



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

Ana Paula Ferreira de Almeida Cabral da Silva

**Determinação da ancestralidade em populações da América do Sul**

Rio de Janeiro

2017

Ana Paula Ferreira de Almeida Cabral da Silva

**Determinação da ancestralidade em populações da América do Sul**

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

Orientadora: Prof.<sup>a</sup> Dra. Leonor Gusmão

Coorientadora: Prof.<sup>a</sup> Dra. Dayse Aparecida da Silva

Rio de Janeiro

2017

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

S586 Silva, Ana Paula Ferreira de Almeida Cabral da.  
Determinação da ancestralidade em populações da América do Sul / Ana  
Paula Ferreira de Almeida Cabral da Silva. – 2017.  
85 f.

Orientadora: Leonor Gusmão.  
Coorientadora: Dayse Aparecida da Silva.

Dissertação (Mestrado) - Universidade do Estado do Rio de Janeiro,  
Instituto de Biologia Roberto Alcântara Gomes. Pós-graduação em  
Biociências.

1. Genética de populações – Teses. 2. Polimorfismo (Genética) - Teses.  
3. América do Sul - Teses. 4. Grupos populacionais – Genética. 5. População  
– Aspectos genéticos. I. Gusmão, Leonor. II. Silva, Dayse Aparecida da. III.  
Universidade do Estado do Rio de Janeiro. Instituto de Biologia Roberto  
Alcântara Gomes. IV. Título.

CDU 616.411(8)

Bibliotecária: Ana Rachel Fonseca de Oliveira  
CRB7/6382

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

---

Assinatura

---

Data

Ana Paula Ferreira de Almeida Cabral da Silva

**Determinação da ancestralidade em populações da América do Sul**

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

Aprovada em 26 de julho de 2017.

Banca Examinadora:

---

Prof.<sup>a</sup> Dra. Leonor Gusmão (Orientadora)

Instituto de Biologia Roberto Alcântara Gomes - UERJ

---

Prof. Dr. André Luiz Mencialha

Instituto de Biologia Roberto Alcântara Gomes - UERJ

---

Prof.<sup>a</sup> Dra. Fernanda de Souza Gomes Kehdy

Fundação Oswaldo Cruz

Rio de Janeiro

2017

## **AGRADECIMENTOS**

Gostaria de agradecer a minha família pelo apoio de sempre.

A todos do Laboratório de Diagnóstico por DNA da UERJ que contribuíram direta ou indiretamente para a realização deste trabalho.

A Prof<sup>a</sup> Dra. Dayse A. Silva por ter me convidado para participar do projeto que se tornaria este trabalho.

E especialmente a minha orientadora Prof<sup>a</sup> Dra. Leonor Gusmão por toda a atenção e dedicação a este trabalho e pela valiosa amizade.

Quem sabe concentrar-se numa coisa e insistir nela como único objetivo, obtém, ao fim e ao cabo, a capacidade de fazer qualquer coisa.

*Mahatma Gandhi*

## RESUMO

SILVA, Ana Paula Ferreira de Almeida Cabral da. **Determinação da ancestralidade em populações da América do Sul**. 2017. 85f. Dissertação (Mestrado em Biociências) - Instituto de Biologia Roberto Alcântara Gomes, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2017.

A América Latina apresenta uma população híbrida formada a partir de três grupos ancestrais: Ameríndios, Europeus e Africanos. Os Ameríndios foram os primeiros a ocuparem o continente americano seguido pelos europeus (na América do Sul, chegam inicialmente espanhóis e portugueses) e por último, pelos africanos que foram trazidos para a América como mão de obra escrava. Esta região possui registros geográficos, arqueológicos e históricos únicos que são distintos de outras regiões da América, o que faz com que as populações deste subcontinente apresentem características interessantes para a compreensão de processos de migração e miscigenação. As proporções relativas das populações ancestrais que contribuem para a formação de uma população miscigenada podem ser estimadas por meio de marcadores informativos de ancestralidade (AIMs) caracterizados por uma alta diversidade entre grupos de distintos continentes. No entanto, os resultados obtidos com base no estudo deste tipo de marcadores apresentam um grau de incerteza que irá depender das diferenças de frequência alélica apresentadas por estes nas populações ancestrais, bem como no número de marcadores utilizados. Assim sendo, quanto maiores forem as diferenças alélicas entre populações de referência e o número de marcadores usados, menores serão os intervalos de confiança associados às estimativas de ancestralidade obtidas. No presente trabalho foram avaliadas as proporções de ancestralidade em diferentes populações latino-americanas (Brasil - Estado do Rio de Janeiro -, Colômbia e Paraguai) e possíveis níveis de estruturação populacional. Além disso, foram testadas diferentes estratégias de inferência de ancestralidade: (i) utilizando 46 AIM-Indels (Pereira *et al.*, 2012) e (ii) utilizando 165 AIM-SNPs (HID-Ion™ AmpliSeq – *Ion Torrent*). Os resultados obtidos permitiram a ampliação da base de dados para as populações do Estado do Rio de Janeiro e de cinco regiões da Colômbia. Pela primeira vez, foi avaliado o perfil de ancestralidade genética da população miscigenada do Paraguai. O painel de ancestralidade HID-Ion™ AmpliSeq – *Ion Torrent* demonstrou ser eficiente em detectar a variabilidade das amostras e reduzir as diferenças entre pares de irmãos. E por fim, foi observado a heterogeneidade já esperada entre estes grupos populacionais, que refletem o complexo processo de miscigenação que originou as populações atuais da América Latina.

Palavras-chave: Ancestralidade. Rio de Janeiro. Colômbia. Paraguai. AIMs. Famílias. América do Sul.

## ABSTRACT

SILVA, Ana Paula Ferreira de Almeida Cabral da. **Ancestry estimation in South American populations**. 2017. 85f. Dissertação (Mestrado em Biociências) - Instituto de Biologia Roberto Alcântara Gomes, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2017.

