction of the population.

Overall, considering the urban population samples that we studied representative of their States and taking into account the number of inhabitants in the IBGE demographic census (see Table S5 for details), we attained near complete coverage in the South and Southeast regions (with the exception of Native and Afro-descendant communities), 70% coverage in the North and only ~20% in the Northeast and the less populous Central-West regions. As a whole, the 14 populations that were analysed in the present study represent approximately 70% of the global Brazilian population.

Genetic ancestry of the Brazilian populations revisited

Many attempts have been made to determine the ethnic diversity of Brazil, to infer patterns of variation throughout the country and to note differences among communities with particular histories. Furthermore, many different types of markers have been used to pursue this objective.

Studies on uniparental markers were especially useful to discern male- and female-specific features. Nevertheless, for most studies on Y chromosome diversity in Brazilian populations, only Y-STRs (e.g., [19,30,31,32,33]) or just a restricted number of SNP markers (e.g., [26,34,35]) were studied, which often limited the accuracy of ancestry estimates derived from the three continents. Regarding the studies on the mtDNA variability in Brazilian populations, a very high percentage has been dedicated to small communities of Native Americans and Afro-descendants (e.g., [30,31]), but only few describe the composition of urban admixed populations [16,36]. In general, lineage markers have shown that in almost all Brazilian populations studied until now, the admixture was characterised by an asymmetric mating pattern occurring preferentially between European men and Native American or African women (e.g., [37,38]). In the Afro-descendant communities (known as Quilombos) in Amazonia and Maranhão, an asymmetric mating pattern was also observed between African men and Native American women (e.g., [15,39]).

Various studies were undertaken using autosomal markers, including blood groups and proteins and, more recently, polymorphisms representing different types of DNA variation, namely, STRs, SNPs and Indels (e.g., [3,11,12,20,40]). In general, they indicated a stronger Native American influence in the north. Significant African input in the northeast region was demonstrated, whereas the European contribution was rather global, following an increasing north-south gradient.

The same pattern of genetic variation throughout the country was observed in the present study, although we could find slight differences between our estimates and those from Callegari-Jacques *et al.* [10], Godinho *et al.* [11], Lins *et al.* [20] and Pena *et al.* [40] (see Figure 5). The observed discrepancies can be attributed to the number and/or type of markers and the different sampling strategies that were used. Unfortunately, such conceptual particularities in the design of each study that we reviewed hindered a direct and more comprehensive comparison between ancestry studies performed in Brazil. Nonetheless, a global overview of all studies shows a major European contribution across all regions of the country, despite some variation in their estimations. Furthermore, a general concordance extends to the regional level. The Northeast, Center-West, Southeast and South all reveal an admixed pattern of mainly European ancestry followed by African and Native American genetic influences. The exception occurs in the North, where the Native American membership proportion was higher than the African proportion in some studies [10,40, this study] compared with others [11,20]. Our study had the peculiarity of yielding lower European ancestry estimates while at the same time depicting stronger Native American contributions. As already noted, some of these differences may be due to the number and/or type of markers used and different sampling strategies. We highlight the fact that we have used carefully selected AIMs to capture the greatest amount of genetic differentiation among ancestral parental groups; we have also collected the greatest quantity of samples involved in a single ancestry assessment study in Brazil.

The relevance of sampling strategies

In any genetic study, samples always represent a reference group and not a real population, unless they are absolutely randomly collected. Nevertheless, perfectly random population samples are not necessarily or not always the most useful ones, and sampling strategies must reflect the objective of a research project or application. For example, samples that include information on the grandparents' birthplace, although very useful to learn about

the history of a population, do not represent the actual living or residential population; thus, they have limited value for forensic or medical purposes.

Most studies on the ethnicity of South American populations aim to evaluate admixture processes occurring in separate groups within a population. Nevertheless, the classification of groups is complicated, and different studies can use different criteria (e.g., self-declaration or ethnic classification by the researchers). Moreover, phenotypic classifications are usually subjective, and a person can self-report himself into different groups at different time points. Alternatively, people can classify others in different ways (e.g., [26,28,41,42]). A poor correlation between phenotypic/ethnic classifications and genetic ancestry has been found in most studies of Brazilian population groups, which prevents extrapolation of results based on some groups to the overall population. For instance, it was demonstrated that groups with equivalent self-reported proportions, but from populations with different levels of urbanisation [28] or from different regions of Brazil [40], can have different genetic ancestry profiles.

For the reasons mentioned above, the samples included in our study were randomly selected from unrelated individuals living in one of the targeted populations without using any criteria related to individual ethnicity.

In many publications that evaluate population ancestry, a frequent issue is the nature of the sampling. Samples are frequently selected from hospitals, universities, genetic diagnostic units, or other entities that do not necessarily represent the whole population, as the European, Native American and African genetic ancestry proportions can be affected by educational or socioeconomic status (e.g., [10,37]). This was also the case in the present study, and we believe, therefore, that this was the main reason for the discrepant results we found for the population of Espírito Santo. The samples had been collected from students and professors in a university, which was pointed out before to lead to an increased European proportion (e.g., [10]). For that reason, we have to carefully interpret the lower African and higher European ancestries that we found in Espírito Santo in comparison to the other three populations from the Southeast region, namely Minas Gerais, Rio de Janeiro and São Paulo. Thus, it would be desirable to better scrutinise the ethnic substructure of the Southeast by studying other source samples from Espírito Santo. The consistency of the results obtained among populations and within groups of populations, in addition to the expectations based on the available genetic and demographic data, precludes any significant distortion of the results obtained in the remaining populations we have studied.

Conclusion

Overall, the present results highlight the considerable amount of ethnic admixture that occurred throughout the country and that the current Brazilian population is significantly differentiated from the ancestral Native American, European and African populations that have been incorporated into its genepool during the last five centuries. The arrival of the Europeans stands out as the major event that shaped the genetic landscape of Brazil, which is especially evident in urban populations. Although demographically less representative of Brazil and spread throughout the country, there are still many rural populations and native communities that have maintained a strong Native American background. The African legacy is also a hallmark of all Brazilian populations, and in the present study represents more than 25% of the genetic makeup of most urban populations with the exception of the Northern and Southern regions (from 11 to 18%).

Apart from the variation found among populations throughout the country, our study also underscores the tremendous diversity that can be found among individuals from the same population in terms of the three different ethnic contributions (a wide spectrum can be observed, ranging from a more or less balanced contribution from two or three sources to an almost single source ancestry).

In accordance with previous studies, our results emphasise the heterogeneity of the Brazilian population from within and among populations, which justifies the need for additional studies involving more markers in new populations to allow an accurate assessment of the genetic ancestry for each subpopulation.

Author Contributions

Conceived and designed the work: FSNM, RP, EFC, LG. Collected samples: ARBA, DG, DAS, EVWS, IMP, JA, MO. Performed the experiments: FSNM, RV. Analysed and interpreted the data: FSNM, RP, LG. Drafting the article: FSNM, RP, LG. All authors reviewed the article critically for important intellectual content and final approval of the version to be published.

References

1. Parra EJ, Marcini A, Akey J, Martinson J, Batzer MA, *et al.* (1998) Estimating African American admixture proportions by use of population-specific alleles. *Am J Hum Genet* 63: 1839-1851.
2. Kosoy R, Nassir R, Tian C, White PA, Butler LM, *et al.* (2009) Ancestry informative marker sets for determining continental origin and admixture proportions in common populations in America. *Hum Mutat* 30: 69-78.
3. Pereira R, Phillips C, Pinto N, Santos C, Santos SEB, *et al.* (2012) Straightforward inference of ancestry and admixture proportions through ancestry-informative insertion deletion multiplexing. *PLoS One* 7: e29684.
4. Pritchard JK, Donnelly P (2001) Case-control studies of association in structured or admixed populations. *Theor Popul Biol* 60: 227-237.
5. Zembrzuski VM, Callegari-Jacques SM, Hutz MH (2006) Application of an African Ancestry Index as a genomic control approach in a Brazilian population. *Ann Hum Genet* 70: 822-828.
6. Tian C, Gregersen PK, Seldin MF (2008) Accounting for ancestry: population substructure and genome-wide association studies. *Hum Mol Genet* 17: R143-150.
7. Amorim CEG, Falcão-Alencar G, Godinho NMO, Diniz MECG, Gontijo CC, *et al.* (2009) Forensic application of an individual ancestry index in Brazilian populations. *Forensic Science International: Genetics Supplement Series* 2: 479-480.
8. Kayser M, de Knijff P (2011) Improving human forensics through advances in genetics, genomics and molecular biology. *Nat Rev Genet* 12: 179-192.
9. Instituto Brasileiro de Geografia e Estatística (2000) *Brasil: 500 anos de povoamento/Brazil - 500 years of settlement*. Rio de Janeiro: IBGE. 232 p.
10. Callegari-Jacques SM, Grattapaglia D, Salzano FM, Salamoni SP, Crossetti SG, *et al.* (2003) Historical genetics: spatiotemporal analysis of the formation of the Brazilian population. *Am J Hum Biol* 15: 824-834.
11. Godinho NMO, Gontijo CC, Diniz MECG, Falcão-Alencar G, Dalton GC, *et al.* (2008) Regional patterns of genetic admixture in South America. *Forensic Science International: Genetics Supplement Series* 1: 329-330.
12. Santos NP, Ribeiro-Rodrigues EM, Ribeiro-Dos-Santos AK, Pereira R, Gusmão L, *et al.* (2010) Assessing individual interethnic admixture and population substructure using a

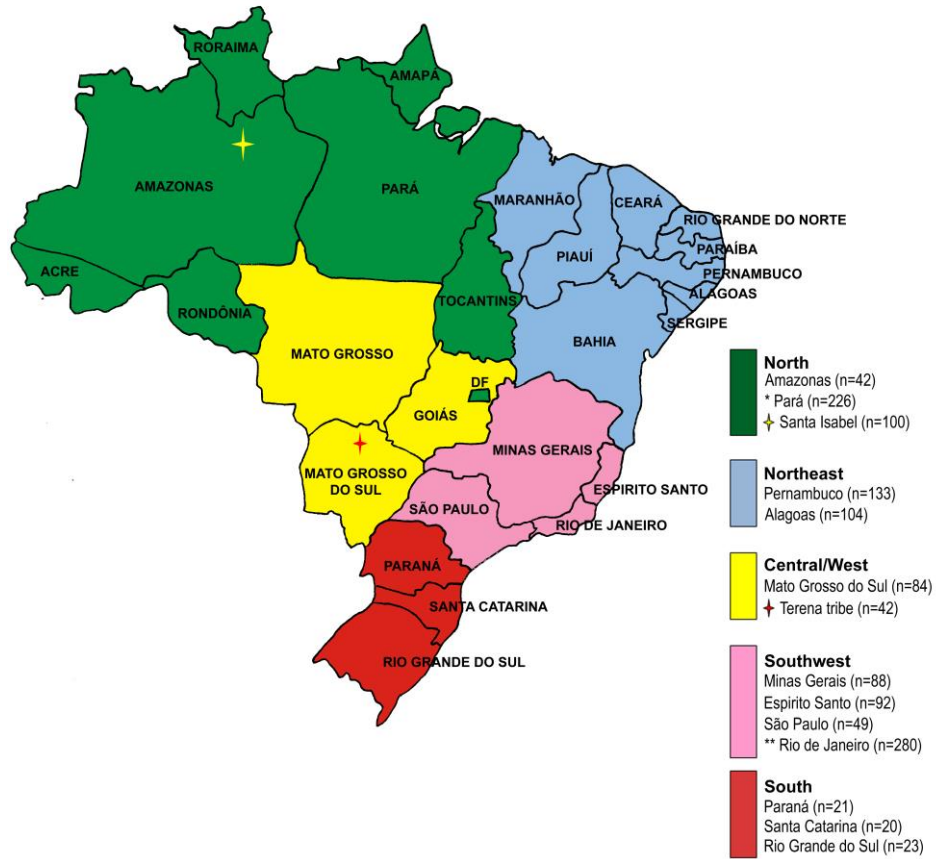
- 48-insertion-deletion (INSEL) ancestry-informative marker (AIM) panel. *Hum Mutat* 31: 184-190.
13. Schneider H, Salzano FM (1979) Gm allotypes and racial admixture in two Brazilian populations. *Hum Genet* 53: 101-105.
 14. Ribeiro-dos-Santos AK, Pereira JM, Lobato MR, Carvalho BM, Guerreiro JF, *et al.* (2002) Dissimilarities in the process of formation of Curiaú, a semi-isolated Afro-Brazilian population of the Amazon region. *Am J Hum Biol* 14: 440-447.
 15. Bortolini MC, Da Silva WAJWA, De Guerra DC, Remonato G, Mirandola R, *et al.* (1999) African-derived South American populations: A history of symmetrical and asymmetrical matings according to sex revealed by bi- and uni-parental genetic markers. *Am J Hum Biol* 11: 551-563.
 16. Alves-Silva J, da Silva Santos M, Guimaraes PE, Ferreira AC, Bandelt HJ, *et al.* (2000) The ancestry of Brazilian mtDNA lineages. *Am J Hum Genet* 67: 444-461.
 17. Sans M (2000) Admixture studies in Latin America: from the 20th to the 21st century. *Hum Biol* 72: 155-177.
 18. Silva DA, Carvalho E, Costa G, Tavares L, Amorim A, *et al.* (2006) Y-chromosome genetic variation in Rio de Janeiro population. *Am J Hum Biol* 18: 829-837.
 19. Francez PA, Ramos LP, Palha TdJBF, Santos SEB (2012) Haplotype diversity of 17 Y-str *loci* in an admixed population from the Brazilian Amazon. *Genet Mol Biol* 35: 45-52.
 20. Lins TC, Vieira RG, Abreu BS, Grattapaglia D, Pereira RW (2010) Genetic composition of Brazilian population samples based on a set of twenty-eight ancestry informative SNPs. *Am J Hum Biol* 22: 187-192.
 21. Manta FS, Pereira R, Caiafa A, Silva DA, Gusmão L, *et al.* (2013) Analysis of genetic ancestry in the admixed Brazilian population from Rio de Janeiro using 46 autosomal ancestry-informative Indel markers. *Ann Hum Biol* 40: 94-98.
 22. Manta F, Caiafa A, Pereira R, Silva D, Amorim A, *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.
 23. Excoffier L, Lischer HE (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* 10: 564-567.
 24. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155: 945-959.

25. Parra EJ, A. KR, Shriver MD (2004) Implications of correlations between skin color and genetic ancestry for biomedical research. *Nature Genetics* 36: S54-S56.
26. Marrero AR, Das Neves Leite FP, De Almeida Carvalho B, Peres LM, Kommers TC, *et al.* (2005) Heterogeneity of the genome ancestry of individuals classified as White in the state of Rio Grande do Sul, Brazil. *Am J Hum Biol* 17: 496-506.
27. Leite TK, Fonseca RM, de Franca NM, Parra EJ, Pereira RW (2011) Genomic ancestry, self-reported "color" and quantitative measures of skin pigmentation in Brazilian admixed siblings. *PLoS One* 6: e27162.
28. Lins TC, Vieira RG, Abreu BS, Gentil P, Moreno-Lima R, *et al.* (2011) Genetic heterogeneity of self-reported ancestry groups in an admixed Brazilian population. *J Epidemiol* 21: 240-245.
29. Cardena MMSG, Ribeiro-dos-Santos Â, Santos S, Mansur AJ, Pereira AC, *et al.* (2013) Assessment of the Relationship between Self-Declared Ethnicity, Mitochondrial Haplogroups and Genomic Ancestry in Brazilian Individuals. *PLoS ONE* 8: e62005.
30. Leite FP, Callegari-Jacques SM, Carvalho BA, Kommers T, Matte CH, *et al.* (2008) Y-STR analysis in Brazilian and South Amerindian populations. *Am J Hum Biol* 20: 359-363.
31. Palha TJ, Ribeiro-Rodrigues EM, Ribeiro-dos-Santos A, Guerreiro JF, Moura LS, *et al.* (2011) Male ancestry structure and interethnic admixture in African-descent communities from the Amazon as revealed by Y-chromosome Strs. *Am J Phys Anthropol* 144: 471-478.
32. Palha T, Gusmão L, Ribeiro-Rodrigues E, Guerreiro JF, Ribeiro-Dos-Santos A, *et al.* (2012) Disclosing the genetic structure of Brazil through analysis of male lineages with highly discriminating haplotypes. *PLoS One* 7: e40007.
33. Wiezel CE, Luizon MR, Sousa SM, Santos LM, Muniz YC, *et al.* (2013) Y-Linked microsatellites in Amazonian Amerindians applied to ancestry estimates in Brazilian Afro-derived populations. *Am J Hum Biol* 25: 313-317.
34. Santos SEB, Rodrigues JD, Ribeiro-dos-Santos AK, Zago MA (1999) Differential contribution of indigenous men and women to the formation of an urban population in the Amazon region as revealed by mtDNA and Y-DNA. *Am J Phys Anthropol* 109: 175-180.