Latin America has an admixed population made up of three ancestral groups: Amerindians, Europeans and Africans. The Native Americans were the first to occupy the American continent followed by the Europeans (in South America initially arrived the Spaniards and the Portuguese) and finally by the Africans who were brought to America as slave labor. This region has unique geographical, archaeological and historical records that are distinct from other regions of the Americas, which makes the populations of this subcontinent present interesting characteristics for the understanding of migration processes and admixture. The relative proportions of the ancestral populations that contribute to the formation of an admixed population can be estimated through ancestry informative markers (AIMs) characterized by a high diversity among groups of different continents. However, the results obtained based on the study of this type of markers present a degree of uncertainty that will depend on their allelic frequency differences, presented by these in the ancestral populations, as well as on the number of markers used. Thus, the greater the allelic differences between reference populations and the number of markers used, the lower the confidence intervals associated with the ancestry estimates obtained. In the present study, we evaluate the proportions of ancestry in different Latin American populations (Brazil - RJ, Colombia and Paraguay) and possible levels of population sub-structure. In addition, we test different ancestry inference strategies: (i) using 46 AIM-Indels (Pereira *et al.*, 2012) and (ii) using 165 AIM-SNPs (HID-Ion™ AmpliSeq-Ion Torrent). The results obtained allowed the enlargement of the database for the populations of State of Rio de Janeiro and five regions of Colombia. For the first time, the ancestry genetic profile of the admixed population of Paraguay was evaluated. The *HID-Ion™ AmpliSeq-Ion Torrent* Ancestry Panel proved to be efficient in detecting sample variability and reducing the differences between sibling pairs. And finally, was observed the heterogeneity already expected among these population groups, which reflect the complex process of admixture that originated the present populations of Latin America.

Keywords: Ancestry. Rio de Janeiro. Colombia. Paraguay. AIMs. Families. South America.

## LISTA DE SIGLAS E ABREVIATURAS

- AIM - *Ancestry Informative Marker* - Marcadores Informativos de Ancestralidade
- AFR - África
- ATP - Adenosina trifosfato
- CEPAL - Comissão Econômica para a América Latina e o Caribe
- CNVs - *Copy Number Variants* - Variação no número de cópias
- DANE - *Departamento Administrativo Nacional de Estadística*
- DGEEC - *Dirección General de Estadística, Encuestas y Censos de La Secretaría Técnica de Planificación de La Presidencia de La República de Paraguay*
- DNA - *Deoxyribonucleic Acid* - Ácido desoxirribonucléico
- DTT - Ditionitrito
- EAAF - *Laboratorio de Genética Forense del Equipo Argentino de Antropología Forense - Forensic Genetic Laboratory*
- emPCR - *Emulsion PCR* - PCR em emulsão
- EUR - Europa
- HGDP - *Human Genome Diversity Panel* - Painel de diversidade do genoma humano
- IBGE - Instituto Brasileiro de Geografia e Estatística
- Indel - Polimorfismo de inserção/deleção
- ISPs - *Ion Sphere™*
- LDD - Laboratório de Diagnósticos por DNA
- LGM - *Last Glacial Maximum* - Último máximo glacial
- MID - *Marshfield Insertion Deletion Polymorphism* - Polimorfismos de inserção/deleção Marshfield
- MCMC - *Markov Chain Monte Carlo*
- NAM - Nativos americanos
- NGS - *Next Generation Sequencing* - Sequenciamento de nova geração
- pb - Pares de bases
- PCR- *Polymerase Chain Reaction* - Reação em cadeia da polimerase
- REDOME - Registro Nacional de Doadores de Medula Óssea
- SSC - *Saline-Sodium Citrate* – Solução Salina de Citrato de Sódio
- SNPs - *Single nucleotide polymorphisms* - Polimorfismo de um único nucleotídeo
- STRs - *Short tandem repeats* - Repetições curtas em sequência
- TE - Tris-EDTA

UERJ - Universidade do Estado do Rio de Janeiro

VNTRs - *Variable Number Tandem Repeats* -Variação no número de repetições em sequência

## SUMÁRIO

	<b>INTRODUÇÃO</b> .....	11
1	<b>OBJETIVO</b> .....	28
1.1	<b>Objetivos específicos</b> .....	28
2	<b>MATERIAL E MÉTODOS</b> .....	29
2.1	<b>Amostras</b> .....	29
2.1.1	<u>Brasil</u> .....	29
2.1.2	<u>Colômbia</u> .....	30
2.1.3	<u>Paraguai</u> .....	30
2.2	<b>Métodos de extração de DNA</b> .....	30
2.2.1	<u>Extração orgânica de DNA</u> .....	30
2.2.2	<u>Extração de DNA com Chelex®</u> .....	31
2.2.3	<u>Extração de DNA com Dodecil Sulfato de Sódio (SDS) à 0,05%</u> .....	32
2.3	<b>Quantificação de DNA</b> .....	32
2.4	<b>Inferência de ancestralidade utilizando 46 AIM-Indels</b> .....	33
2.5	<b>Análise de fragmentos por Eletroforese capilar</b> .....	36
2.6	<b>Análise de dados</b> .....	36
2.7	<b>Inferência de ancestralidade utilizando 165 AIM-SNPs</b> .....	37
3	<b>RESULTADOS</b> .....	40
3.1	<b>Rio de Janeiro</b> .....	40
3.1.1	<u>Perfil de ancestralidade do Rio de Janeiro</u> .....	40
3.1.2	<u>Comparação da eficiência de diferentes painéis de AIMS</u> .....	42
3.2	<b>Colômbia</b> .....	48
3.3	<b>Paraguai</b> .....	52
3.4	<b>Comparação dos perfis de ancestralidade</b> .....	53
4	<b>DISCUSSÃO</b> .....	54
	<b>CONCLUSÃO</b> .....	61
	<b>REFERÊNCIAS</b> .....	63
	<b>ANEXO A - Termo de Consentimento Livre e Esclarecido</b> .....	68
	<b>ANEXO B - Pôster apresentado no 62º Congresso Brasileiro de Genética</b> .....	71
	<b>ANEXO C - Trabalho que será apresentado no 27º Congresso da</b>	

Sociedade Internacional de Genética Forense.....	77
<b>ANEXO D</b> - Tabela contendo os 165 AIM-SNPs presentes no painel de ancestralidade Ion Torrent.....	84

## INTRODUÇÃO

Genética de populações é a área da ciência que busca compreender as variações genéticas individuais e populacionais relacionadas com fatores subjacentes como seleção, deriva gênica, migração e mutação. Como fonte de variação genética, as mutações podem ocorrer espontaneamente ao longo do genoma. Os fenômenos de seleção, deriva gênica e migração vão contribuir para moldar a variabilidade genética nas populações e sua distribuição mundial. Como consequência de tais processos, as populações humanas apresentam diferentes padrões de diversidade, refletindo diferentes relações de ancestralidade a nível individual e populacional (Bamshad *et al.*, 2004).

A compreensão das variações genéticas tanto a nível individual quanto a nível populacional tem sido objeto de interesse de diversos grupos de pesquisa em todo o mundo, por ser uma importante ferramenta capaz de revelar aspectos históricos e evolutivos das populações. As informações obtidas através deste tipo de análise podem ser de grande valia para várias áreas do conhecimento como a genética clínica, conservação de espécies, genética forense, entre outras.

Uma das técnicas mais utilizada em estudos genéticos é baseada em marcadores polimórficos microssatélites ou STRs (*Short Tandem Repeats*). No entanto, uma elevada taxa de mutação faz com que este tipo de marcador não apresente diferenças genéticas significativas entre populações, mas sim dentro das populações, o que dificulta o entendimento das diferenças entre grupos ancestrais (Phillips *et al.*, 2007). Marcadores de linhagem ligados ao cromossomo Y ou ao DNA mitocondrial têm sido utilizados com frequência na reconstrução da história evolutiva de populações humanas e suas correlações geográficas. No entanto, devido a natureza haplóide, a ausência de recombinação e transmissão exclusivamente materna ou paterna, esses marcadores só permitem conhecer parte da história (Phillips *et al.*, 2007), não sendo assim possível determinar a ancestralidade global da população. Neste contexto, o estudo envolvendo marcadores autossômicos informativos de ancestralidade (AIMs – *Ancestry Informative Markers*) tornou-se relevante no entendimento das heterogeneidades entre populações (Kidd *et al.*, 2011) devido a alta divergência nas frequências alélicas entre os principais grupos étnicos.

A América Latina apresenta uma população híbrida formada a partir de três grupos ancestrais: Ameríndios, Europeus e Africanos. Os Ameríndios foram os primeiros a ocuparem

o continente americano seguido pelos europeus (inicialmente espanhóis e portugueses) e por último, pelos africanos que foram trazidos para a América como mão de obra escrava (Sans, 2000; Salzano e Sans, 2014; Ruiz-Linares *et al.*, 2014).

A América do Sul apresenta registros geográficos, arqueológicos e históricos únicos que são distintos de outras regiões da América (Rothhammer e Dillehay, 2009), o que faz com que as populações deste subcontinente apresentem características interessantes para a compreensão de processos de migração e miscigenação.

As proporções relativas das populações ancestrais que contribuem para a formação de uma população miscigenada podem ser estimadas por meio de marcadores informativos de ancestralidade (AIMs) caracterizados por uma alta diversidade entre grupos de distintos continentes. No entanto, os resultados obtidos com base no estudo deste tipo de marcadores apresentam um grau de incerteza que irá depender das diferenças de frequência alélica apresentadas por estes nas populações ancestrais, bem como no número de marcadores utilizados. Assim sendo, quanto maiores forem as diferenças alélicas entre populações de referência e o número de marcadores usados, menores serão os intervalos de confiança associados às estimativas de ancestralidade obtidas.

Muitas das metodologias até agora descritas incluem um número de marcadores de ancestralidade relativamente elevado (Pereira *et al.*, 2012; Philips *et al.*, 2007). No entanto, geralmente, tratam-se de estratégias desenvolvidas para utilização em plataformas de genotipagem por determinação do tamanho de fragmento (no caso de marcadores do tipo Indel) ou de sequenciamento pelo método clássico de Sanger (no caso de SNPs), não tirando partido das metodologias mais recentes de sequenciamento de nova geração (*Next Generation Sequencing* - NGS), as quais permitem aumentar significativamente o número de marcadores a serem estudados.

Homburger e colaboradores (2015) afirmaram que caracterizar a extensão da subestrutura e contribuição diferencial dos componentes ancestrais é crucial para a compreensão da heterogeneidade genética das populações sul-americanas. Nesse sentido, no presente trabalho foram avaliadas as proporções de ancestralidade em diferentes populações latino-americanas (Brasil, Colômbia e Paraguai) e possíveis níveis de estruturação populacional. Além disso, foram testadas diferentes estratégias de inferência de ancestralidade: (1) utilizando 46 AIM-Indels pelo método convencional de análise de fragmentos (Pereira *et al.*, 2012) e (2) utilizando 165 marcadores SNPs (Kidd *et al.*, 2014; Nassir *et al.*, 2009) pelo método de Sequenciamento de Nova Geração (NGS) na Plataforma