35. Hunemeier T, Carvalho C, Marrero AR, Salzano FM, Junho Pena SD, *et al.* (2007) Niger-Congo speaking populations and the formation of the Brazilian gene pool: mtDNA and Y-chromosome data. *Am J Phys Anthropol* 133: 854-867.
36. Barbosa AB, da Silva LA, Azevedo DA, Balbino VQ, Mauricio-da-Silva L (2008) Mitochondrial DNA control region polymorphism in the population of Alagoas state, north-eastern Brazil. *J Forensic Sci* 53: 142-146.
37. Salzano FM (2004) Interethnic variability and admixture in Latin America--social implications. *Rev Biol Trop* 52: 405-415.
38. Pena SD, Bastos-Rodrigues L, Pimenta JR, Bydlowski SP (2009) DNA tests probe the genomic ancestry of Brazilians. *Braz J Med Biol Res* 42: 870-876.
39. Carvalho BM, Bortolini MC, Santos SEBd, Ribeiro-dos-Santos ÂKC (2008) Mitochondrial DNA mapping of social-biological interactions in Brazilian Amazonian African-descendant populations. *Genetics and Molecular Biology* 31: 12-22.
40. Pena SD, Di Pietro G, Fuchshuber-Moraes M, Genro JP, Hutz MH, *et al.* (2011) The genomic ancestry of individuals from different geographical regions of Brazil is more uniform than expected. *PLoS One* 6: e17063.
41. Domingues PM, Gusmão L, Silva DA, Amorim A, Pereira RW, *et al.* (2007) Sub-Saharan Africa descendents in Rio de Janeiro (Brazil): population and mutational data for 12 Y-STR *loci*. *Int J Legal Med* 121: 238-241.
42. Guerreiro-Junior V, Bisso-Machado R, Marrero A, Hunemeier T, Salzano FM, *et al.* (2009) Genetic signatures of parental contribution in black and white populations in Brazil. *Genet Mol Biol* 32: 1-11.

Figures:

Figure 1. Map of Brazil showing the geographical location of the populations considered in the present study.



* Pereira *et al.*, 2012 ** Manta *et al.*, 2013

Figure 2. MDS plot of the F_{ST} pairwise genetic distances between the studied populations (SI: Santa Isabel do Rio Negro; MA: Manaus; BE: Belém; PE: Pernambuco; AL: Alagoas; MS: Mato Grosso do Sul; TE: Terena; MG: Minas Gerais; ES: Espírito Santo; RJ: Rio de Janeiro; SP: São Paulo; PR: Paraná; SC: Santa Catarina; RS: Rio Grande do Sul).

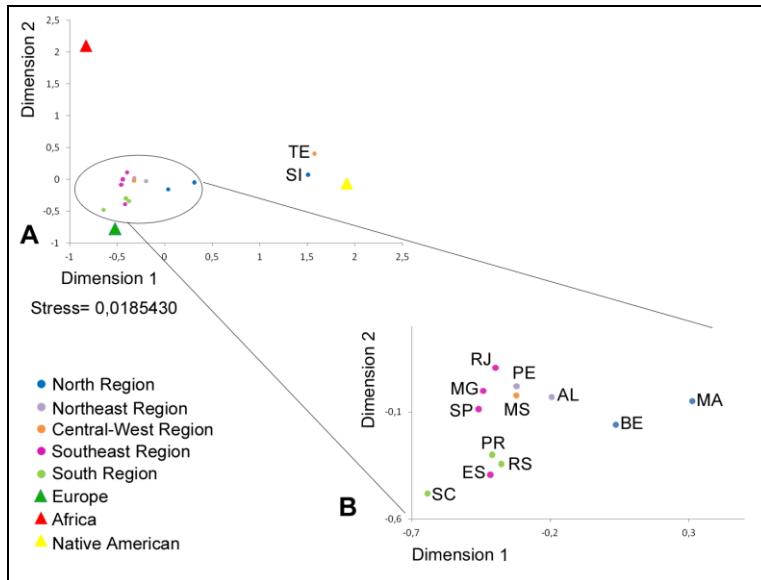
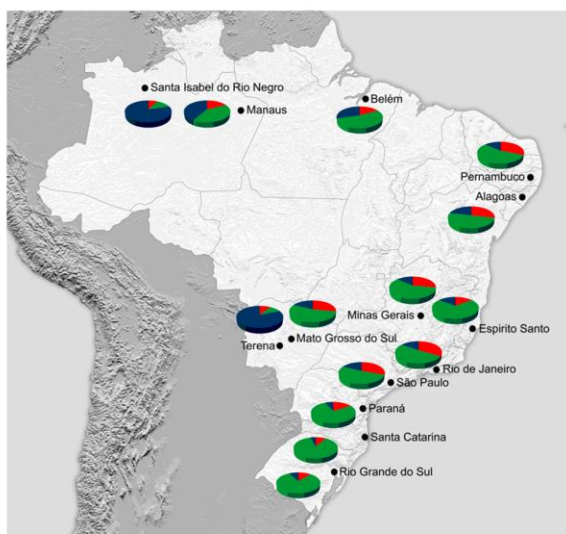


Figure 3. Average ancestral membership proportions obtained for the Brazilian testing populations using 46 AIM-Indels.



	Population samples	AFR	EUR	NAM	Ref.
North	Santa Isabel	0.074	0.168	0.758	This study
	Manaus	0.163	0.459	0.378	This study
	Belém	0.168	0.537	0.295	Pereira <i>et al.</i> , 2012
Northeast	Pernambuco	0.279	0.568	0.153	This study
	Alagoas	0.266	0.547	0.187	This study
Central-West	Mato Grosso do Sul	0.259	0.588	0.153	This study
	Terena	0.086	0.134	0.780	This study
Southeast	Minas Gerais	0.289	0.592	0.119	This study
	Espirito Santo	0.134	0.741	0.125	This study
	Rio de Janeiro	0.311	0.552	0.137	Manta <i>et al.</i> , 2013
	São Paulo	0.255	0.629	0.116	This study
South	Paraná	0.175	0.710	0.115	This study
	Santa Catarina	0.114	0.797	0.089	This study
	Rio Grande do Sul	0.140	0.729	0.130	This study

■ European
■ African
■ Native Americans

Figure 4. Individual ancestry estimates obtained for the HGDP-CEPH reference samples and individuals tested from Brazilian populations using 46 AIM-Indels (AFR: Africa; EUR: Europe; NAM: Native American; SI: Santa Isabel do Rio Negro; MA: Manaus; BE: Belém; PE: Pernambuco; AL: Alagoas; MS: Mato Grosso do Sul; TE: Terena; MG: Minas Gerais; ES: Espírito Santo; RJ: Rio de Janeiro; SP: São Paulo; PR: Paraná; SC: Santa Catarina; RS: Rio Grande do Sul).

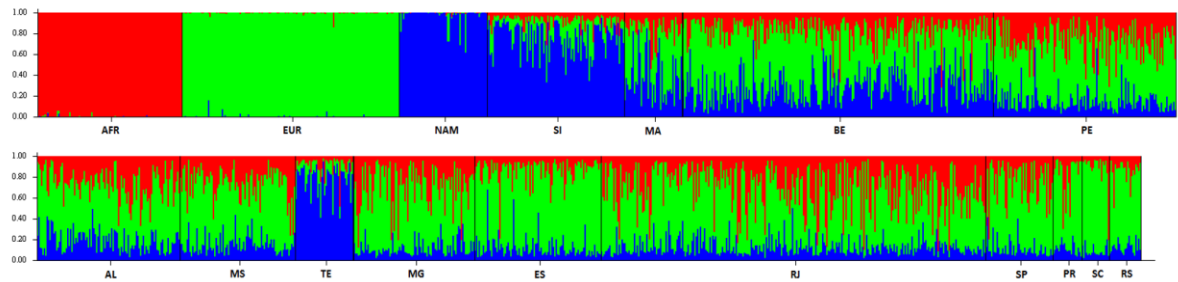
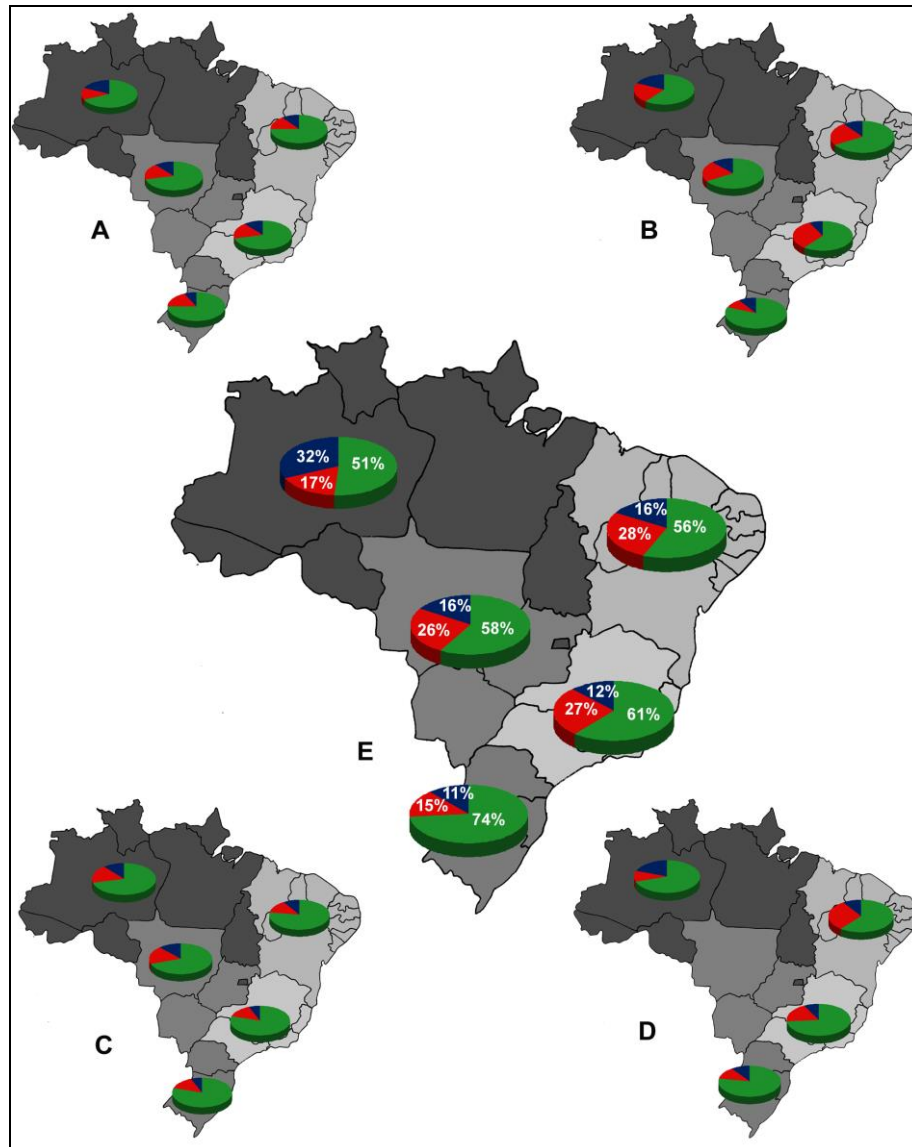


Figure 5. Comparison of the European, African and Native American ancestry estimates obtained in the present work and in previous studies for the five regions of Brazil. In the present work, the overall values indicated for each region are a weighted average of ancestry estimates of the population samples studied in that region considering their respective representation among inhabitants. A: Callegari-Jacques *et al.* [10]; B: Godinho *et al.* [11]; C: Lins *et al.* [20]; D: Pena *et al.* [40]; E: this study.



Supplementary file legends (ANEXO D)

Table S1. List of genotypes found in 12 different populations included in the present work as well as data from Rio de Janeiro, Brasil, previously published by Manta *et al.*, 2013.

Table S2. Allele frequencies for 46 AIM-Indels in twelve Brazilian populations.

Table S3. Gene diversities for 46 AIM-Indels in twelve Brazilians populations.

Table S4. Genetic distances (F_{ST}) between African, European, Native American and Brazilian populations (lower diagonal) and corresponding non-differentiation P values (upper diagonal).

Table S5. Demographic significance of sampled populations based in 2010 IBGE data for Brazilian Regions and States.

4 DISCUSSÃO

Os polimorfismos denominados Indels são variações de comprimento resultantes de inserção ou deleção de um ou mais nucleotídeos em uma sequência de DNA. Os marcadores Indels autossômicos apresentam uma elevada frequência no genoma humano e grandes vantagens em estudos populacionais e forenses em relação ao uso de outros tipos de marcadores, tais como a baixa taxa de mutação, *amplicons* menores (50 pares de base ou menos) e menor risco de interpretações incorretas observadas frequentemente em sistemas multialélicos. Além disso, por serem marcadores bialélicos, a genotipagem é facilitada, pois é necessário apenas observar se ocorreu a inserção ou a deleção. Com isso, o número de reações químicas necessárias para uma análise é menor, o que facilita os processos e diminui os custos (Pereira & Gusmão, 2011).

Neste estudo, utilizamos os Indels como ferramentas que permitissem verificar tanto a identificação individual, como possibilitasse avaliar proporções de ancestralidade a nível populacional e individual na população brasileira.

Na primeira etapa deste trabalho, foi avaliada a eficiência do sistema Indelplex-HID direcionado a identificação humana (Pereira *et al.*, 2009) na rotina forense da população brasileira. A utilização de um sistema multiplex composto por mini-Indels representa um ganho na rotina de análises criminais dentro de um laboratório forense, como é o caso do Laboratório de Diagnósticos por DNA da UERJ, onde o sucesso na amplificação de DNA degradado está relacionado com o menor comprimento dos *amplicons* gerados pelo sistema utilizado. A utilização de sistemas que reúnem *primers* relacionados com a detecção de pequenos fragmentos de DNA eleva as chances de se obter genotipagens positivas (Pereira & Gusmão, 2011).

A partir da possibilidade de amplificação de *loci* Indels em *amplicons* de comprimento reduzido, avaliamos a eficiência deste sistema na genotipagem de amostras *post-mortem*. Os resultados obtidos mostraram ser este sistema uma potente ferramenta para a amplificação de DNA degradado. Tal conclusão é reforçada pela elevação da eficiência da genotipagem de amostras que apresentaram amplificações parciais ou negativas, quando analisadas inicialmente com sistemas multiplexes de marcadores STRs (Identifiler® e Minifiler®), usuais em

laboratórios de rotina forense. Desse modo, ficou caracterizada a possibilidade de se aplicar o sistema Indelplex-HID para fins de genotipagem de amostras DNA de restos mortais como metodologia adicional ou complementar.

Ainda no contexto forense, este multiplex mostrou ser altamente informativo para a população do Rio de Janeiro, apresentando um alto poder de discriminação e uma maior eficiência em comparação com os valores encontrados para europeus e africanos utilizando os mesmos marcadores (Pereira *et al.*, 2009). Uma maior eficiência também foi observada quando comparados com valores obtidos para outras populações utilizando o sistema comercial DIPplex® kit (Qiagen) formado por 30 marcadores Indels de Identificação (LaRue *et al.*, 2013).

O poder de discriminação do sistema mostrou-se igualmente elevado para qualquer uma das amostras populacionais testadas, inclusive para os índios Terena, mesmo esta apresentando uma redução da diversidade, o que evidencia a eficiência deste multiplex até mesmo em populações ameríndias. Vale ressaltar que este resultado é de grande importância, além desta ser a primeira população ameríndia caracterizada para este sistema.

Os altos valores de diversidade média tanto da amostra correspondente à população geral do Rio de Janeiro como a composta apenas por indivíduos autodeclarados afrodescendentes estão de acordo com o esperado para populações que apresentam mistura genética, baseado nos dados apresentados por Pereira e colaboradores (2009) para populações ancestrais. Além disso, as diversidades genéticas não diferiram significativamente em ambas as amostras apesar de apresentarem diferentes proporções de mistura dos componentes ancestrais. Este resultado pode ser explicado pelos critérios utilizados na seleção dos marcadores Indels escolhidos para apontar diversidades elevadas, tanto em africanos quanto em europeus (Pereira *et al.*, 2009).

Na segunda etapa deste trabalho, foram utilizados marcadores autossômicos informativos de ancestralidade (46-AI-Indels) para investigar conjuntamente a história de ambas as linhagens. Até pouco tempo, a maioria dos estudos em relação à proporção de ancestralidade eram exclusivamente baseados em marcadores uniparentais do cromossomo Y e do DNA mitocondrial (Alves-Silva *et al.*, 2000; Góes *et al.*, 2002; Silva *et al.*, 2006, Domingues *et al.*, 2007; Bernardo *et al.*, 2013), o que

permitia a análise de apenas um lado da história, ou seja, a herança patrilinea ou matrilinea.

Pelas leis da genética, espera-se que as proporções de ancestralidade determinadas a partir de marcadores autossômicos correspondam aproximadamente aos valores médios indicados por estudos realizados por mtDNA e marcadores do cromossomo Y. Para confirmar esse princípio, os resultados obtidos a partir de Indels autossômicos foram comparados com aqueles observados com a utilização de marcadores do mtDNA e Y-SNPs na população do Rio de Janeiro.

Os dados relativos às ancestralidades materna e paterna no estado do Rio de Janeiro analisados por estudos isolados evidenciam um padrão de povoamento calçado na reprodução assimétrica, representada pela forte predominância da paternidade do homem europeu (88%) (Silva *et al.*, 2006) e da maternidade de mulheres africanas (60%) (Bernardo *et al.*, 2013). De fato, as estimativas de ancestralidade global obtidas com os 46-AI-Indels autossômicos (55% de contribuição europeia, 31% de contribuição africana e 14% de contribuição ameríndia) estão de acordo com os resultados esperados a partir dos valores obtidos pelos marcadores uniparentais.