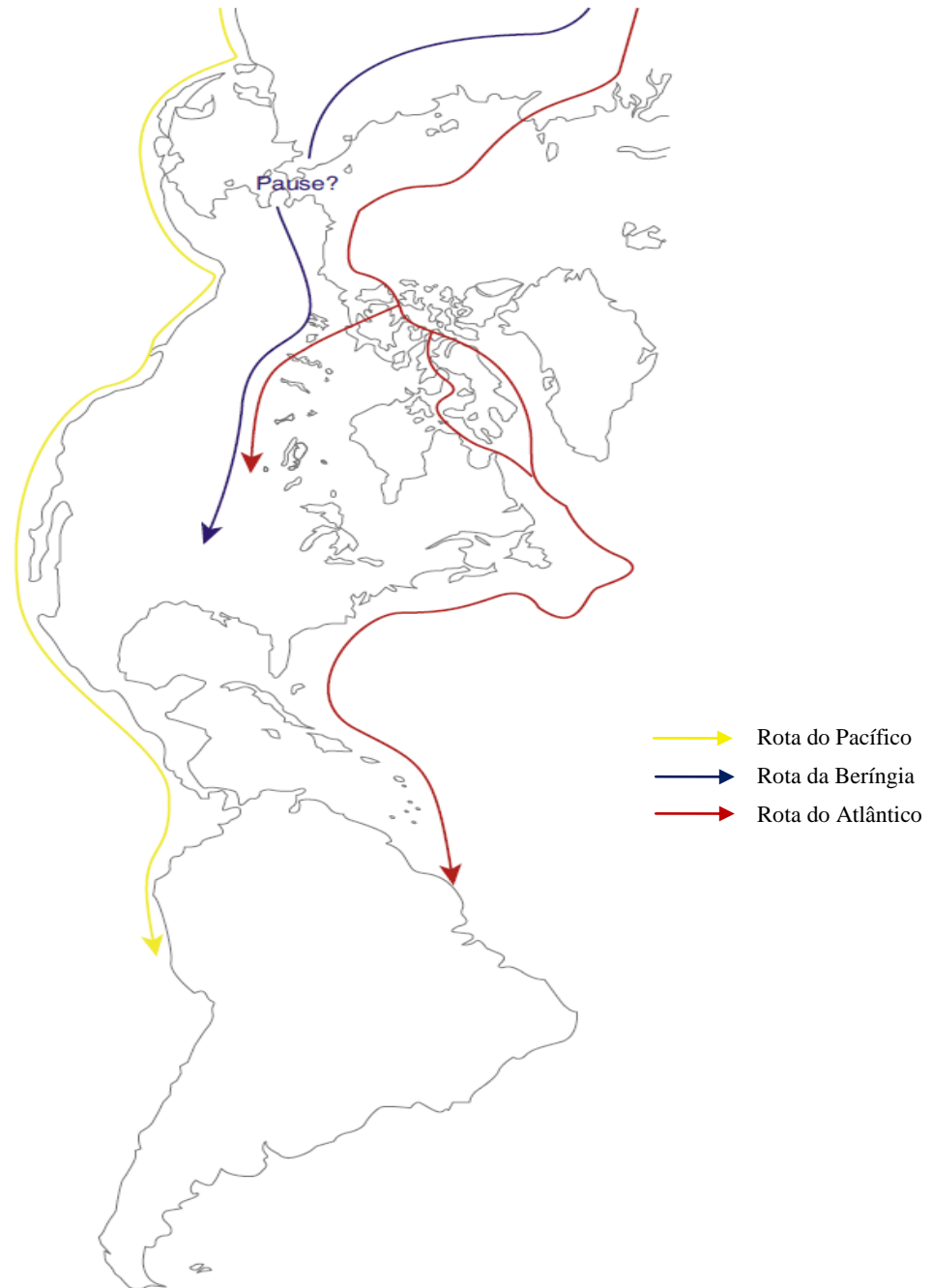
*Ion Torrent*. Pretendeu-se com este trabalho contribuir para um melhor entendimento das variações das proporções de ancestralidade nas populações miscigenadas da América Latina.

### **Origem das populações nativas da América do Sul**

A chegada dos seres humanos no Novo Mundo é amplamente debatida pela comunidade científica. Estima-se que a cerca de 15.000-18.000 anos atrás (Salzano e Sans, 2014), indivíduos de origem asiática chegaram à América através da região conhecida como estreito de Bering (Bourgeon *et al.*, 2017). Na década de 1980 Greenberg e colaboradores (1986) e Turner (1987) analisaram dados linguísticos, dentários e genéticos e sugeriram que o povoamento da América ocorreu por três migrações diferentes: (1) os primeiros ameríndios chegaram ao interior do Alasca e do Yukon e depois se espalharam para alguns territórios do Canadá e eventualmente para o resto do Novo Mundo; (2), posteriormente, a costa noroeste do Pacífico foi colonizada por falantes Na-Dene; e (3) finalmente, o Ártico foi ocupado por esquimós (Rothhammer e Dillehay, 2009). No entanto, a data de chegada dos primeiros indivíduos, as rotas migratórias e o tempo de dispersão em direção ao sul do continente, ainda são discutidas.

Levando em consideração que nem os dados arqueológicos nem os dados genéticos conseguiram resolver muitas das questões relacionadas a ocupação inicial do continente americano, O'Rourke e Raff (2010) propuseram rotas de migração alternativa do norte da Ásia para as Américas (**Figura 1**) e afirmaram que é essencial a geração de novos modelos e hipóteses onde novas e mais poderosas análises possam ser aplicadas. O cenário hipotético proposto por estes autores considera que os indivíduos que ocupavam a Beríngia estivessem explorando recursos costeiros e que este movimento poderia facilmente ter resultado na ocupação da costa norte do Alasca antes do Último Máximo Glacial (LGM) e, conseqüentemente, no movimento para o interior do continente através da drenagem do Rio MacKenzie. Além disso, condições climáticas poderiam ter favorecido um movimento para o leste da América do Norte e, neste caso, estas populações teriam sido capazes de explorar zonas rurais ecologicamente ricas. Pode-se esperar também que o movimento costeiro para a América do Sul tenha avançado mais rapidamente do que a migração de grupos localizados no interior do continente.

Figura 1 - Rotas hipotéticas para a migração inicial do norte asiático para as Américas propostas por O'Rourke e Raff (2010)



Fonte: O'Rourke e Raff, 2010 (Adaptado).

Legenda: As rotas hipotéticas do Pacífico e da Beríngia (amarela e azul respectivamente) poderiam ter ocorrido após o Último Máximo Glacial (LGM). Já a migração para o norte pela rota do Atlântico (vermelha) implicaria no movimento de uma população pré-LGM. Essas rotas migratórias não são mutuamente exclusivas (O'Rourke e Raff, 2010).

Arqueólogos observaram que o povoamento da América do Sul difere em muitos aspectos do observado na América do Norte principalmente no fato de nenhuma cultura isolada ter dominado o continente sul-americano como ocorreu na América do Norte com a cultura Clovis. Vários fatores provavelmente explicam esta diversidade inicial, como as barreiras geográficas ao movimento humano (montanhas andinas e Rio Amazonas), mudanças climáticas, mudanças nas áreas de recursos naturais, novos padrões de assentamento, novos estilos de vida, tecnologias inovadoras (domesticação de plantas), entre outros. Além disso, a ausência de migrações posteriores e um isolamento populacional, que parece ter ocorrido na América do Norte, podem ter contribuído para uma menor diversificação biológica e complexidade cultural dos nativos americanos desta região (Rothhammer e Dillehay, 2009).

Rothhammer e Dillehay, em 2009, fizeram uma revisão abordando aspectos interdisciplinares para a colonização da América do Sul e apontaram diferenças entre os lados Oeste e Leste, que possuem registros arqueológicos e histórias humanas diferentes. Estas diferenças refletem-se em padrões distintos de variação genética, morfológica e cultural. Além disso, delinearam um possível esquema de dispersão humana na América do Sul (**Figura 2**). Neste modelo, indivíduos poderiam ter entrado no planalto andino através do Istmo do Panamá. Uma vez adaptados neste local, teriam migrado para o sul do continente. Alguns grupos poderiam ter migrado para o leste e chegado em regiões como a Venezuela, Guianas e nordeste do Brasil e outros poderiam ter migrado para próximo de sistemas fluviais na Bacia Amazônica. Outra rota possível seria através da costa do Pacífico até o Chile seguindo locais de pesca favoráveis e utilizando embarcações rudimentares.

Em relação a diversidade genética, as análises da variação do mtDNA e do Cromossomo Y apontam a presença de cinco haplogrupos de mtDNA (A, B, C, D e X) (Schurr *et al.*, 1990; Brown *et al.*, 1998) e dois haplogrupos de Cromossomo Y (C e Q) (Zegura *et al.*, 2004) nas populações nativas americanas existentes. Estes haplogrupos estão presentes nas populações nativas da Sibéria, o que reforça, portanto, a origem asiática destas populações (Rothhammer e Dillehay, 2009).

Denevan, em 1976, estimou que quando os primeiros europeus chegaram à América, 57,3 milhões de indígenas viviam em todo o continente, dos quais 47 milhões habitavam os países hoje denominados latino-americanos. No entanto, estima-se que 130 anos depois, essa população havia diminuído em 90% e que a população indígena do Caribe quase foi exterminada em menos de meio século, em consequência de guerras, maus-tratos e epidemias que ocorreram com a chegada dos europeus (Comissão Econômica para a América Latina e o Caribe – CEPAL, 2014).

Atualmente, o relatório “*Los pueblos indígenas en América Latina. Avances en el último decenio y retos pendientes para la garantía de sus derechos*”, publicado em 2014 pela CEPAL, estima que haja 45 milhões de nativos americanos divididos em mais de 800 comunidades que representam 8,3% da população Latino-americana.

Figura 2 - Rotas migratórias hipotéticas para a ocupação da América do Sul proposta por Rothhammer e Dillehay (2009) e principais sítios arqueológicos que datam do Pleistoceno tardio

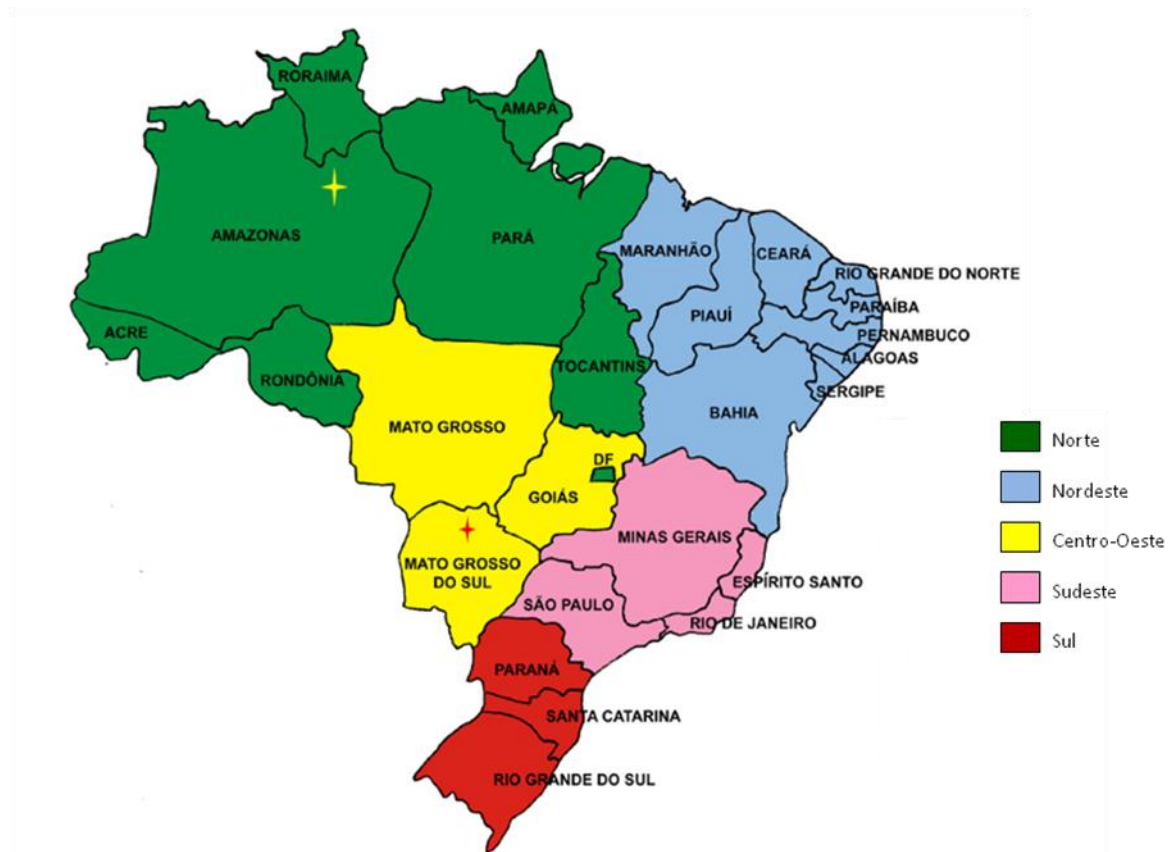


Fonte: Rothhammer e Dillehay, 2009.

## Brasil

O Brasil é o maior país da América Latina ocupando uma área de aproximadamente 8.511.960 Km<sup>2</sup> e possui uma população estimada pelo Instituto Brasileiro de Geografia e Estatística (IBGE), em 2016, em 206.081.432 habitantes. O país é dividido administrativamente em 26 estados mais a capital Brasília e, continentalmente, em 5 macrorregiões: Norte, Nordeste, Centro-oeste, Sudeste e Sul (**Figura 3**). A região sudeste é a mais populosa e se destaca também em termos econômicos. Já a região norte é a menos populosa segundo o censo 2010 (IBGE).

Figura 3 - Mapa do Brasil dividido em suas macrorregiões



Fonte: Saloum *et al.*, 2013 (Adaptado).