Na continuação do estudo da ancestralidade, com o intuito de revelar detalhes sobre diferentes populações brasileiras e gerar dados não só para avaliar o grau de miscigenação, mas também a natureza histórica do processo, amostras de indivíduos residentes em diferentes estados foram genotipadas com o sistema 46-AI-Indels.

Em um contexto geral, em relação à ancestralidade europeia foi observada a existência de um gradiente crescente do norte ao sul do Brasil, onde esta foi a mais prevalente em todas as populações urbanas, como também demonstrado por diferentes grupos de pesquisa (Callegari-Jacques *et al.*, 2003; Godinho *et al.*, 2008; Lins *et al.*; 2010; Pena *et al.*, 2011). As pequenas diferenças, encontradas em relação às contribuições ancestrais nos estudos citados, estão provavelmente relacionadas à natureza da amostra, aos marcadores, assim como, à eficiência do sistema utilizado.

Em nível regional, as populações do Norte apresentaram maior proporção de ancestralidade ameríndia, a qual foi cerca de duas vezes superior em relação à contribuição africana. Por outro lado, no Nordeste, Centro-Oeste e Sudeste, a

ancestralidade africana foi a segunda mais prevalente. Em nível intrapopulacional, todas as populações urbanas se apresentaram altamente miscigenadas e a maioria da variação nas proporções de ancestralidade foi observada entre os indivíduos dentro de cada população, em vez de entre as populações.

Em relação às amostras da tribo Terena e da população de Santa Isabel foi observada uma alta proporção de ancestralidade indígena ratificando os dados históricos (IBGE, 2000).

Como esperado para uma população miscigenada, os presentes resultados não forneceram subsídios que nos permitam concluir que seja possível diferenciar "populações" de brasileiros brancos ou de brasileiros negros ou qualquer outro grupo. De maneira semelhante, embora existam denominações como afrodescendentes, eurodescendentes ou ameríndio descendentes, a maioria dos brasileiros estudados apresentou simultaneamente uma proporção de ancestralidade africana, europeia e ameríndia, quando analisados Indels autossômicos. Desta forma, o presente trabalho ressalta que a melhor classificação para indivíduos da população brasileira se encontra no âmbito de sua individualidade, pois cada indivíduo apresenta um genoma singular genealogicamente heterogêneo, característico da sua própria história evolutiva.

Assumindo um caráter eminentemente técnico e metodológico, os sistemas multiplexes formados por pequenos marcadores de inserção deleção se tornaram um sistema de análise genética simples, de fácil implementação e útil em uma ampla variedade de aplicações nos campos de Genética Populacional, Clínica e Forense.

5 CONCLUSÕES

- O sistema Indelplex-HID possibilitou a amplificação eficiente de 38 marcadores Indels autossômicos em uma única reação de PCR, permitindo a genotipagem e caracterização genética de duas populações brasileiras (Rio de Janeiro e Índios Terena) de forma rápida e requerendo pouca quantidade de amostra.
- O sistema multiplex destinado à Identificação Humana (Indelplex-HID) mostrou-se altamente informativo para as populações brasileiras consideradas miscigenadas, apresentando um alto poder de discriminação e uma elevada eficiência em casos forenses e em casos de investigação de parentesco biológico.
- Foram genotipados 798 indivíduos representantes das cinco macrorregiões geográficas do Brasil para 46 marcadores autossômicos informativos de ancestralidade, contribuindo assim, de maneira significativa para a caracterização da ancestralidade da população Brasileira.
- O sistema multiplex informativo de ancestralidade (46-AIM-Indel) demonstrou ser um método rápido e barato para estimar proporções de ancestralidade, mesmo em uma população miscigenada como a brasileira, tanto a nível individual quanto populacional. Demonstrando que independentemente do fenótipo cor da pele, a vasta maioria dos brasileiros tem um grau significativo de ancestralidade européia, ameríndia e africana.
- A mistura interétnica do Rio de Janeiro, capturada pelos AIM-Indels autossômicos mostrou-se similar ao que era esperado a partir de informações do DNA mitocondrial e do cromossomo Y.
- O presente estudo e outros realizados em diversas populações de distintas regiões do Brasil confirmam os relatos históricos sobre as diferenças regionais que ocorreram no processo de colonização do Brasil, assim como a formação tri-híbrida da nossa população.

- Foi comprovada a eficiência na classificação de pessoas de diversas origens biogeográficas, e por isso, o painel 46-AI-Indels além de se mostrar totalmente apto a ser utilizado em estudos de genética clínica e em contextos históricos, também se mostra interessante em aplicações forenses, em especial como ferramenta complementar durante uma investigação criminal.

REFERÊNCIAS

- Alves-Silva J, da Silva Santos M, Guimaraes PE, Ferreira AC, Bandelt HJ, Pena SD, *et al.* The ancestry of Brazilian mtDNA lineages. *Am J Hum Genet.* 2000; 67:444-461.
- Banco de dados de Indels. Disponível em: <http://www.marshfieldclinic.org/mgs/>
- Banco de dados de SNPs. Disponível em: <http://www.ncbi.nlm.nih.gov/projects/SNP/>
- Barnholtz-Sloan JS, McEvoy B, Shriver MD, Rebbeck TR. Ancestry estimation and correction for population stratification in molecular epidemiologic association studies. *Cancer Epidemiol Biomarkers Prev.* 2008; 17: 471-477.
- Bastos-Rodrigues L, Pimenta JR, Pena SD. The genetic structure of human populations studied through short insertion-deletion polymorphisms. *Ann Hum Genet.* 2006; 70: 658-665.
- Bernardo S, Hermida R, Desidério M, Silva D, Carvalho E F. mtDNA ancestry of Rio de Janeiro population, Brazil. *Mol Biol Reports.* 2013. (submetido)
- Britten RJ, Rowen L, Williams J. Majority of divergence between closely related DNA samples is due to Indels. *Proc Natl Acad Sci U S A.* 2003; 100: 4661-4665.
- Budowle B. SNP typing strategies. *Forensic Sci Int.* 2004; 146:139-142.
- Budowle B, van Daal A. Forensically relevant SNP classes. *Biotechniques.* 2008; 44: 603-608, 610.
- Butler JM. *Methods in molecular biology: Forensic DNA profiling protocols* Humana Press:Totowa,New Jersey. 2003.
- Butler JM. *Forensic DNA Typing – Biology, Technology and genetics of STR markers*, Burlington, London Elsevier Academic Press. 2005a.
- Butler JM. Constructing STR multiplex assays. *Methods in Molecular Biology.* 2005b; 297: 53-66.
- Butler JM, Coble MD, Vallone PM. STR vs SNPs: thoughts on the future of forensic DNA testing. *Forensic Science, Medicine, and Pathology.* 2007; 3: 200:205.
- Butler JM. *Fundamentals of Forensic DNA Typing*: Academic Press/Elsevier. 2009.
- Callegari-Jacques SM, Salzano FM. Brazilian Indian/ non-Indian interactions and their effects. *Ciênc Cult.* 1999; 51:166-174.

Callegari-Jacques SM, Grattapaglia D, Salzano FM, Salamoni SP, Crossetti SG, Ferreira ME, *et al.* Historical genetics: spatiotemporal analysis of the formation of the Brazilian population. *Am J Hum Biol.* 2003; 15:824-834.

Campbell MC, Tishkoff SA. African Genetic Diversity: Implications for Human Demographic History, Modern Human Origins, and Complex Disease Mapping. *Annu Rev Genomics Hum Genet.* 2008; 9: 403–433

Carracedo A, Sánchez-Diz P. Forensic DNA-Typing Technologies: A Review IN Carracedo A (Ed.) *Forensic DNA Typing Protocols. Methods Mol Biol.* 2005; 297:1-12.

Carvalho-Silva DR, Santos FR, Rocha J, Pena SDJ. The phylogeography of brazilian Y-chromosome lineages. *Am. J. Hum. Genet.* 2001; 68: 281-286.

Cavalli-Sforza LL, Menozzi P, Piazza A. History and geography of human genes. Princeton: Princeton University Press, 1994.

Charlesworth, B, Sniegowski P, Stephan W. The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature.* 1994; 371: 215-220.

Cunha MC, Salzano FM. História dos Índios no Brasil, Fundação de Amparo a Pesquisa do estado de São Paulo. 1992.

Da Silva GK, Guimarães R, Mattevi VS, Lazzaretti RK, Sprinz E, Kuhmmer R, *et al.* The role of mannose-binding lectin gene polymorphisms in susceptibility to HIV-1 infection in Southern Brazilian patients. *AIDS.* 2011; 25: 411-418.

Domingues PM, Gusmao L, da Silva DA, Amorim A, Pereira RW, de Carvalho EF. Sub-Saharan Africa descendants in Rio de Janeiro (Brazil): population and mutational data for 12 Y-STR *loci.* *Int J Legal Med.* 2007; 121:238-241.

Excoffier L, Laval G, Schneider S. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evol Bioinform Online.* 2005; 1:47-50.

Fundação Nacional do Índio - FUNAI. Disponível em: http://www.funai.gov.br/indios/fr_conteudo.htm. Acesso em: outubro de 2012.

Francez PAC, Rodrigues EM, Velasco AM, Santos SE. Insertion-deletion polymorphisms--utilization on forensic analysis. *Int J Legal Med.* 2012; 126:491-496

Gill P, Werrett DJ, Budowle B, Guerrieri R. An assessment of whether SNPs will replace STRs in national DNA databases—joint considerations of the DNA working group of the European Network of Forensic Science Institutes (ENFSI) and the Scientific Working group on DNA Analysis Methods (SWGDM). *Sci Justice.* 2004; 44: 51-53.

Góes ACS, Silva DA, Alvarez JC, Entrala C, Lorente JA, Carvalho EF. Identification of human remain by mitochondrial DNA analysis – detection of human remain by mitochondrial DNA analysis-detection of 25 polymorphisms. *Forensica*. 2002; 1: 45-48.

Góes ACS, Silva DA, Gil EHF, Silva MTD, Pereira RW, Carvalho EF. Allele frequencies data and statistic parameters of 16 *loci* – D19S433, D2S1338, CSF1PO, D16S539, D7S820, D21S11, D18S51, D13S317, D5S818, FGA, Penta E, TH01, vWA, D8S1179, TPOX, D3S1358 – in the Rio de Janeiro population, Brazil. *Forensic Sci Int*. 2004; 141:131-132.

Gusmão, L, Sánchez-Diz P, Alves C, Gomes I, Zarrabeitia MT, Abovich M, *et al*. A GEP-ISFG collaborative study on the optimization of an X-STR decaplex: data on 15 Iberian and Latin American populations. *International Journal of Legal Medicine*. 2009; 123: 227-234.

Hartl & Clark. *Principles of population genetics*. 3rd.ed. Sauer Associates, Sunderland. 1997.

Instituto Brasileiro de Geografia Estatística (IBGE). Disponível em: <http://www.ibge.gov.br>. Acesso em: dezembro de 2012.

Instituto Brasileiro de Geografia Estatística (IBGE). *BRASIL 500 Anos de povoamento*, IBGE, Rio de Janeiro. 2000.

Instituto Brasileiro de Geografia Estatística (IBGE) *Censo Demográfico 2000- Características da População e dos Domicílios- Resultados do universo*. Rio de Janeiro, 2012.

Instituto Brasileiro de Geografia Estatística (IBGE), *Os indígenas no Censo Demográfico 2010*. Rio de Janeiro, 2012.

International HapMap Consortium. The International HapMap Project. *Nature*. 2003; 426: 789-796.

International HapMap Consortium. A haplotype map of the human genome. *Nature* 2005; 437: 1299-1320.

International Human Genome Sequencing Consortium *Finishing the euchromatic sequence of the human genome*. *Nature*. 2004; 431: 931-945

Jeffreys AJ, Wilson V, Thein SL. Hypervariable 'minisatellite' regions in human DNA. *Nature*. 1985a; 314:67-73.

Jeffreys AJ, Wilson V, Thein SL. Individual-specific 'fingerprints' of human DNA. *Nature*. 1985b; 316:76-79.

Jobling MA, Pandey AA, Tyler-Smith C. The Y chromosome in forensic analysis and paternity testing. *Int. J. Legal Med*. 1997; 110: 118-124.

Jobling MA, Tyler-Smith C. The human Y chromosome: an evolutionary marker comes of age. *Nature Gen.* 2003; 4: 598-612.

Jobling MA, Gill P. Encoded evidence: DNA in forensic analysis. *Nat Rev Genet.* 2004; 5: 739-751.

Jobling, MA, Hurles ME, Tyler-Smith C. *Human Evolutionary Genetics: Origins, Peoples and Disease.* New York, Garland Science. 2004.

Jorde LB, Wooding SP. Genetic variation, classification and 'race'. *Nat Genet.* 2004; 36: S28-33.

Kayser M, de Knijff P. Improving human forensics through advances in genetics, genomics and molecular biology. *Nat Rev Genet.* 2011; 12: 179-192.

Kiittles RA, Weiss KM. Race, ancestry, and genes: implications for defining disease risk. *Annu Rev.Genomics Hum Genet.* 2003; 4:33-67.

Kobilinsky L, Liotti TF, Oeser-Sweat J. *DNA forensic and legal applications.* Wiley-Interscience, New Jersey. 2005

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J *et al.* Initial sequencing and analysis of the human genome. *Nature.* 2001; 409: 860-921.

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; 126:533-540.

Manta FSN. *Desenvolvimento e validação de sistemas multiplex de tipagem para identificação humana por DNA [Dissertação de mestrado].* Rio de Janeiro, RJ:Universidade do Estado do Rio de Janeiro, 2008; 128p.

Manta FSN, Silva DA, Sanchez-Diz P, Aguiar JI, Cabello PH, Carvalho EF. Terena Amerindian group autosomal STR data: comparison studies with other Brazilian populations. *Mol Biol Rep.* 2012; 39: 4455-4459.

Melton L. On the trail of SNPs. *Nature.* 2003; 422: 917-923.

Miller SA. Simple Salting out Procedure for extraction DNA from human nucleated cells. *Nucl. Acids Res.* 1988; 16: 1215.

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: 1182-1190.

Nei M. Genetic Distance between populations. *Amer. Nat.* 1972; 106: 283-292.

Nei M, Roychoudhury AK. Sampling variances of heterozygosity and genetic distance. *Genetics.* 1973; 76: 379-390.

Pakendorf B, Stoneking M. Mitochondrial DNA and human evolution. *Annu Rev Genomics Hum Genet.* 2005;6:165-183.

Parra FC, Amado RC, Lambertucci JR, Rocha J, Antunes CM, Pena SDJ. Color and genomic ancestry in Brazilians. *Proc Natl Acad Sci U S A.* 2003; 100: 177-82.

Pena SDJ, Carvalho-Silva DR, Alves-Silva J, Prado VF, Santos FR. Retrato molecular do Brasil. *Ciência Hoje.* 2000; 159:16-25.

Pena, S. Ministério da Ciência e Tecnologia. Centro de Gestão e Estudos Estratégicos (CGGE) Segurança Pública: Determinação de Identidade Genética Pelo DNA. 2004.

Pena SDJ, Bortolini MC. *Estudos Avançados.* 2004; 18.

Pena SDJ. Segurança Pública: determinação de identidade genética pelo DNA. *Parcerias estratégicas.* 2005; 20: 447-460.

Pena SDJ, Birchal TS. A inexistência biológica versus a existência social de raças humanas: pode a ciência instruir o etos social *REVISTA USP, São Paulo,* 2006; 68: 10-21.

Pena SD, Bastos-Rodrigues L, Pimenta and Bydlowski. DNA tests probe the genomic ancestry of Brazilians. *Braz J Med Biol Res.* 2009; 42: 870-876.

Pena SDJ, Di Pietro G, Fuchshuber-Moraes M, Genro JP, Hutz MH, Kehdy FSG. The Genomic Ancestry of Individuals from Different Geographical Regions of Brazil Is More Uniform Than Expected. *PLoS ONE.* 2011; 6: e17063.

Pereira R, Phillips C, Alves C, Amorim A, Carracedo A, Gusmão L. A new multiplex for human identification using insertion/deletion polymorphisms. *Electrophoresis.* 2009; 30: 3682-3690.

Pereira R, Gusmão L. Capillary electrophoresis of 38 non-coding biallelic mini-Indels for degraded samples and as complementary tool in paternity testing. *DNA In: Alonso A. Electrophoresis Protocols for Forensic Genetics. Humana Press.* 2011; 141-158.

Pereira R, Phillips C, Pinto N, Santos C, Santos SEB, Amorim A, *et al.* Straightforward inference of ancestry and admixture proportions through ancestry-informative insertion deletion multiplexing. *PLoS One.* 2012; 7: e29684.

Pereira RW, Pena SD. Phylogeography of haplotypes of five microsatellites located in a low-recombination region of the X chromosome: studies worldwide and in Brazilian populations. *Genetica.* 2006; 126: 243-250.

Phillips C, Lareu M, Salas A, Fondevila M, Berniell Lee G, Carracedo A, *et al.* Population specific single nucleotide polymorphisms. *Int Cong Ser.* 2004; 1261:233-235.

Phillips C, Salas A, Sánchez JJ, Fondevila M, Gómez-Tato A, Alvarez-Dios J, *et al.* SNPforID Consortium. Inferring ancestral origin using a single multiplex assay of ancestry-informative marker SNPs. *Forensic Sci Int Genet.* 2007; 1: 273-280.

Price AL, Patterson N, Yu F, Cox DR, Waliszewska A, McDonald GJ, *et al.* A genomewide admixture map for Latino populations. *Am J Hum Genet.* 2007; 80: 1024-1036.

Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics.* 2000; 155: 945-959.

Royal CDM, Duston GM. Changing the paradigm from 'race' to human genome variation. *Nat genet.* 2004; 36: S5-S7

Rosenberg NA, Mahajan S, Ramachandran S, Zhao C, Pritchard JK, Feldman MW. Clines, clusters, and the effect of study design on the inference of human population structure. *PLoS Genet.* 2005; 1: e70.

Salzano FM, Callegari-Jaques SM. *South American Indians: A case study in evolution.* Oxford: Clarendon Press.1988.

Sanchez JJ, Phillips C, Borsting C, Balogh K, Bogus M, Fondevila M, *et al.* A multiplex assay with 52 single nucleotide polymorphisms for human identification. *Electrophoresis.* 2006; 27:1713-1724.

Santos FR, Tyler-Smith C. Reading the human Y-chromosome: the emerging DNA markers and human genetic history. *Braz J Genet.* 1996; 19: 665-670.

Santos SEB, Ribeiro-dos-Santos AKC, Santos EJM e Guerreiro JF. The Amazonian microcosm. *Ciência e Cultura.* 1999; 51: 181-190.

Santos RV, Fry PH, Monteiro S, Maio MC, Rodrigues JC, Bastos-Rodrigues L, Pena SD. Color, race, and genomic ancestry in Brazil: dialogues between anthropology and genetics. *Curr Anthropol.* 2009; 50:787-819.

Santos NP, Ribeiro-Rodrigues EM, Ribeiro-dos-Santos AK, Pereira R, Gusmão L, Amorim A, *et al.* Assessing individual interethnic admixture and population substructure using a 48-insertion-deletion (INDEL) ancestry-informative marker (AIM) panel. *Hum Mutat.* 2010; 31: 184-190.

Schaffner SF. The X chromosome in population genetics. *Nat Rev Genet* 2004; 5: 43-51.

Schwartz M, Vissing J. New patterns of inheritance in mitochondrial disease. *Biochem Biophys Res Commun.* 2003; 310-247-51.

Shriver MD, Smith MW, Jin L, Marcini A, Akey JM, Deka R, *et al.* Ethnic-affiliation estimation by use of population-specific DNA markers. *Am J Hum Genet.* 1997; 60: 957-964.

- Shriver MD, Parra EJ, Dios S, Bonilla C, Norton H, Jovel C, *et al.* Skin pigmentation, biogeographical ancestry and admixture mapping. *Hum Genet.* 2003;112: 387-99.
- Silva DA, Carvalho E, Costa G, Tavares L, Amorim A.; Gusmão L. Y-chromosome genetic variation in Rio de Janeiro Population. *M. J. Hum. Biol.* 2006; 18: 829-837.
- Szibor R, Hering S, Kuhlisch E, Plate I, Demberger S, Krawczak M, *et al.* Haplotyping of STR cluster DXS6801-DXS6809-DXS6789 on Xq21 provides a powerful tool for kinship testing. *Int J Legal Med.* 2005; 119: 363-369
- Tian C, Gregersen PK, Seldin MF. Accounting for ancestry: population substructure and genome-wide association studies. *Hum Mol Genet.* 2008; 17: 143-150.
- Tishkoff S, Kidd KK. Implications of biogeography of human populations for “race” and medicine. *Nature Genetics Supplement.* 2004; 11: S21-S27.
- Tsai HJ, Kho JY, Shaikh N, Choudhry S, Naqvi M, Navarro D, *et al.* Admixture-matched case-control study: a practical approach for genetic association studies in admixed populations. *Hum Genet.* 2006; 118:626-39.
- Walsh PS, Metzgor D, Higuchi R. Chelex TM 100 as a medium for the simple extraction of DNA for PCR- based typing from forensic material. *Biotechniques.* 1991; 1: 91-98.
- Weber JL, David D, Heil J, Fan Y, Zhao C, Marth G. Human diallelic insertion/deletion polymorphisms. *Am J Hum Genet.* 2002; 71: 854-862.
- Vullo C, Borosky A, Catelli M, Romanini C, Fondevila M, Santos C *et al.* Population data for 38 autosomal insertion/deletion (Indels) and 50 SNPS polymorphisms in Argentinean population. *Forensic Sci Int Genetics.* 2011; 3: 419-420.

ANEXO A - Termo de Consentimento Livre e Esclarecido



UNIVERSIDADE DO ESTADO DO RIO DE JANEIRO
Instituto de Biologia Roberto Alcântara Gomes
LABORATÓRIO DE DIAGNÓSTICOS POR DNA
 CONVÊNIO: TRIBUNAL DE JUSTIÇA DO ESTADO DO RIO DE JANEIRO
 MINISTÉRIO PÚBLICO DO ESTADO DO RIO DE JANEIRO

FORMULÁRIO DE IDENTIFICAÇÃO E CONSENTIMENTO INFORMADO PARA UTILIZAÇÃO DE AMOSTRA BIOLÓGICA PARA ANÁLISE DE VÍNCULO GENÉTICO E DE ESTUDO POPULACIONAL

ORIGEM:	TJC	EXAME CODIGO:
DATA:	HORÁRIO:	
VARA:	PG/DNA:	PROT:
PROCESSO/ESC:	PROT. PRINCIPAL:	
OBS:		
OBJETIVO DESTE EXAME? CONFIRMAR OU NÃO A PATERNIDADE DE SP SOBRE SF		
HÁ ALGUM VÍNCULO GENÉTICO ENTRE SP E M ? () NÃO () SIM		
ALGUM PARENTE DE SP PODE TAMBÉM A SER CONSIDERADO SP ? () NÃO () SIM		
SÃO VIVOS? () Mãe de SP () Pai de SP () Mãe da Mãe () Pai da Mãe		
AUTORIZAÇÃO E CONCORDÂNCIA QUE AS AMOSTRAS BIOLÓGICAS COLETADAS SE DESTINAM À INVESTIGAÇÃO DE PATERNIDADE/MATERNIDADE PODENDO FAZER PARTE DE FORMA NÃO IDENTIFICADA DE ESTUDOS POPULACIONAIS E ANCESTRALIDADE		
SP	DATA DE NASC:	
RG	ÓRG. EMISSOR:	NATURALIDADE:
ASSINATURA	TEL:	
M	DATA DE NASC:	
RG	ÓRG. EMISSOR:	NATURALIDADE:
ASSINATURA	TEL:	
SF	DATA DE NASC:	
RG	ÓRG. EMISSOR:	NATURALIDADE:
ASSINATURA	TEL:	
ADVOGADOS:		
AUTOR:		
REU:		

ANEXO B - Aprovação do Comitê de Ética em Pesquisa



UNIVERSIDADE DO ESTADO DO RIO DE JANEIRO
HOSPITAL UNIVERSITÁRIO PEDRO ERNESTO
COMITÊ DE ÉTICA EM PESQUISA



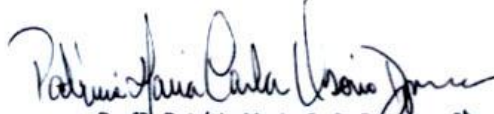
Rio de Janeiro, 14 de outubro de 2009

Do: Comitê de Ética em Pesquisa
Prof^a. Patrícia Maria C. O. Duque
Para: Coord. Prof. Elizeu Fagundes de Carvalho

O Comitê de Ética em Pesquisa do Hospital Universitário Pedro Ernesto, após avaliação, considerou o projeto (2485-CEP/HUPE – CAAE: 0067.0.228.000-09) "REGIÕES POLIMÓRFICAS DO DNA COMO FERRAMENTAS PARA ESTUDOS DE GENÉTICA DE POPULAÇÕES E FORENSE: UTILIZAÇÃO COMO MARCADORES DE ANCESTRALIDADE, MOVIMENTOS MIGRATÓRIOS, ESTUDOS SOBRE ESTRUTURAÇÃO GÊNICA DA POPULAÇÃO BRASILEIRA E INVESTIGAÇÕES CRIMINAIS E DE VÍNCULOS GENÉTICOS" aprovado, encontrando-se este dentro dos padrões éticos da pesquisa em seres humanos, conforme Resolução n.º196 sobre pesquisa envolvendo seres humanos de 10 de outubro de 1996, do Conselho Nacional de Saúde, bem como o consentimento livre e esclarecido.

O pesquisador deverá informar ao Comitê de Ética qualquer acontecimento ocorrido no decorrer da pesquisa.

O Comitê de Ética solicita a V. S^a., que ao término da pesquisa encaminhe a esta comissão um sumário dos resultados do projeto.


Prof^a Patrícia Maria C. O. Duque
Membro do Comitê de Ética em Pesquisa
COMITÊ DE ÉTICA EM PESQUISA
HUPE/UERJ

CEP - COMITÊ DE ÉTICA EM PESQUISA
AV. VINTE E OITO DE SETEMBRO, 77 TÉRREO - VILA ISABEL - CEP 20551-030
TEL: 21 2587-6353 - FAX: 21 2264-0853 - E-mail: cep-hupe@uerj.br

ANEXO C - Lista de soluções e tampões utilizadosTampão de lise de hemáceas (TLH)

NH ₄ CL	155 mM
KHCO ₃	10 mM
EDTA	1 mM

Tampão TE (pH 7,6)

Tris-Cl pH 7,6	10 mM
EDTA pH 8,0	1 mM

Tampão de Lise de núcleo (NLB)

Tris-Cl pH 8	10 mM
NaCl	0,4 mM
EDTA	2 mM

Tampão fosfato PBS (pH 7,4)

NH ₄ CL	155 mM
KHCO ₃	10 mM
NaCl	136 mM

Tampão de extração SEB (pH 8)

EDTA pH 8,0	10 mM
NaCl	100 mM
Tris-Cl pH 8,0	10 mM
SDS	2%

Multiplex Qiagen Master Mix 2x

HotStar Taq^R DNA Polimerase

Multiplex PCR Buffer pH 8,7

dNTP Mix

ANEXO D – Material Suplementar dos artigos publicados e submetido**ARTIGO I:****Indel markers: genetic diversity of 38 polymorphisms in Brazilian populations and application in a paternity investigation with post mortem material**

Forensic Science International: Genetics. 2012; 6 (5): 658-661.
<http://dx.doi.org/10.1016/j.fsigen.2011.12.008>,

Electronic Supplementary Material: **Manta et al. 2012**

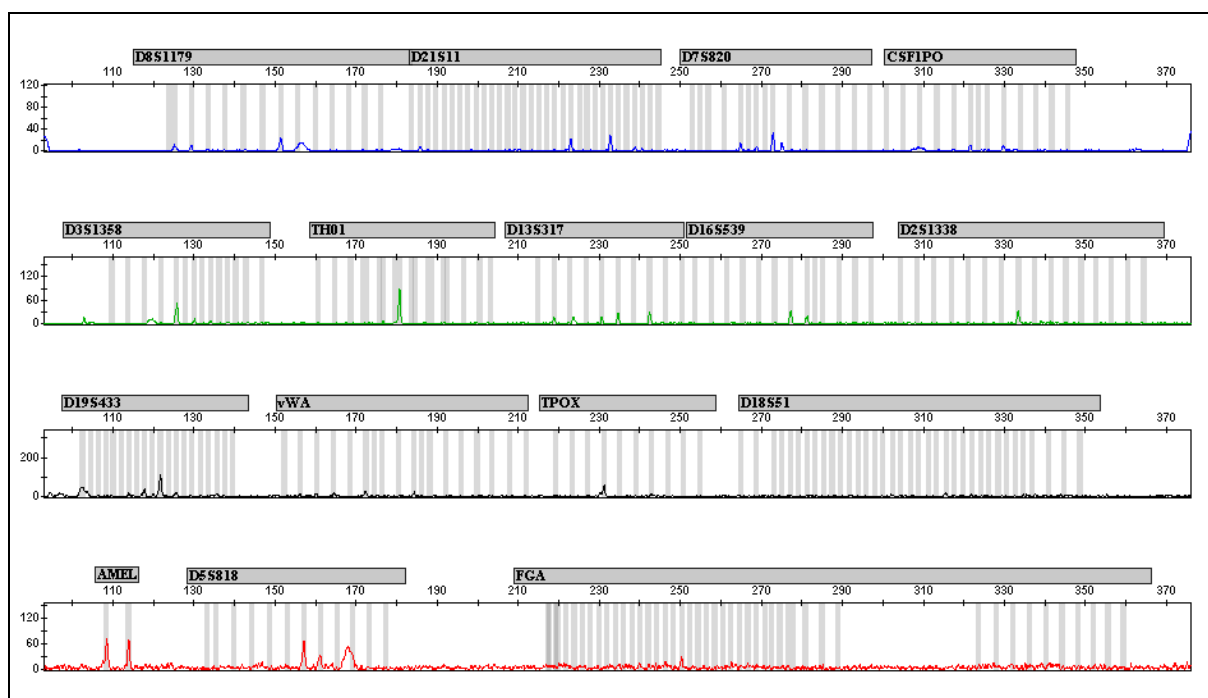


Figure S1 Alleged father (tooth sample) electropherogram using the AmpF/STR® Identifier kit (Applied Biosystems).

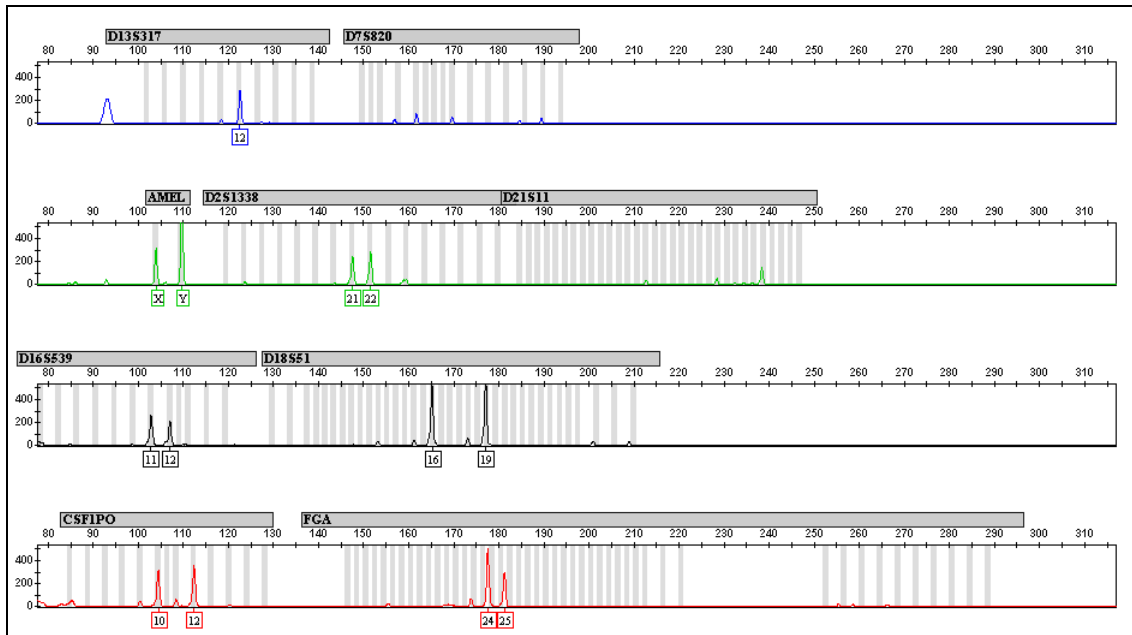


Figure S2 Alleged father (tooth sample) electropherogram using the AmpF/STR® Minifiler kit (Applied Biosystems).