A população brasileira é considerada uma das mais heterogêneas do mundo sendo resultado da mistura entre três principais populações ancestrais: Ameríndia, Europeia e Africana. Os europeus (portugueses) chegaram ao Brasil em 1500, mas a ocupação efetiva do litoral brasileiro só ocorreu a partir de 1534. Já os africanos foram trazidos para o Brasil como

escravos para trabalhar, inicialmente, na economia açucareira. Estima-se que durante os séculos XVI e XIX tenham chegado ao Brasil cerca de quatro milhões de indivíduos de origem africana, que correspondem a mais de um terço do comércio de escravos daquele período (IBGE, 2007). A história recente do país também é marcada por migrações de indivíduos de outros países da Europa e da Ásia, que ocuparam principalmente o sul e sudeste do Brasil, além de migrações internas.

Devido ao processo de colonização, as migrações europeias recentes e as migrações internas, os brasileiros têm padrões de miscigenação regional variáveis. Esta característica é especialmente relevante para os estudos de associação caso-controle, pois a estratificação populacional pode produzir resultados espúrios (Pritchard e Donnelly, 2001).

Em 2014, Salzano e Sans, em um trabalho de revisão da literatura, reuniram dados de trabalhos realizados tanto com marcadores de linhagens quanto com marcadores autossômicos na população brasileira e destacaram o predomínio da proporção de ancestralidade europeia que é mais acentuada nas regiões sul e sudeste, seguida pela proporção de ancestralidade africana, em maior proporção no nordeste, e por fim o componente nativo, que é mais evidente na região norte. A região centro-oeste, de acordo com estes autores, se assemelha a região norte do país.

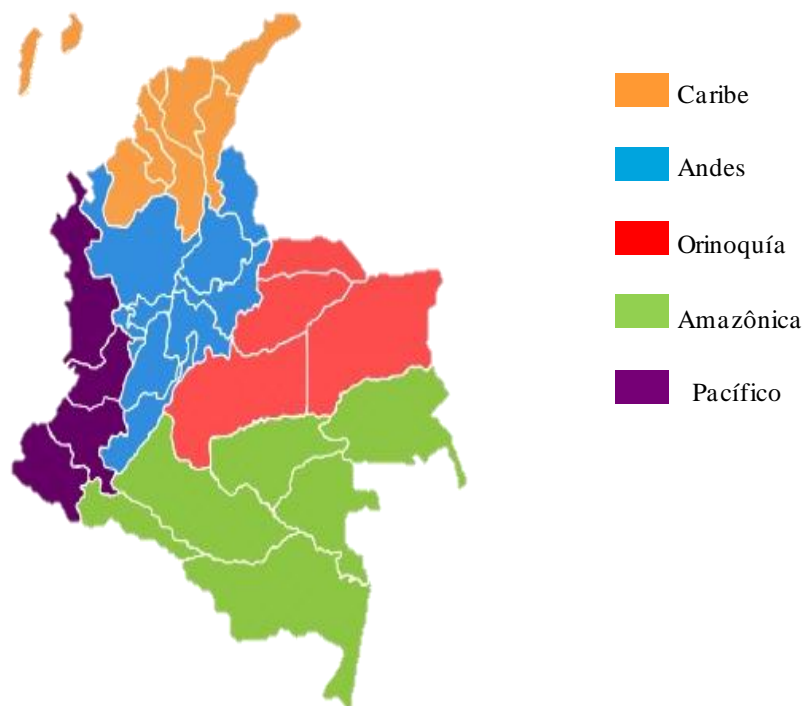
## **Colômbia**

A Colômbia está localizada no extremo norte do continente sul-americano, ocupando uma área de aproximadamente 1.141.748 Km<sup>2</sup> e com uma população estimada em 49.183.407 habitantes, segundo o Departamento Administrativo Nacional de Estadística – DANE (Censo geral 2005). O país é dividido administrativamente em 31 departamentos, mais o distrito da capital Bogotá, e faz fronteira com Panamá, Venezuela, Brasil, Peru e Equador. Continentalmente a Colômbia é dividida em cinco macrorregiões: Andes, Caribe, Orinoquia, Amazônia, regiões do Pacífico, além da região insular do Mar do Caribe (**Figura 4**). Os Andes e a região das Caraíbas são os locais mais densamente povoados e economicamente ativos do país. As outras três regiões continentais, que incluem as regiões do Pacífico, Orinoquia e Amazônia, ocupam a maior parte do território colombiano e, no entanto, são as regiões menos povoadas e desenvolvidas economicamente (DANE censo general 2005; <http://www.dane.gov.co>).

Devido a sua localização, como porta de entrada para o continente sul americano, a Colômbia tem sido sujeita a um intenso fluxo genético e cultural por migrações de nativos americanos desde os tempos pré-históricos, que resultou em uma alta diversidade de grupos étnicos e uma heterogeneidade notável entre as diferentes regiões geográficas (Arango *et al.*, 2004).

No período de colonização, a Colômbia experimentou diferentes níveis de mistura entre a população nativa, os europeus e os africanos. Estes diferentes níveis de miscigenação estão associados com diferentes padrões de migração, efeitos de deriva e/ou um maior ou menor isolamento geográfico das populações.

Figura 4 - Mapa da Colômbia dividido em macrorregiões continentais: Caribe, Andes, Orinoquia, Amazônica e Pacífico. A capital Bogotá está localizada na região dos Andes



A ancestralidade genética da Colômbia foi estudada por diversos autores que avaliaram marcadores do cromossomo Y e do mtDNA (Carvajal-Carmona *et al.*, 2000; Rojas *et al.*, 2010; Xavier *et al.*, 2015) e apontaram uma diferença entre a ancestralidade materna e paterna nas populações miscigenadas, onde há um predomínio de ancestralidade nativa na

análise do mtDNA e uma maior prevalência de ancestralidade europeia na análise do cromossomo Y. Os haplogrupos de mtDNA europeus são os segundos mais frequentes em populações não-nativas da Colômbia, exceto em algumas populações de descendentes africanos (Rodas *et al.*, 2003). As populações nativas parecem preservar o componente materno nativo, e a presença de haplótipos do cromossomo Y europeu parece depender do isolamento cultural e geográfico do grupo (Xavier *et al.*, 2015).