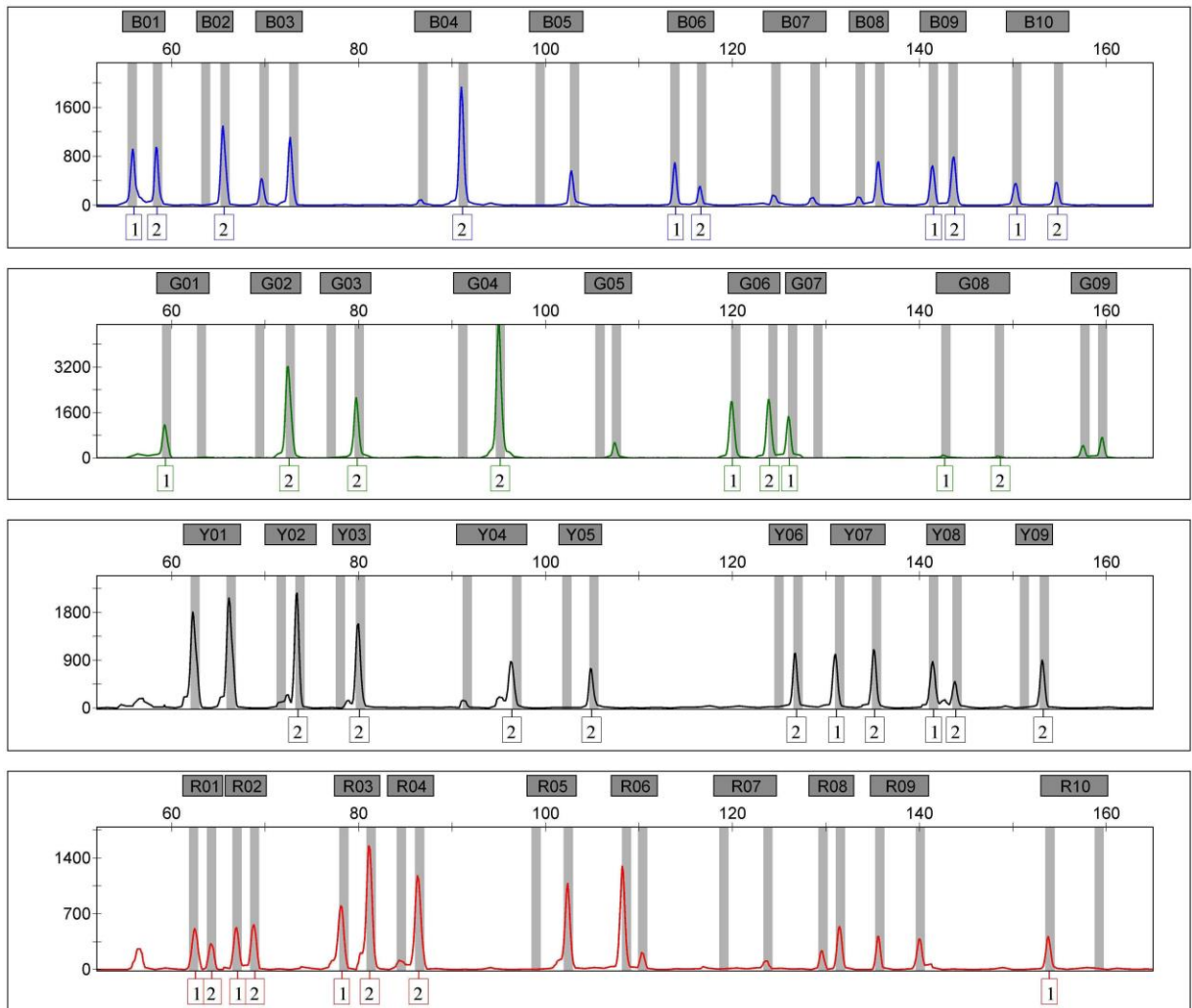


Figure S3 Alleged father (tooth sample) electropherogram using the HID Indel-plex [5]

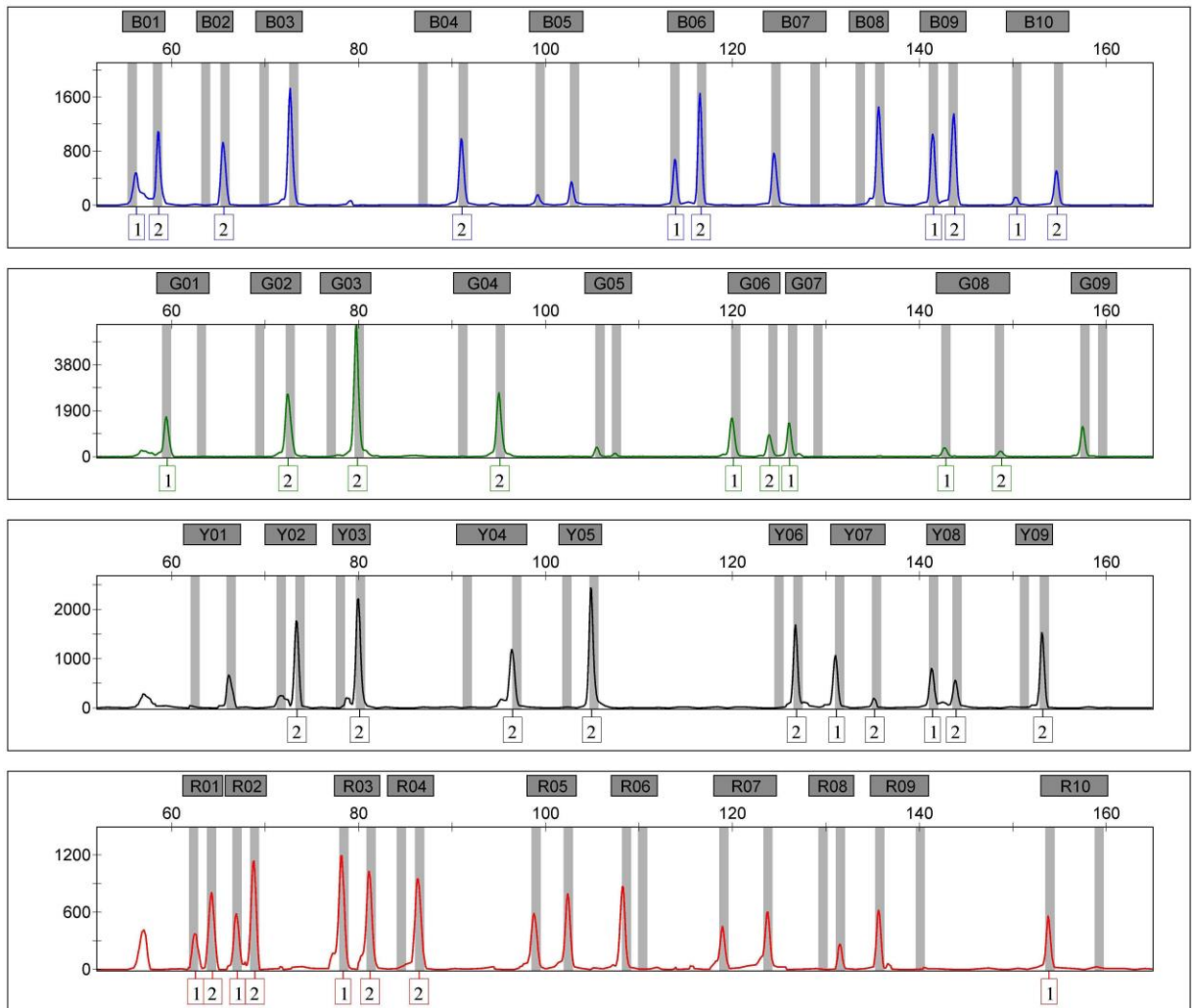


Figure S4 Alleged father (bone sample) electropherogram using the HID Indel-plex [5]

ARTIGO II:**Analysis of Genetic Ancestry in the Admixed Brazilian Population from Rio de Janeiro using 46 Autosomal Ancestry-Informative Indel Markers**

Annals of Human Biology, 2013; 40(1): 94-8. doi: 10.3109/03014460.2012.742138.

Electronic Supplementary Material: **Manta *et al.* 2013**

Supplementary Table I. Allele frequencies for the 46 AIM-Indels in Rio de Janeiro general population and self-declared Afro-descendants.

MID	rs number	Rio de Janeiro			Rio de Janeiro self-declared Afro- descendants		
		Short allele	Long allele	Variant allele	Short allele	Long allele	Variant allele
MID-1470	rs2307666	0.420	0.580	-	0.279	0.721	-
MID-777	rs1610863	0.313	0.687	-	0.336	0.664	-
MID-196	rs16635	0.512	0.488	-	0.473	0.527	-
MID-881	rs1610965	0.725	0.275	-	0.610	0.390	-
MID-3122	rs35451359	0.882	0.118	-	0.779	0.221	-
MID-548	rs140837	0.309	0.691	-	0.332	0.668	-
MID-659	rs1160893	0.157	0.843	-	0.208	0.792	-
MID-2011	rs2308203	0.602	0.398	-	0.478	0.522	-
MID-2929	rs33974167	0.739	0.261	-	0.735	0.265	-
MID-593	rs1160852	0.234	0.766	-	0.385	0.615	-
MID-798	rs1610884	0.436	0.564	-	0.429	0.571	-
MID-1193	rs2067280	0.136	0.864	-	0.208	0.792	-
MID-1871	rs2308067	0.213	0.787	-	0.119	0.881	-
MID-17	rs4183	0.459	0.541	-	0.588	0.412	-
MID-2538	rs3054057	0.266	0.734	-	0.212	0.788	-
MID-1644	rs2307840	0.687	0.313	-	0.527	0.473	-
MID-3854	rs60612424	0.220	0.780	-	0.345	0.655	-
MID-2275	rs3033053	0.216	0.784	-	0.274	0.726	-
MID-94	rs16384	0.179	0.821	-	0.137	0.863	-
MID-3072	rs34611875	0.707	0.293	-	0.460	0.540	-
MID-772	rs1610859	0.945	0.055	-	0.965	0.035	-
MID-2313	rs3045215	0.250	0.750	-	0.226	0.774	-
MID-397	rs25621	0.761	0.239	-	0.748	0.252	-
MID-1636	rs2307832	0.596	0.404	-	0.487	0.513	-
MID-51	rs16343	0.446	0.554	-	0.358	0.642	-
MID-2431	rs3031979	0.129	0.871	-	0.156	0.844	-
MID-2264*	rs34122827*	0.341	0.645	0.014	0.353	0.611	0.036
MID-2256	rs133052	0.213	0.787	-	0.177	0.823	-
MID-128	rs6490	0.241	0.759	-	0.226	0.774	-
MID-15	rs4181	0.416	0.584	-	0.367	0.633	-
MID-2241	rs3030826	0.400	0.600	-	0.571	0.429	-
MID-419	rs140708	0.846	0.154	-	0.850	0.150	-
MID-943	rs1611026	0.589	0.411	-	0.482	0.518	-
MID-159	rs16438	0.617	0.383	-	0.661	0.339	-
MID-2005	rs2308161	0.446	0.554	-	0.345	0.655	-
MID-250	rs16687	0.657	0.343	-	0.765	0.235	-
MID-1802	rs2307998	0.141	0.859	-	0.274	0.726	-
MID-1607	rs2307803	0.257	0.743	-	0.302	0.698	-
MID-1734	rs2307930	0.766	0.234	-	0.823	0.177	-
MID-406	rs25630	0.514	0.486	-	0.292	0.708	-
MID-1386	rs2307582	0.196	0.804	-	0.192	0.808	-
MID-1726	rs2307922	0.504	0.496	-	0.407	0.593	-
MID-3626	rs11267926	0.507	0.493	-	0.365	0.635	-
MID-360*	rs25584*	0.712	0.270	0.018	0.703	0.270	0.027
MID-1603	rs2307799	0.345	0.655	-	0.228	0.772	-
MID-2719	rs34541393	0.429	0.571	-	0.415	0.585	-

*For details on the variant alleles see Pereira *et al.* (2012).

Supplementary Table II. Pairwise F_{ST} genetic distances (below diagonal) and respective p-values (above diagonal) between HGDP-CEPH ancestral populations. Rio de Janeiro general population and self-declared Afro-descendants using 46 AIM-Indels.

	AFR	EUR	AME	RJ	RJ-AFD
AFR	-	$\leq 10^{-5}$	$\leq 10^{-5}$	$\leq 10^{-5}$	$\leq 10^{-5}$
EUR	0.36515	-	$\leq 10^{-5}$	$\leq 10^{-5}$	$\leq 10^{-5}$
AME	0.44273	0.29768	-	$\leq 10^{-5}$	$\leq 10^{-5}$
RJ	0.18190	0.05176	0.22025	-	$\leq 10^{-5}$
RJ-AFD	0.09974	0.12738	0.26216	0.02026	-

(AFR: African; EUR: European; AME: Amerindian; RJ: Rio de Janeiro; RJ-AFD: Rio de Janeiro self-declared Afro-descendants)

Supplementary Table III. Linkage Disequilibrium exact test significant results for Rio de Janeiro population (Bonferroni corrected significance level $\alpha < 0.05/1035$; 1035 pairwise tests performed).

<i>Loci pair</i>	<i>p</i> value	Standard error	Switches
MID3122 - MID3072**	0.000004	0.000001	51494152
MID2011 - MID3072	0.000032	0.000002	80578180
MID2538 - MID3072	0.000018	0.000001	74558953
MID3854 - MID3072	0.000023	0.000002	75987351
MID94 - MID2256**	0.000000	0.000000	58970353
MID3854 - MID2241**	0.000000	0.000000	79688890
MID2011 - MID419	0.000009	0.000001	67227966
MID881 - MID943**	0.000001	0.000001	79598562
MID1644 - MID1802	0.000045	0.000002	62608841
MID2005 - MID1802	0.000026	0.000002	68500776
MID1644 - MID1726**	0.000012	0.000001	81850023

** represent *loci* pairs located in contiguous positions in a same chromosome

ARTIGO III –**Revisiting the genetic ancestry of Brazilians using autosomal AIM-Indels**

Submetido à revista PLOS ONE (em maio de 2013).

A tabela Suplementar I não foi apresentada no corpo dessa tese devido a sua dimensão.

Table S1. List of genotypes found in 12 different populations included in the present work as well as data from Rio de Janeiro, Brazil, previously published by Manta *et al.*, 2013.

Table S2. Allele frequencies for 46 AIM-Indels in twelve Brazilian populations.

Markers			Santa Isabel			Manaus		
	MID	rs number	short allele	long allele	variant allele	short allele	long allele	variant allele
1	MID-1470	rs2307666	0,0700	0,9300	-	0,2619	0,7381	-
2	MID-777	rs1610863	0,6900	0,3100	-	0,4268	0,5732	-
3	MID-196	rs16635	0,1050	0,8950	-	0,3171	0,6829	-
4	MID-881	rs1610965	0,9350	0,0650	-	0,8056	0,1944	-
5	MID-3122	rs35451359	0,9850	0,0150	-	0,9524	0,0476	-
6	MID-548	rs140837	0,2100	0,7900	-	0,2250	0,7750	-
7	MID-659	rs1160893	0,1950	0,8050	-	0,1548	0,8452	-
8	MID-2011	rs2308203	0,7800	0,2200	-	0,7738	0,2262	-
9	MID-2929	rs33974167	0,9550	0,0450	-	0,7805	0,2195	-
10	MID-593	rs1160852	0,0150	0,9850	-	0,0854	0,9146	-
11	MID-798	rs1610884	0,1650	0,8350	-	0,2619	0,7381	-
12	MID-1193	rs2067280	0,2000	0,8000	-	0,1310	0,8691	-
13	MID-1871	rs2308067	0,4150	0,5850	-	0,2619	0,7381	-
14	MID-17	rs4183	0,8300	0,1700	-	0,5952	0,4048	-
15	MID-2538	rs3054057	0,1400	0,8600	-	0,2024	0,7976	-
16	MID-1644	rs2307840	0,3550	0,6450	-	0,5238	0,4762	-
17	MID-3854	rs60612424	0,0300	0,9700	-	0,0714	0,9286	-
18	MID-2275	rs3033053	0,4100	0,5900	-	0,2805	0,7195	-
19	MID-94	rs16384	0,4550	0,5450	-	0,2805	0,7195	-
20	MID-3072	rs34611875	0,9400	0,0600	-	0,8691	0,1310	-
21	MID-772	rs1610859	0,7424	0,2576	-	0,8718	0,1282	-
22	MID-2313	rs3045215	0,7500	0,2500	-	0,3537	0,6463	-
23	MID-397	rs25621	0,3800	0,6200	-	0,5488	0,4512	-
24	MID-1636	rs2307832	0,9750	0,0250	-	0,8333	0,1667	-
25	MID-51	rs16343	0,8100	0,1900	-	0,7143	0,2857	-
26	MID-2431	rs3031979	0,2700	0,7300	-	0,1829	0,8171	-
27	MID-2264	rs34122827*	0,1650	0,8200	0,0150	0,2024	0,7857	0,0119
28	MID-2256	rs133052	0,7539	0,2462	-	0,4762	0,5238	-
29	MID-128	rs6490	0,1900	0,8100	-	0,2976	0,7024	-
30	MID-15	rs4181	0,8050	0,1950	-	0,4524	0,5476	-
31	MID-2241	rs3030826	0,1000	0,9000	-	0,2683	0,7317	-
32	MID-419	rs140708	0,6250	0,3750	-	0,7073	0,2927	-
33	MID-943	rs1611026	0,4850	0,5150	-	0,5732	0,4268	-
34	MID-159	rs16438	0,8450	0,1550	-	0,6548	0,3452	-
35	MID-2005	rs2308161	0,2750	0,7250	-	0,3691	0,6310	-
36	MID-250	rs16687	0,4350	0,5650	-	0,5238	0,4762	-
37	MID-1802	rs2307998	0,0150	0,9850	-	0,1548	0,8452	-
38	MID-1607	rs2307803	0,4950	0,5050	-	0,3415	0,6585	-
39	MID-1734	rs2307930	0,5663	0,4337	-	0,7683	0,2317	-
40	MID-406	rs25630	0,7500	0,2500	-	0,7250	0,2750	-
41	MID-1386	rs2307582	0,5650	0,4350	-	0,3659	0,6342	-
42	MID-1726	rs2307922	0,6150	0,3850	-	0,6071	0,3929	-
43	MID-3626	rs11267926	0,0600	0,9400	-	0,3659	0,6342	-
44	MID-360	rs25584	0,5550	0,4450	0,0000	0,7683	0,2317	0,0000
45	MID-1603	rs2307799	0,7750	0,2250	-	0,5875	0,4125	-
46	MID-2719	rs34541393	0,6150	0,3850	-	0,5128	0,4872	-

Markers			Pernambuco			Alagoas		
	MID	rs number	short allele	long allele	variant allele	short allele	long allele	variant allele
1	MID-1470	rs2307666	0,4023	0,5977	-	0,4471	0,5529	-
2	MID-777	rs1610863	0,4248	0,5752	-	0,4327	0,5673	-
3	MID-196	rs16635	0,4173	0,5827	-	0,4327	0,5673	-
4	MID-881	rs1610965	0,7739	0,2261	-	0,8069	0,1931	-
5	MID-3122	rs35451359	0,8797	0,1203	-	0,8942	0,1058	-
6	MID-548	rs140837	0,2556	0,7444	-	0,2789	0,7212	-
7	MID-659	rs1160893	0,1654	0,8346	-	0,1635	0,8365	-
8	MID-2011	rs2308203	0,6353	0,3647	-	0,6346	0,3654	-
9	MID-2929	rs33974167	0,7368	0,2632	-	0,7115	0,2885	-
10	MID-593	rs1160852	0,2444	0,7556	-	0,2548	0,7452	-
11	MID-798	rs1610884	0,4737	0,5263	-	0,3884	0,6117	-
12	MID-1193	rs2067280	0,1805	0,8196	-	0,1408	0,8592	-
13	MID-1871	rs2308067	0,2331	0,7669	-	0,2212	0,7789	-
14	MID-17	rs4183	0,4549	0,5451	-	0,4615	0,5385	-
15	MID-2538	rs3054057	0,3196	0,6805	-	0,2837	0,7164	-
16	MID-1644	rs2307840	0,7269	0,2731	-	0,6893	0,3107	-
17	MID-3854	rs60612424	0,2068	0,7932	-	0,1923	0,8077	-
18	MID-2275	rs3033053	0,2099	0,7901	-	0,2308	0,7692	-
19	MID-94	rs16384	0,1880	0,8120	-	0,2981	0,7019	-
20	MID-3072	rs34611875	0,7105	0,2895	-	0,7933	0,2067	-
21	MID-772	rs1610859	0,9436	0,0564	-	0,9078	0,0922	-
22	MID-2313	rs3045215	0,2970	0,7030	-	0,3173	0,6827	-
23	MID-397	rs25621	0,6617	0,3384	-	0,7500	0,2500	-
24	MID-1636	rs2307832	0,6278	0,3722	-	0,6923	0,3077	-
25	MID-51	rs16343	0,4774	0,5226	-	0,4615	0,5385	-
26	MID-2431	rs3031979	0,1504	0,8496	-	0,1442	0,8558	-
27	MID-2264	rs34122827*	0,3722	0,5865	0,0414	0,3606	0,6010	0,0385
28	MID-2256	rs133052	0,1880	0,8120	-	0,3495	0,6505	-
29	MID-128	rs6490	0,2707	0,7293	-	0,2115	0,7885	-
30	MID-15	rs4181	0,4549	0,5451	-	0,4567	0,5433	-
31	MID-2241	rs3030826	0,4023	0,5977	-	0,3606	0,6394	-
32	MID-419	rs140708	0,8271	0,1729	-	0,8510	0,1490	-
33	MID-943	rs1611026	0,5611	0,4389	-	0,5769	0,4231	-
34	MID-159	rs16438	0,6429	0,3571	-	0,6154	0,3846	-
35	MID-2005	rs2308161	0,4351	0,5649	-	0,4471	0,5529	-
36	MID-250	rs16687	0,6955	0,3045	-	0,6731	0,3269	-
37	MID-1802	rs2307998	0,1316	0,8684	-	0,2019	0,7981	-
38	MID-1607	rs2307803	0,2481	0,7519	-	0,3269	0,6731	-
39	MID-1734	rs2307930	0,7731	0,2269	-	0,7959	0,2041	-
40	MID-406	rs25630	0,5338	0,4662	-	0,5721	0,4279	-
41	MID-1386	rs2307582	0,2143	0,7857	-	0,2644	0,7356	-
42	MID-1726	rs2307922	0,5379	0,4621	-	0,5865	0,4135	-
43	MID-3626	rs11267926	0,4659	0,5341	-	0,4856	0,5144	-
44	MID-360	rs25584	0,6917	0,2857	0,0226	0,7260	0,2308	0,0433
45	MID-1603	rs2307799	0,3872	0,6128	-	0,3894	0,6106	-
46	MID-2719	rs34541393	0,3459	0,6541	-	0,4039	0,5962	-

Markers		Mato Grosso do Sul			Terena			
	MID	rs number	short allele	long allele	variant allele	short allele	long allele	variant allele
1	MID-1470	rs2307666	0,4643	0,5357	-	0,0488	0,9512	-
2	MID-777	rs1610863	0,3691	0,6310	-	0,7738	0,2262	-
3	MID-196	rs16635	0,4691	0,5309	-	0,0833	0,9167	-
4	MID-881	rs1610965	0,7434	0,2566	-	0,9643	0,0357	-
5	MID-3122	rs35451359	0,8691	0,1310	-	0,9762	0,0238	-
6	MID-548	rs140837	0,3155	0,6845	-	0,4286	0,5714	-
7	MID-659	rs1160893	0,1429	0,8571	-	0,1548	0,8452	-
8	MID-2011	rs2308203	0,6548	0,3452	-	0,9643	0,0357	-
9	MID-2929	rs33974167	0,7083	0,2917	-	0,9524	0,0476	-
10	MID-593	rs1160852	0,2143	0,7857	-	0,0476	0,9524	-
11	MID-798	rs1610884	0,4643	0,5357	-	0,2262	0,7738	-
12	MID-1193	rs2067280	0,1369	0,8631	-	0,2738	0,7262	-
13	MID-1871	rs2308067	0,2619	0,7381	-	0,3333	0,6667	-
14	MID-17	rs4183	0,5417	0,4583	-	0,8929	0,1071	-
15	MID-2538	rs3054057	0,3036	0,6964	-	0,0476	0,9524	-
16	MID-1644	rs2307840	0,7744	0,2256	-	0,3659	0,6342	-
17	MID-3854	rs60612424	0,2083	0,7917	-	0,0595	0,9405	-
18	MID-2275	rs3033053	0,2073	0,7927	-	0,4881	0,5119	-
19	MID-94	rs16384	0,2441	0,7560	-	0,7619	0,2381	-
20	MID-3072	rs34611875	0,7202	0,2798	-	0,9286	0,0714	-
21	MID-772	rs1610859	0,9286	0,0714	-	0,7738	0,2262	-
22	MID-2313	rs3045215	0,2917	0,7083	-	0,7976	0,2024	-
23	MID-397	rs25621	0,7381	0,2619	-	0,3452	0,6548	-
24	MID-1636	rs2307832	0,6488	0,3512	-	0,8929	0,1071	-
25	MID-51	rs16343	0,4881	0,5119	-	0,8537	0,1463	-
26	MID-2431	rs3031979	0,1786	0,8214	-	0,3929	0,6071	-
27	MID-2264	rs34122827*	0,3659	0,5915	0,0427	0,1786	0,8095	0,0119
28	MID-2256	rs133052	0,2857	0,7143	-	0,7381	0,2619	-
29	MID-128	rs6490	0,3133	0,6868	-	0,2024	0,7976	-
30	MID-15	rs4181	0,3929	0,6071	-	0,8095	0,1905	-
31	MID-2241	rs3030826	0,3333	0,6667	-	0,1500	0,8500	-
32	MID-419	rs140708	0,8810	0,1191	-	0,6429	0,3571	-
33	MID-943	rs1611026	0,6191	0,3810	-	0,4524	0,5476	-
34	MID-159	rs16438	0,6369	0,3631	-	0,7561	0,2439	-
35	MID-2005	rs2308161	0,4451	0,5549	-	0,2143	0,7857	-
36	MID-250	rs16687	0,6386	0,3615	-	0,4524	0,5476	-
37	MID-1802	rs2307998	0,1325	0,8675	-	0,0238	0,9762	-
38	MID-1607	rs2307803	0,2901	0,7099	-	0,3810	0,6191	-
39	MID-1734	rs2307930	0,7308	0,2692	-	0,5357	0,4643	-
40	MID-406	rs25630	0,5127	0,4873	-	0,5952	0,4048	-
41	MID-1386	rs2307582	0,2560	0,7441	-	0,4881	0,5119	-
42	MID-1726	rs2307922	0,5741	0,4259	-	0,8333	0,1667	-
43	MID-3626	rs11267926	0,5188	0,4813	-	0,0714	0,9286	-
44	MID-360	rs25584	0,6369	0,3452	0,0179	0,6071	0,3810	0,0119
45	MID-1603	rs2307799	0,3210	0,6790	-	0,7262	0,2738	-
46	MID-2719	rs34541393	0,3563	0,6438	-	0,5714	0,4286	-

Markers			Minas Gerais			Espírito Santo		
	MID	rs number	short allele	long allele	variant allele	short allele	long allele	variant allele
1	MID-1470	rs2307666	0,4602	0,5398	-	0,5598	0,4402	-
2	MID-777	rs1610863	0,3239	0,6761	-	0,3587	0,6413	-
3	MID-196	rs16635	0,5227	0,4773	-	0,5330	0,4670	-
4	MID-881	rs1610965	0,7443	0,2557	-	0,8933	0,1067	-
5	MID-3122	rs35451359	0,9261	0,0739	-	0,9396	0,0604	-
6	MID-548	rs140837	0,3011	0,6989	-	0,2088	0,7912	-
7	MID-659	rs1160893	0,1705	0,8296	-	0,1196	0,8804	-
8	MID-2011	rs2308203	0,6307	0,3693	-	0,6868	0,3132	-
9	MID-2929	rs33974167	0,6818	0,3182	-	0,7065	0,2935	-
10	MID-593	rs1160852	0,2330	0,7671	-	0,1304	0,8696	-
11	MID-798	rs1610884	0,4943	0,5057	-	0,5000	0,5000	-
12	MID-1193	rs2067280	0,1477	0,8523	-	0,1848	0,8152	-
13	MID-1871	rs2308067	0,1875	0,8125	-	0,2802	0,7198	-
14	MID-17	rs4183	0,4830	0,5171	-	0,3791	0,6209	-
15	MID-2538	rs3054057	0,3239	0,6761	-	0,3500	0,6500	-
16	MID-1644	rs2307840	0,6875	0,3125	-	0,8820	0,1180	-
17	MID-3854	rs60612424	0,2557	0,7443	-	0,0815	0,9185	-
18	MID-2275	rs3033053	0,1932	0,8068	-	0,1250	0,8750	-
19	MID-94	rs16384	0,1818	0,8182	-	0,1902	0,8098	-
20	MID-3072	rs34611875	0,7046	0,2955	-	0,8478	0,1522	-
21	MID-772	rs1610859	0,9602	0,0398	-	0,9076	0,0924	-
22	MID-2313	rs3045215	0,2159	0,7841	-	0,2609	0,7391	-
23	MID-397	rs25621	0,7386	0,2614	-	0,7120	0,2880	-
24	MID-1636	rs2307832	0,6307	0,3693	-	0,7222	0,2778	-
25	MID-51	rs16343	0,4943	0,5057	-	0,5879	0,4121	-
26	MID-2431	rs3031979	0,0852	0,9148	-	0,1576	0,8424	-
27	MID-2264	rs34122827*	0,4659	0,4830	0,0511	0,4775	0,4494	0,0730
28	MID-2256	rs133052	0,2273	0,7727	-	0,1868	0,8132	-
29	MID-128	rs6490	0,2955	0,7046	-	0,3846	0,6154	-
30	MID-15	rs4181	0,4375	0,5625	-	0,4451	0,5550	-
31	MID-2241	rs3030826	0,3921	0,6080	-	0,2582	0,7418	-
32	MID-419	rs140708	0,8182	0,1818	-	0,7554	0,2446	-
33	MID-943	rs1611026	0,6364	0,3636	-	0,6868	0,3132	-
34	MID-159	rs16438	0,6149	0,3851	-	0,6576	0,3424	-
35	MID-2005	rs2308161	0,4943	0,5057	-	0,5337	0,4663	-
36	MID-250	rs16687	0,6477	0,3523	-	0,7337	0,2663	-
37	MID-1802	rs2307998	0,1421	0,8580	-	0,0598	0,9402	-
38	MID-1607	rs2307803	0,2841	0,7159	-	0,2167	0,7833	-
39	MID-1734	rs2307930	0,8068	0,1932	-	0,7989	0,2011	-
40	MID-406	rs25630	0,5171	0,4830	-	0,6902	0,3098	-
41	MID-1386	rs2307582	0,1818	0,8182	-	0,2228	0,7772	-
42	MID-1726	rs2307922	0,5341	0,4659	-	0,6522	0,3478	-
43	MID-3626	rs11267926	0,4773	0,5227	-	0,5989	0,4011	-
44	MID-360	rs25584	0,6989	0,2898	0,0114	0,7826	0,2174	0,0000
45	MID-1603	rs2307799	0,3182	0,6818	-	0,3901	0,6099	-
46	MID-2719	rs34541393	0,3750	0,6250	-	0,4130	0,5870	-

Markers			São Paulo			Paraná		
	MID	rs number	short allele	long allele	variant allele	short allele	long allele	variant allele
1	MID-1470	rs2307666	0,5000	0,5000	-	0,5476	0,4524	-
2	MID-777	rs1610863	0,3469	0,6531	-	0,2857	0,7143	-
3	MID-196	rs16635	0,5306	0,4694	-	0,3571	0,6429	-
4	MID-881	rs1610965	0,7755	0,2245	-	0,7619	0,2381	-
5	MID-3122	rs35451359	0,8469	0,1531	-	0,8571	0,1429	-
6	MID-548	rs140837	0,3163	0,6837	-	0,3095	0,6905	-
7	MID-659	rs1160893	0,1429	0,8571	-	0,1667	0,8333	-
8	MID-2011	rs2308203	0,5510	0,4490	-	0,7381	0,2619	-
9	MID-2929	rs33974167	0,6939	0,3061	-	0,7619	0,2381	-
10	MID-593	rs1160852	0,2143	0,7857	-	0,0952	0,9048	-
11	MID-798	rs1610884	0,4796	0,5204	-	0,5000	0,5000	-
12	MID-1193	rs2067280	0,1225	0,8776	-	0,1191	0,8810	-
13	MID-1871	rs2308067	0,1939	0,8061	-	0,3095	0,6905	-
14	MID-17	rs4183	0,4184	0,5816	-	0,3810	0,6191	-
15	MID-2538	rs3054057	0,3061	0,6939	-	0,4048	0,5952	-
16	MID-1644	rs2307840	0,7449	0,2551	-	0,8095	0,1905	-
17	MID-3854	rs60612424	0,1735	0,8265	-	0,1667	0,8333	-
18	MID-2275	rs3033053	0,1633	0,8367	-	0,1191	0,8810	-
19	MID-94	rs16384	0,1633	0,8367	-	0,0750	0,9250	-
20	MID-3072	rs34611875	0,8367	0,1633	-	0,8571	0,1429	-
21	MID-772	rs1610859	0,9184	0,0816	-	0,9762	0,0238	-
22	MID-2313	rs3045215	0,3061	0,6939	-	0,2750	0,7250	-
23	MID-397	rs25621	0,6633	0,3367	-	0,7143	0,2857	-
24	MID-1636	rs2307832	0,6939	0,3061	-	0,7857	0,2143	-
25	MID-51	rs16343	0,4286	0,5714	-	0,5476	0,4524	-
26	MID-2431	rs3031979	0,1122	0,8878	-	0,1905	0,8095	-
27	MID-2264	rs34122827*	0,4286	0,4898	0,0816	0,3333	0,5952	0,0714
28	MID-2256	rs133052	0,2041	0,7959	-	0,2857	0,7143	-
29	MID-128	rs6490	0,2551	0,7449	-	0,2857	0,7143	-
30	MID-15	rs4181	0,4082	0,5918	-	0,5238	0,4762	-
31	MID-2241	rs3030826	0,3776	0,6225	-	0,2619	0,7381	-
32	MID-419	rs140708	0,7857	0,2143	-	0,9286	0,0714	-
33	MID-943	rs1611026	0,5918	0,4082	-	0,6905	0,3095	-
34	MID-159	rs16438	0,6327	0,3674	-	0,7250	0,2750	-
35	MID-2005	rs2308161	0,4184	0,5816	-	0,5238	0,4762	-
36	MID-250	rs16687	0,7347	0,2653	-	0,6667	0,3333	-
37	MID-1802	rs2307998	0,0918	0,9082	-	0,1191	0,8810	-
38	MID-1607	rs2307803	0,2041	0,7959	-	0,2143	0,7857	-
39	MID-1734	rs2307930	0,8061	0,1939	-	0,7381	0,2619	-
40	MID-406	rs25630	0,6020	0,3980	-	0,6905	0,3095	-
41	MID-1386	rs2307582	0,1122	0,8878	-	0,1905	0,8095	-
42	MID-1726	rs2307922	0,5714	0,4286	-	0,6905	0,3095	-
43	MID-3626	rs11267926	0,5306	0,4694	-	0,6191	0,3810	-
44	MID-360	rs25584	0,7755	0,2245	0,0000	0,8333	0,1429	0,0238
45	MID-1603	rs2307799	0,2959	0,7041	-	0,3095	0,6905	-
46	MID-2719	rs34541393	0,4184	0,5816	-	0,4762	0,5238	-