Os trabalhos utilizando marcadores autossômicos são ainda mais escassos do que os trabalhos com marcadores de linhagem, o que levou Ossa e colaboradores (2016) a realizarem um trabalho mais abrangente utilizando marcadores recombinantes para a análise da ancestralidade genética em populações miscigenadas da Colômbia. Neste trabalho, observou-se uma elevada diversidade e uma heterogeneidade das populações analisadas, salientando a necessidade de se aprofundar o estudo de amostras de outras regiões do país.

## **Paraguai**

O Paraguai ocupa uma área de aproximadamente 406.750 Km<sup>2</sup> divididos em 237 distritos que estão organizados em 17 departamentos mais a capital Assunção (**Figura 5**) (Dipierri, 2011). Possui cerca de 6.600.000 habitantes vivendo em sua maioria (61%) em áreas urbanas, de acordo com o censo populacional de 2012 (Dirección General de Estadística, Encuestas y Censos de La Secretaría Técnica de Planificación de La Presidencia de La República de Paraguay – DGEEC). O país faz fronteira com Bolívia, Brasil e Argentina.

A distribuição populacional pelo território paraguaio se apresenta de forma desigual (Vullo *et al.*, 2016). A região Norte (região do Chaco), onde estão localizados os departamentos de Alto Paraguay, Boquerón e Presidente Hayes, que cobrem mais de 50% da superfície do país, é povoada por somente 2,5 % da população (Dipierri *et al.*, 2011). O restante da população reside no leste do país (Região Oriental) onde está localizada a capital Assunção.

Dados históricos apontam que antes da chegada dos primeiros europeus a região situada entre os rios Paraguai e Paraná era ocupada por diversas etnias indígenas, dentre estas se destacavam as tribos guarani seminômades, enquanto que a região do Gran Chaco era percorrida por grupos nômades como os guaicurús (Colombo, 2012).

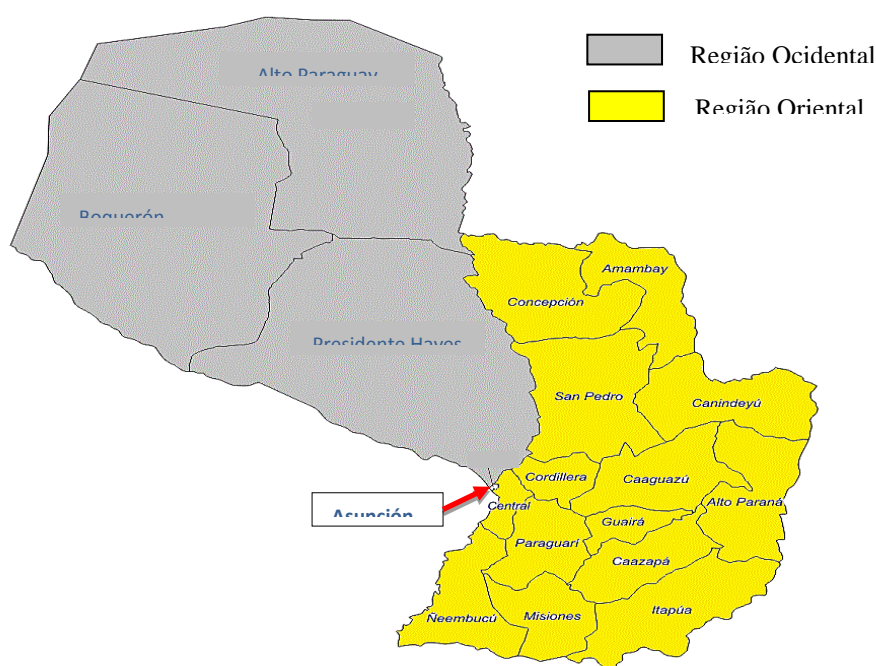
Os primeiros europeus, colonos espanhóis liderados por Aleixo Garcia e Sebastião Caboto, chegaram ao território paraguaio em 1524. Aleixo Garcia vindo de uma expedição

que se deslocou por terra a partir da costa brasileira e Sebastião Caboto, através do rio Paraná. Os primeiros núcleos coloniais foram estabelecidos por Domingos Martínez de Irala entre os anos de 1536 – 1556. Em 1537, o espanhol Juan Salazar de Espinosa construiu, numa colina próxima ao rio Paraguai, o forte de Nossa Senhora Santa Maria da Assunção, que em poucos anos se converteu em centro de irradiação dos conquistadores (Colombo, 2012). Esta região corresponde à atual capital do país. Domingos Martínez de Irala estabeleceu a política de colonização que consistiu na delimitação das fronteiras com o Brasil, na fundação de vilas e na intensa miscigenação de espanhóis com índios guarani, principal fator de formação da população do país (Paoli, 2010).

Atualmente, a população do Paraguai é composta por descendentes da população tradicional hispano-guarani que sobreviveu a Guerra do Paraguai (1864-1870) e imigrantes, principalmente europeus, que chegaram após a guerra com o objetivo de repovoar o país.

Não existem dados oficiais sobre a composição étnica da população do Paraguai já que o DGEEC não inclui o item raça ou etnia em seus levantamentos, embora incluam a percentagem de indígenas puros que habitam o país. Da mesma forma, até o presente momento, as informações sobre a ancestralidade genética desta população ainda são bastante escassas na literatura.

Figura 5 - Mapa do Paraguai com seus 17 departamentos mais a capital Assunção



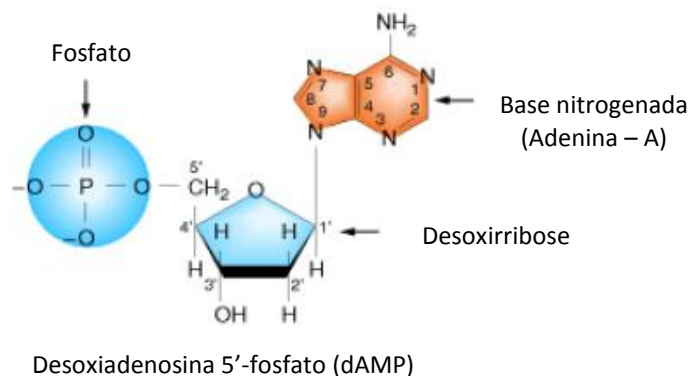
Dipierre e colaboradores, em 2011, publicaram um trabalho sobre a população do Paraguai avaliando a distribuição de sobrenomes. Benitez e colaboradores, no mesmo ano, publicaram um trabalho avaliando polimorfismos do sistema HLA em indígenas guaranis e indivíduos miscigenados. O primeiro trabalho aponta a prevalência de sobrenomes de grupos miscigenados e a relativa ausência de sobrenomes indígenas que está de acordo com a difusão dos espanhóis sobre uma área de baixa densidade povoada por grupos indígenas. O segundo trabalho demonstrou que os mestiços do Paraguai são geneticamente muito distantes dos indígenas guarani e muito mais próximos dos espanhóis. No entanto, nenhum destes trabalhos utilizou marcadores informativos de ancestralidade para determinar a contribuição africana, europeia e nativa para esta população.