Markers			Santa Catarina			Rio Grande do Sul		
	MID	rs number	short allele	long allele	variant allele	short allele	long allele	variant allele
1	MID-1470	rs2307666	0,4750	0,5250	-	0,5217	0,4783	-
2	MID-777	rs1610863	0,3750	0,6250	-	0,3044	0,6957	-
3	MID-196	rs16635	0,4000	0,6000	-	0,5652	0,4348	-
4	MID-881	rs1610965	0,8158	0,1842	-	0,7500	0,2500	-
5	MID-3122	rs35451359	1,0000	0,0000	-	1,0000	0,0000	-
6	MID-548	rs140837	0,1750	0,8250	-	0,3044	0,6957	-
7	MID-659	rs1160893	0,1250	0,8750	-	0,1087	0,8913	-
8	MID-2011	rs2308203	0,6000	0,4000	-	0,6364	0,3636	-
9	MID-2929	rs33974167	0,7250	0,2750	-	0,6087	0,3913	-
10	MID-593	rs1160852	0,0250	0,9750	-	0,0652	0,9348	-
11	MID-798	rs1610884	0,6000	0,4000	-	0,4565	0,5435	-
12	MID-1193	rs2067280	0,1500	0,8500	-	0,1136	0,8864	-
13	MID-1871	rs2308067	0,1316	0,8684	-	0,3261	0,6739	-
14	MID-17	rs4183	0,3000	0,7000	-	0,4565	0,5435	-
15	MID-2538	rs3054057	0,5750	0,4250	-	0,3913	0,6087	-
16	MID-1644	rs2307840	0,9250	0,0750	-	0,8044	0,1957	-
17	MID-3854	rs60612424	0,1250	0,8750	-	0,0652	0,9348	-
18	MID-2275	rs3033053	0,1750	0,8250	-	0,2174	0,7826	-
19	MID-94	rs16384	0,1500	0,8500	-	0,1522	0,8478	-
20	MID-3072	rs34611875	0,9750	0,0250	-	0,8636	0,1364	-
21	MID-772	rs1610859	0,9750	0,0250	-	0,9565	0,0435	-
22	MID-2313	rs3045215	0,1750	0,8250	-	0,3913	0,6087	-
23	MID-397	rs25621	0,8250	0,1750	-	0,7391	0,2609	-
24	MID-1636	rs2307832	0,8500	0,1500	-	0,6522	0,3478	-
25	MID-51	rs16343	0,5250	0,4750	-	0,6304	0,3696	-
26	MID-2431	rs3031979	0,1000	0,9000	-	0,0870	0,9130	-
27	MID-2264	rs34122827*	0,5000	0,4500	0,0500	0,5870	0,3478	0,0652
28	MID-2256	rs133052	0,2500	0,7500	-	0,2609	0,7391	-
29	MID-128	rs6490	0,4000	0,6000	-	0,3696	0,6304	-
30	MID-15	rs4181	0,4250	0,5750	-	0,4348	0,5652	-
31	MID-2241	rs3030826	0,1500	0,8500	-	0,2609	0,7391	-
32	MID-419	rs140708	0,8000	0,2000	-	0,7609	0,2391	-
33	MID-943	rs1611026	0,6750	0,3250	-	0,6522	0,3478	-
34	MID-159	rs16438	0,6579	0,3421	-	0,6522	0,3478	-
35	MID-2005	rs2308161	0,7500	0,2500	-	0,5217	0,4783	-
36	MID-250	rs16687	0,7000	0,3000	-	0,5652	0,4348	-
37	MID-1802	rs2307998	0,0250	0,9750	-	0,1087	0,8913	-
38	MID-1607	rs2307803	0,3158	0,6842	-	0,3044	0,6957	-
39	MID-1734	rs2307930	0,8500	0,1500	-	0,7609	0,2391	-
40	MID-406	rs25630	0,6500	0,3500	-	0,6957	0,3044	-
41	MID-1386	rs2307582	0,0500	0,9500	-	0,1739	0,8261	-
42	MID-1726	rs2307922	0,6000	0,4000	-	0,7826	0,2174	-
43	MID-3626	rs11267926	0,7000	0,3000	-	0,5870	0,4130	-
44	MID-360	rs25584	0,7250	0,2500	0,0250	0,6304	0,3696	0,0000
45	MID-1603	rs2307799	0,4000	0,6000	-	0,4130	0,5870	-
46	MID-2719	rs34541393	0,4250	0,5750	-	0,4783	0,5217	-

Table S3. Gene diversities for 46 AIM-Indels in twelve Brazilians populations.

Markers			Santa Isabel			Manaus			Pernambuco		
Locus	MID	rs number	Obs. Het	Exp.Het.	P-value	Obs. Het.	Exp.Het.	P-value	Obs. Het.	Exp. Het.	P-value
1	MID-1470	rs2307666	0,120	0,131	0,387	0,333	0,391	0,424	0,474	0,483	0,858
2	MID-777	rs1610863	0,380	0,430	0,250	0,463	0,495	0,755	0,489	0,491	1,000
3	MID-196	rs16635	0,150	0,189	0,070	0,390	0,438	0,490	0,504	0,488	0,725
4	MID-881	rs1610965	0,130	0,122	1,000	0,333	0,318	1,000	0,330	0,351	0,593
5	MID-3122	rs35451359	0,030	0,030	1,000	0,095	0,092	1,000	0,241	0,212	0,216
6	MID-548	rs140837	0,380	0,333	0,228	0,250	0,353	0,080	0,436	0,382	0,114
7	MID-659	rs1160893	0,290	0,316	0,521	0,214	0,265	0,231	0,286	0,277	1,000
8	MID-2011	rs2308203	0,360	0,345	0,775	0,119	0,354	0,000	0,474	0,465	0,854
9	MID-2929	rs33974167	0,090	0,086	1,000	0,244	0,347	0,072	0,346	0,389	0,262
10	MID-593	rs1160852	0,030	0,030	1,000	0,171	0,158	1,000	0,338	0,371	0,346
11	MID-798	rs1610884	0,250	0,277	0,460	0,333	0,391	0,423	0,511	0,501	0,863
12	MID-1193	rs2067280	0,300	0,322	0,532	0,214	0,230	0,528	0,331	0,297	0,244
13	MID-1871	rs2308067	0,510	0,488	0,684	0,238	0,391	0,017	0,301	0,359	0,088
14	MID-17	rs4183	0,280	0,284	1,000	0,381	0,488	0,202	0,549	0,498	0,296
15	MID-2538	rs3054057	0,220	0,242	0,398	0,357	0,327	1,000	0,474	0,437	0,424
16	MID-1644	rs2307840	0,430	0,460	0,521	0,333	0,505	0,034	0,346	0,399	0,183
17	MID-3854	rs60612424	0,060	0,058	1,000	0,143	0,134	1,000	0,278	0,329	0,107
18	MID-2275	rs3033053	0,440	0,486	0,407	0,317	0,409	0,239	0,344	0,333	0,797
19	MID-94	rs16384	0,530	0,498	0,548	0,366	0,409	0,698	0,271	0,306	0,250
20	MID-3072	rs34611875	0,120	0,113	1,000	0,214	0,230	0,528	0,368	0,413	0,214
21	MID-772	rs1610859	0,354	0,384	0,437	0,256	0,226	1,000	0,113	0,107	1,000
22	MID-2313	rs3045215	0,340	0,377	0,421	0,366	0,463	0,189	0,399	0,419	0,677
23	MID-397	rs25621	0,420	0,474	0,291	0,512	0,501	1,000	0,421	0,449	0,561
24	MID-1636	rs2307832	0,050	0,049	1,000	0,286	0,281	1,000	0,489	0,469	0,713
25	MID-51	rs16343	0,300	0,309	0,748	0,429	0,413	1,000	0,534	0,501	0,488
26	MID-2431	rs3031979	0,400	0,396	1,000	0,268	0,303	0,594	0,256	0,256	1,000
27	MID-2264*	rs34122827*	0,290	0,302	0,502	0,333	0,346	1,000	0,526	0,518	1,000
28	MID-2256	rs133052	0,464	0,369	0,011	0,429	0,505	0,365	0,316	0,306	1,000
29	MID-128	rs6490	0,280	0,309	0,339	0,357	0,423	0,457	0,391	0,396	1,000
30	MID-15	rs4181	0,330	0,316	0,757	0,381	0,501	0,131	0,519	0,498	0,726
31	MID-2241	rs3030826	0,160	0,181	0,243	0,341	0,397	0,431	0,399	0,483	0,049
32	MID-419	rs140708	0,490	0,471	0,831	0,341	0,419	0,268	0,256	0,287	0,226
33	MID-943	rs1611026	0,410	0,502	0,074	0,415	0,495	0,347	0,450	0,494	0,375
34	MID-159	rs16438	0,250	0,263	0,699	0,357	0,458	0,180	0,444	0,461	0,707
35	MID-2005	rs2308161	0,410	0,401	1,000	0,357	0,471	0,182	0,473	0,493	0,722
36	MID-250	rs16687	0,490	0,494	1,000	0,333	0,505	0,033	0,414	0,425	0,838
37	MID-1802	rs2307998	0,030	0,030	1,000	0,167	0,265	0,039	0,218	0,229	0,699
38	MID-1607	rs2307803	0,490	0,502	0,843	0,439	0,455	1,000	0,389	0,375	0,815
39	MID-1734	rs2307930	0,398	0,494	0,064	0,317	0,360	0,414	0,346	0,352	0,806
40	MID-406	rs25630	0,360	0,377	0,789	0,450	0,404	0,693	0,466	0,500	0,489
41	MID-1386	rs2307582	0,430	0,494	0,226	0,390	0,470	0,321	0,353	0,338	0,796
42	MID-1726	rs2307922	0,570	0,476	0,058	0,548	0,483	0,516	0,455	0,499	0,380
43	MID-3626	rs11267926	0,120	0,113	1,000	0,390	0,470	0,321	0,432	0,500	0,161
44	MID-360	rs25584	0,370	0,496	0,015	0,317	0,360	0,413	0,436	0,441	0,192
45	MID-1603	rs2307799	0,270	0,351	0,039	0,625	0,491	0,107	0,459	0,476	0,716
46	MID-2719	rs34541393	0,470	0,476	1,000	0,410	0,506	0,337	0,436	0,454	0,701

Markers			Alagoas			Mato Grosso do Sul			Terena		
Locus	MID	rs number	Obs.Het.	Exp.Het.	P-value	Obs.Het.	Exp.Het.	P-value	Obs.Het.	Exp.Het.	P-value
1	MID-1470	rs2307666	0,471	0,497	0,690	0,429	0,500	0,196	0,098	0,094	1,000
2	MID-777	rs1610863	0,538	0,493	0,425	0,381	0,468	0,102	0,310	0,354	0,405
3	MID-196	rs16635	0,519	0,493	0,690	0,494	0,501	1,000	0,167	0,155	1,000
4	MID-881	rs1610965	0,327	0,313	1,000	0,382	0,384	1,000	0,071	0,070	1,000
5	MID-3122	rs35451359	0,192	0,190	1,000	0,190	0,229	0,137	0,048	0,047	1,000
6	MID-548	rs140837	0,442	0,404	0,462	0,345	0,434	0,076	0,524	0,496	0,760
7	MID-659	rs1160893	0,269	0,275	0,731	0,262	0,246	1,000	0,310	0,265	0,566
8	MID-2011	rs2308203	0,500	0,466	0,526	0,500	0,455	0,469	0,071	0,070	1,000
9	MID-2929	rs33974167	0,365	0,412	0,337	0,393	0,416	0,607	0,095	0,092	1,000
10	MID-593	rs1160852	0,375	0,382	1,000	0,333	0,339	1,000	0,095	0,092	1,000
11	MID-798	rs1610884	0,466	0,477	0,837	0,476	0,500	0,667	0,452	0,354	0,090
12	MID-1193	rs2067280	0,262	0,243	0,686	0,250	0,238	1,000	0,452	0,402	0,697
13	MID-1871	rs2308067	0,346	0,346	1,000	0,333	0,389	0,256	0,429	0,450	1,000
14	MID-17	rs4183	0,577	0,499	0,118	0,560	0,500	0,282	0,214	0,194	1,000
15	MID-2538	rs3054057	0,413	0,408	1,000	0,393	0,425	0,606	0,095	0,092	1,000
16	MID-1644	rs2307840	0,505	0,430	0,104	0,354	0,352	1,000	0,683	0,470	0,005
17	MID-3854	rs60612424	0,269	0,312	0,201	0,393	0,332	0,106	0,119	0,113	1,000
18	MID-2275	rs3033053	0,365	0,357	1,000	0,390	0,331	0,174	0,405	0,506	0,224
19	MID-94	rs16384	0,365	0,420	0,238	0,345	0,371	0,556	0,429	0,367	0,406
20	MID-3072	rs34611875	0,356	0,330	0,553	0,274	0,405	0,006	0,143	0,134	1,000
21	MID-772	rs1610859	0,184	0,168	0,596	0,143	0,133	1,000	0,452	0,354	0,089
22	MID-2313	rs3045215	0,481	0,435	0,365	0,393	0,416	0,605	0,262	0,327	0,327
23	MID-397	rs25621	0,346	0,377	0,436	0,357	0,389	0,570	0,500	0,458	0,733
24	MID-1636	rs2307832	0,423	0,428	1,000	0,464	0,458	1,000	0,167	0,194	0,381
25	MID-51	rs16343	0,519	0,499	0,699	0,524	0,503	0,827	0,244	0,253	1,000
26	MID-2431	rs3031979	0,231	0,248	0,438	0,286	0,295	0,718	0,452	0,483	0,750
27	MID-2264*	rs34122827*	0,510	0,510	0,879	0,463	0,518	0,309	0,381	0,316	0,439
28	MID-2256	rs133052	0,311	0,457	0,002	0,357	0,411	0,285	0,333	0,391	0,423
29	MID-128	rs6490	0,308	0,335	0,390	0,458	0,433	0,798	0,405	0,327	0,171
30	MID-15	rs4181	0,490	0,499	1,000	0,595	0,480	0,038	0,286	0,312	0,621
31	MID-2241	rs3030826	0,510	0,463	0,392	0,476	0,447	0,626	0,300	0,258	0,568
32	MID-419	rs140708	0,240	0,255	0,694	0,190	0,211	0,319	0,429	0,465	0,738
33	MID-943	rs1611026	0,519	0,491	0,687	0,500	0,474	0,649	0,714	0,501	0,011
34	MID-159	rs16438	0,423	0,476	0,300	0,488	0,465	0,812	0,341	0,373	0,674
35	MID-2005	rs2308161	0,471	0,497	0,693	0,500	0,497	1,000	0,286	0,341	0,358
36	MID-250	rs16687	0,558	0,442	0,008	0,458	0,464	1,000	0,476	0,501	0,764
37	MID-1802	rs2307998	0,346	0,324	0,760	0,217	0,231	0,624	0,048	0,047	1,000
38	MID-1607	rs2307803	0,423	0,442	0,660	0,457	0,414	0,424	0,571	0,477	0,322
39	MID-1734	rs2307930	0,388	0,327	0,067	0,359	0,396	0,400	0,548	0,503	0,756
40	MID-406	rs25630	0,471	0,492	0,693	0,443	0,503	0,369	0,429	0,488	0,524
41	MID-1386	rs2307582	0,260	0,391	0,002	0,417	0,383	0,567	0,452	0,506	0,547
42	MID-1726	rs2307922	0,538	0,487	0,313	0,457	0,492	0,651	0,190	0,281	0,064
43	MID-3626	rs11267926	0,452	0,502	0,332	0,413	0,502	0,121	0,095	0,134	0,173
44	MID-360	rs25584	0,500	0,420	0,154	0,524	0,478	0,762	0,524	0,492	0,483
45	MID-1603	rs2307799	0,394	0,478	0,097	0,494	0,439	0,311	0,452	0,402	0,697
46	MID-2719	rs34541393	0,558	0,484	0,155	0,488	0,462	0,808	0,429	0,496	0,527

Markers			Minas Gerais			Espírito Santo			São Paulo		
Locus	MID	rs number	Obs.Het.	Exp.Het.	P-value	Obs.Het.	Exp.Het.	P-value	Obs.Het.	Exp.Het.	P-value
1	MID-1470	rs2307666	0,511	0,500	0,833	0,467	0,496	0,674	0,429	0,505	0,392
2	MID-777	rs1610863	0,420	0,440	0,808	0,522	0,463	0,261	0,449	0,458	1,000
3	MID-196	rs16635	0,523	0,502	0,831	0,429	0,501	0,206	0,531	0,503	0,778
4	MID-881	rs1610965	0,330	0,383	0,259	0,169	0,192	0,247	0,367	0,352	1,000
5	MID-3122	rs35451359	0,125	0,138	0,380	0,121	0,114	1,000	0,306	0,262	0,576
6	MID-548	rs140837	0,466	0,423	0,449	0,330	0,332	1,000	0,429	0,437	1,000
7	MID-659	rs1160893	0,295	0,284	1,000	0,239	0,212	0,351	0,245	0,247	1,000
8	MID-2011	rs2308203	0,534	0,469	0,252	0,407	0,433	0,629	0,490	0,500	1,000
9	MID-2929	rs33974167	0,455	0,436	0,806	0,348	0,417	0,133	0,449	0,429	1,000
10	MID-593	rs1160852	0,330	0,359	0,549	0,239	0,228	1,000	0,306	0,340	0,667
11	MID-798	rs1610884	0,466	0,503	0,527	0,500	0,503	1,000	0,429	0,504	0,391
12	MID-1193	rs2067280	0,295	0,253	0,199	0,304	0,303	1,000	0,245	0,217	1,000
13	MID-1871	rs2308067	0,216	0,306	0,010	0,407	0,406	1,000	0,347	0,316	0,666
14	MID-17	rs4183	0,443	0,502	0,290	0,363	0,473	0,029	0,592	0,492	0,238
15	MID-2538	rs3054057	0,489	0,440	0,338	0,433	0,458	0,647	0,449	0,429	1,000
16	MID-1644	rs2307840	0,375	0,432	0,225	0,169	0,209	0,094	0,347	0,384	0,477
17	MID-3854	rs60612424	0,352	0,383	0,574	0,163	0,151	1,000	0,265	0,290	0,615
18	MID-2275	rs3033053	0,341	0,314	0,509	0,207	0,220	0,624	0,286	0,276	1,000
19	MID-94	rs16384	0,273	0,299	0,470	0,293	0,310	0,733	0,286	0,276	1,000
20	MID-3072	rs34611875	0,477	0,419	0,210	0,283	0,259	0,685	0,327	0,276	0,326
21	MID-772	rs1610859	0,080	0,077	1,000	0,163	0,169	0,557	0,163	0,151	1,000
22	MID-2313	rs3045215	0,295	0,341	0,219	0,304	0,388	0,056	0,367	0,429	0,330
23	MID-397	rs25621	0,409	0,388	0,783	0,359	0,412	0,214	0,592	0,451	0,051
24	MID-1636	rs2307832	0,443	0,469	0,651	0,400	0,403	1,000	0,367	0,429	0,329
25	MID-51	rs16343	0,489	0,503	0,833	0,516	0,487	0,667	0,327	0,495	0,021
26	MID-2431	rs3031979	0,148	0,157	0,479	0,250	0,267	0,690	0,224	0,201	1,000
27	MID-2264*	rs34122827*	0,489	0,550	0,238	0,517	0,568	0,019	0,694	0,576	0,359
28	MID-2256	rs133052	0,364	0,353	1,000	0,352	0,306	0,182	0,367	0,328	0,661
29	MID-128	rs6490	0,386	0,419	0,607	0,440	0,476	0,511	0,347	0,384	0,477
30	MID-15	rs4181	0,534	0,495	0,518	0,363	0,497	0,012	0,449	0,488	0,766
31	MID-2241	rs3030826	0,511	0,479	0,656	0,429	0,385	0,410	0,388	0,475	0,230
32	MID-419	rs140708	0,273	0,299	0,470	0,337	0,372	0,400	0,429	0,340	0,092
33	MID-943	rs1611026	0,455	0,465	0,822	0,429	0,433	1,000	0,449	0,488	0,765
34	MID-159	rs16438	0,540	0,476	0,258	0,402	0,453	0,356	0,449	0,470	0,767
35	MID-2005	rs2308161	0,489	0,503	0,833	0,348	0,501	0,005	0,510	0,492	1,000
36	MID-250	rs16687	0,432	0,459	0,642	0,424	0,393	0,594	0,449	0,394	0,469
37	MID-1802	rs2307998	0,239	0,245	0,680	0,098	0,113	0,271	0,184	0,169	1,000
38	MID-1607	rs2307803	0,455	0,409	0,431	0,344	0,341	1,000	0,408	0,328	0,176
39	MID-1734	rs2307930	0,250	0,314	0,081	0,337	0,323	1,000	0,347	0,316	0,666
40	MID-406	rs25630	0,420	0,502	0,139	0,337	0,430	0,049	0,429	0,484	0,549
41	MID-1386	rs2307582	0,250	0,299	0,148	0,337	0,348	0,767	0,102	0,201	0,007
42	MID-1726	rs2307922	0,545	0,501	0,518	0,457	0,456	1,000	0,449	0,495	0,568
43	MID-3626	rs11267926	0,500	0,502	1,000	0,473	0,483	1,000	0,490	0,503	1,000
44	MID-360	rs25584	0,420	0,430	0,898	0,391	0,342	0,223	0,327	0,352	0,683
45	MID-1603	rs2307799	0,409	0,436	0,624	0,385	0,478	0,077	0,306	0,421	0,082
46	MID-2719	rs34541393	0,523	0,471	0,366	0,522	0,488	0,526	0,469	0,492	0,775

Markers			Paraná			Santa Catarina			Rio Grande do Sul		
Locus	MID	rs number	Obs.Het.	Exp.Het.	P-value	Obs.Het.	Exp.Het.	P-value	Obs.Het.	Exp.Het.	P-value
1	MID-1470	rs2307666	0,333	0,508	0,185	0,350	0,512	0,197	0,348	0,510	0,207
2	MID-777	rs1610863	0,381	0,418	1,000	0,350	0,481	0,338	0,348	0,433	0,363
3	MID-196	rs16635	0,619	0,470	0,184	0,500	0,492	1,000	0,435	0,502	0,674
4	MID-881	rs1610965	0,286	0,372	0,540	0,263	0,309	0,490	0,318	0,384	0,569
5	MID-3122	rs35451359	0,286	0,251	1,000		0*			0*	
6	MID-548	rs140837	0,524	0,438	0,610	0,150	0,296	0,068	0,348	0,433	0,365
7	MID-659	rs1160893	0,333	0,285	1,000	0,150	0,224	0,246	0,130	0,198	0,215
8	MID-2011	rs2308203	0,333	0,396	0,577	0,500	0,492	1,000	0,364	0,474	0,363
9	MID-2929	rs33974167	0,190	0,372	0,046	0,250	0,409	0,107	0,435	0,487	0,676
10	MID-593	rs1160852	0,095	0,177	0,144	0,050	0,050	1,000	0,130	0,125	1,000
11	MID-798	rs1610884	0,619	0,512	0,404	0,200	0,492	0,015	0,652	0,507	0,219
12	MID-1193	rs2067280	0,238	0,215	1,000	0,300	0,262	1,000	0,227	0,206	1,000
13	MID-1871	rs2308067	0,333	0,438	0,328	0,263	0,235	1,000	0,217	0,449	0,019
14	MID-17	rs4183	0,286	0,483	0,078	0,300	0,431	0,280	0,478	0,507	1,000
15	MID-2538	rs3054057	0,429	0,494	0,659	0,450	0,501	0,675	0,435	0,487	0,676
16	MID-1644	rs2307840	0,286	0,316	1,000	0,150	0,142	1,000	0,304	0,322	1,000
17	MID-3854	rs60612424	0,143	0,285	0,062	0,050	0,224	0,010	0,130	0,125	1,000
18	MID-2275	rs3033053	0,238	0,215	1,000	0,250	0,296	0,467	0,348	0,348	1,000
19	MID-94	rs16384	0,150	0,142	1,000	0,200	0,262	0,354	0,217	0,264	0,414
20	MID-3072	rs34611875	0,095	0,251	0,027	0,050	0,050	1,000	0,182	0,241	0,324
21	MID-772	rs1610859	0,048	0,048	1,000	0,050	0,050	1,000	0,087	0,085	1,000
22	MID-2313	rs3045215	0,150	0,409	0,010	0,150	0,296	0,069	0,435	0,487	0,675
23	MID-397	rs25621	0,381	0,418	1,000	0,250	0,296	0,467	0,348	0,394	0,608
24	MID-1636	rs2307832	0,333	0,345	1,000	0,200	0,262	0,353	0,435	0,464	1,000
25	MID-51	rs16343	0,524	0,508	1,000	0,450	0,512	0,670	0,391	0,476	0,411
26	MID-2431	rs3031979	0,190	0,316	0,117	0,200	0,185	1,000	0,174	0,162	1,000
27	MID-2264*	rs34122827*	0,571	0,542	1,000	0,450	0,559	0,686	0,478	0,542	0,396
28	MID-2256	rs133052	0,381	0,418	1,000	0,400	0,385	1,000	0,348	0,394	0,608
29	MID-128	rs6490	0,381	0,418	1,000	0,500	0,492	1,000	0,391	0,476	0,411
30	MID-15	rs4181	0,286	0,511	0,076	0,350	0,501	0,205	0,522	0,502	1,000
31	MID-2241	rs3030826	0,429	0,396	1,000	0,200	0,262	0,355	0,348	0,394	0,609
32	MID-419	rs140708	0,048	0,136	0,074	0,300	0,328	1,000	0,304	0,372	0,561
33	MID-943	rs1611026	0,238	0,438	0,050	0,250	0,450	0,114	0,348	0,464	0,356
34	MID-159	rs16438	0,450	0,409	1,000	0,368	0,462	0,606	0,435	0,464	1,000
35	MID-2005	rs2308161	0,571	0,511	0,676	0,300	0,385	0,544	0,348	0,510	0,208
36	MID-250	rs16687	0,286	0,455	0,137	0,200	0,431	0,026	0,348	0,502	0,205
37	MID-1802	rs2307998	0,238	0,215	1,000	0,050	0,050	1,000	0,130	0,198	0,215
38	MID-1607	rs2307803	0,333	0,345	1,000	0,526	0,444	0,607	0,348	0,433	0,366
39	MID-1734	rs2307930	0,429	0,396	1,000	0,100	0,262	0,031	0,130	0,372	0,005
40	MID-406	rs25630	0,524	0,438	0,610	0,400	0,467	0,633	0,435	0,433	1,000
41	MID-1386	rs2307582	0,286	0,316	1,000	0,100	0,097	1,000	0,261	0,294	0,519
42	MID-1726	rs2307922	0,524	0,438	0,612	0,400	0,492	0,639	0,435	0,348	0,538
43	MID-3626	rs11267926	0,571	0,483	0,642	0,400	0,431	1,000	0,391	0,496	0,399
44	MID-360	rs25584	0,143	0,292	0,028	0,300	0,422	0,108	0,217	0,476	0,020
45	MID-1603	rs2307799	0,429	0,438	1,000	0,400	0,492	0,639	0,478	0,496	1,000
46	MID-2719	rs34541393	0,381	0,511	0,380	0,450	0,501	0,675	0,522	0,510	1,000

Table S4. Genetic distances (FST) between African, European, Native American and Brazilian populations (lower diagonal) and corresponding non-differentiation P values (upper diagonal)

		PARENTAL			NORTH			NORTHEAST	
		Africa	Europe	Native American	Santa Isabel	Manaus	Belém	Pernambuco	Alagoas
PARENTAL	Africa	-	<5.E-06	<5.E-06	<5 E-06	<5.E-06	<5.E-06	<5.E-06	<5.E-06
	Europe	0.36515	-	<5.E-06	<5.E-06	<5.E-06	<5.E-06	<5.E-06	<5.E-06
	Native American	0.44273	0.29768	-	<5.E-06	<5.E-06	<5.E-06	<5.E-06	<5.E-06
NORTH	Santa Isabel	0.38589	0.22882	0.01992	-	<5.E-06	<5.E-06	<5.E-06	<5.E-06
	Manaus	0.28378	0.08085	0.12616	0.06709	-	0.00495	<5.E-06	<5.E-06
	Belém	0.24460	0.05015	0.14219	0.08961	0.00495	-	<5.E-06	<5.E-06
NORTHEAST	Pernambuco	0.19928	0.04762	0.21225	0.15373	0.03067	0.01300	-	0.07168
	Alagoas	0.21443	0.04950	0.19384	0.13823	0.02030	0.00744	0.00147	-
CENTER-WEST	Mato Grosso do Sul	0.21612	0.04507	0.21900	0.15800	0.03032	0.01120	-0.00027	0.00009
	Terena	0.39929	0.24800	0.02536	0.01654	0.08480	0.10200	0.16179	0.14511
SOUTHEAST	Minas Gerais	0.20782	0.04337	0.23940	0.17799	0.03948	0.02016	0.00006	0.00335
	Espírito Santo	0.29589	0.01014	0.25249	0.18345	0.04479	0.01989	0.01422	0.01530
	Rio de Janeiro	0.18190	0.05176	0.22025	0.16456	0.03604	0.01984	0.00081	0.00343
	São Paulo	0.23356	0.03689	0.25761	0.18498	0.04066	0.01777	0.00050	0.00297
SOUTH	Paraná	0.29837	0.01248	0.26107	0.17640	0.03232	0.01188	0.00686	0.00573
	Santa Catarina	0.34219	0.00807	0.30523	0.21873	0.06608	0.03653	0.02698	0.02713
	Rio Grande do Sul	0.30366	0.01774	0.25299	0.17293	0.03493	0.01276	0.01320	0.01169

Table S4 . Genetic distances (FST) between African, European, Native American and Brazilian populations (lower diagonal) and corresponding non-differentiation P values (upper diagonal). (Continuação)

		CENTER-WEST		SOUTHEAST				SOUTH		
		Mato Grosso do Sul	Terena	Minas Gerais	Espírito Santo	Rio de Janeiro	São Paulo	Paraná	Santa Catarina	Rio Grande do Sul
PARENTAL	Africa	<5.E-06	<5.E-06	<5.E-06	<5.E-06	<5.E-06	<5.E-06	<5.E-06	<5.E-06	<5.E-06
	Europe	<5.E-06	<5.E-06	<5.E-06	<5.E-06	<5.E-06	<5.E-06	0.00139	0.02891	<5.E-06
	Native American	<5.E-06	<5.E-06	<5.E-06	<5.E-06	<5.E-06	<5.E-06	<5.E-06	<5.E-06	<5.E-06
NORTH	Santa Isabel	<5.E-06	<5.E-06	<5.E-06	<5.E-06	<5.E-06	<5.E-06	<5.E-06	<5.E-06	<5.E-06
	Manaus	<5.E-06	<5.E-06	<5.E-06	<5.E-06	<5.E-06	<5.E-06	<5.E-06	<5.E-06	<5.E-06
	Belém	<5.E-06	<5.E-06	<5.E-06	<5.E-06	<5.E-06	<5.E-06	0.00089	<5.E-06	0.00010
NORTHEAST	Pernambuco	0.59360	<5.E-06	0.46877	<5.E-06	0.12682	0.36769	0.03158	<5.E-06	0.00020
	Alagoas	0.43016	<5.E-06	0.00980	<5.E-06	0.00059	0.05188	0.05673	<5.E-06	0.00099
CENTER-WEST	Mato Grosso do Sul	-	<5.E-06	0.73062	<5.E-06	0.27116	0.13939	0.18741	<5.E-06	0.03594
	Terena	0.16385	-	<5.E-06	<5.E-06	<5.E-06	<5.E-06	<5.E-06	<5.E-06	<5.E-06
SOUTHEAST	Minas Gerais	-0.00084	0.18570	-	<5.E-06	0.58311	0.65102	0.04386	0.00040	0.03455
	Espírito Santo	0.01281	0.19722	0.01249	-	<5.E-06	0.00673	0.72577	0.12524	0.67488
	Rio de Janeiro	0.00056	0.17494	-0.00019	0.01868	-	0.18899	0.01000	0.00010	<5.E-06
	São Paulo	0.00199	0.19439	-0.00079	0.00563	0.00136	-	0.47748	0.00356	0.13415
	Paraná	0.00335	0.19253	0.00717	-0.00145	0.00943	0.00045	-	0.49550	0.73973
SOUTH	Santa Catarina	0.02537	0.24345	0.02143	0.00545	0.03081	0.01729	0.00315	-	0.29304
	Rio Grande do Sul	0.00656	0.18759	0.00727	-0.00078	0.01542	0.00508	-0.00102	0.00707	-

Table S5. Demographic significance of the sampled populations based in 2010 IBGE data for Brazilian Regions and States.

	Total population (IBGE)	Population representior	(%)
Santa Isabel	18146	11083182	70%
Manaus (Amazonas)	3483985		
Belém (Pará)	7581051		
North	15864454		
Pernambuco	8796448	11916942	22%
Alagoas	3120494		
Northeast	53081950		
Mato Grosso do Sul	2449024	2472024	18%
Terena	23000		
Center-West	14058094		
Minas Gerais	19597330	80364410	100% *
Eapírito Santo	3514952		
Rio de Janeiro	15989929		
São Paulo	41262199		
Southeast	80364410		
Paraná	10444526	27386891	100% *
Santa Catarina	6248436		
Rio Grande do Sul	10693929		
South	27386891		
Brasil	190755799	133223449	69%

* excluding non-sampled Amerindian and Afro-descendant communities

ANEXO E - Comprovante de submissão do Artigo III à PLOS ONE**PLOS ONE: Notification of co-authorship on manuscript [Revisiting the genetic ancestry of Brazilians using autosomal AIM-Indels]**

Dear Fernanda Saloum de Neves Manta,

You are receiving this email because you have been listed as an author on a manuscript recently submitted to PLOS ONE and entitled "Revisiting the genetic ancestry of Brazilians using autosomal AIM-Indels". The corresponding author for the submission process is: Dr. Elizeu Fagundes de Carvalho

The full author list for the submission is: Fernanda Saloum de Neves Manta; Rui Pereira; Romulo Vianna; Alfredo Rodolfo Beuttenmüller de Araújo; Daniel Gitaí; Dayse Aparecida da Silva; Eldamária de Vargas Wolfgramm; Isabel da Mota Pontes; José Ivan Aguiar; Milton Ozório Moraes; Elizeu Fagundes de Carvalho, Ph.D; Leonor Gusmão

If you are not aware of this submission, or if you should not be listed as a co-author, then please contact the journal office at plosone@plos.org.

Kind regards,
PLOS ONE
<http://pone.edmgr.com/>