### Estrutura do DNA e organização do genoma humano

O genoma é o conjunto total de informação contida nas células que é codificada pela molécula de DNA. Esta molécula é composta por unidades denominadas nucleotídeos (**Figura 6**) que são formados por um grupamento fosfato, uma pentose (Desoxirribose) e uma base nitrogenada que pode ser de quatro tipos: Adenina, Timina, Guanina e Citosina, onde Adenina só se liga a Timina e Citosina só se liga a Guanina.

Para formar uma fita de DNA os nucleotídeos ligam-se através de ligação fosfodiéster; já sua ligação com a fita complementar ocorre através de pontes de hidrogênio entre as bases nitrogenadas. A especificidade de ligação entre as bases nitrogenadas é fundamental para a replicação da molécula de DNA uma vez que a sua replicação é semiconservativa.

Figura 6 - Esquema representativo de um nucleotídeo de DNA



Fonte: <http://www.ebah.com.br/content> (Adaptado).

Nos eucariotos, a maior parte do genoma está concentrada nos núcleos. Nestes, o DNA encontra-se altamente organizado, com o auxílio de proteínas Histonas, responsáveis, basicamente, pela compactação da molécula, e de muitas outras proteínas responsáveis pela regulação dos processos de transcrição, replicação e reparo do DNA.

O DNA nuclear pode ser dividido em regiões codificantes e não codificantes. As regiões codificantes (genes) contêm as regiões reguladoras, os éxons (regiões que codificam proteínas) e os íntrons (segmentos intermediários entre os éxons). Os íntrons ocupam cerca de 5 % do genoma e as regiões que efetivamente codificam proteínas (éxons) representam cerca 1 % (Marian, 2012). Nos eucariotos, o número de genes varia de cerca de 6000, na levedura *Saccharomyces cerevisiae*, a aproximadamente 32000, no milho. No genoma humano estima-se que haja cerca de 21000 genes codificando proteínas distintas (Griffiths *et. al.*, 2013 e Lander, 2011). Nas regiões não codificantes são encontradas sequências nucleotídicas que variam muito de indivíduo para indivíduo. Estas são denominadas de regiões hipervariáveis ou *loci* hipervariáveis.

O genoma humano é composto por dois conjuntos de 23 cromossomos (22 pares de autossomos e 1 par de cromossomos sexuais XX ou XY), um conjunto herdado da mãe e outro herdado do pai, sendo o DNA um mosaico dos antepassados diretos de indivíduo. Portanto, funciona como uma entidade diplóide com fenótipos que surgem da interação complexa entre alelos de genes e/ou seus elementos reguladores funcionais não codificantes (Levy *et al.*, 2007).

Além do DNA presente no núcleo, existe também DNA extra nuclear presente nas mitocôndrias. O DNA mitocondrial (mtDNA) é uma dupla fita circular que possui 16569 pares de bases (pb) que codificam 37 genes, todos envolvidos na produção de energia e no estoque de ATP na célula. Esta molécula apresenta uma região não codificante denominada de região controle ou hipervariável. O número de mitocôndrias em cada célula pode variar de centenas a milhares fazendo com que o haja um número elevado de cópias de mtDNA. Esta característica, associada com a estrutura circular da molécula, faz com que ela seja mais resistente a degradação que o DNA nuclear. Além disso, o mtDNA apresenta uma taxa de mutação elevada quando comparada com a do DNA do núcleo. Esta alta taxa de mutação, que pode ser até dez vezes maior que no DNA nuclear, ocorre devido a ausência de proteínas histonas, exposição a radicais livres oriundos da cadeia respiratória, maior número de replicações por ciclo celular e mecanismos de reparo menos eficientes que os presentes no núcleo.

Em termos de herança genética, o mtDNA só é transmitido pela linhagem materna devido a um mecanismo complexo que ocorre na fecundação, que faz com que as mitocôndrias paternas não sejam incorporadas na formação do zigoto (Mishra e Chan, 2014). Por outro lado, a linhagem paterna pode ser estudada através dos marcadores encontrados no cromossomo Y.

### **Marcadores genéticos**

As mutações são a fonte da variabilidade genética observada nas populações humanas e, em conjunto com a recombinação, fazem com que cada indivíduo apresente características únicas. A diversidade genética das populações pode ser estudada através da análise de diferentes tipos de marcadores genéticos. Denomina-se por marcador genético, ou polimorfismo, qualquer locus que apresente 2 ou mais alelos com frequência de pelo menos 1 % na população.

A análise da variação genética humana teve início a mais de um século com a descoberta dos polimorfismos protéicos presente no sangue (Sistema ABO), por Karl Landsteiner (Landsteiner K., 1900), e até a década de 1980 a diversidade dos grupos sanguíneos e os polimorfismos de proteínas séricas eram utilizados para esta finalidade. No entanto, estes marcadores apresentam uma baixa variabilidade, sendo pouco informativos, além de degradarem rapidamente (Jobling e Gill, 2004).

Em 1984, Alec Jeffreys identificou regiões hipervariáveis no DNA, conhecidas como Minissatélites (ou VNTRs - *Variable Number Tandem Repeats*), que geravam um padrão de bandas altamente variáveis após o tratamento do DNA com enzimas de restrição e sondas específicas. Esta técnica era extremamente informativa de forma que, quando eram usadas duas sondas, a probabilidade de duas amostras coincidirem foi estimada em  $< 5 \times 10^{-19}$  - tão baixa que os únicos indivíduos a compartilharem o mesmo perfil são gêmeos monozigóticos (Jeffreys *et al.*, 1985). A técnica desenvolvida por Alec Jeffreys, que ficou conhecida como DNA *Fingerprinting* ou impressão digital de DNA, impulsionou o desenvolvimento de novas técnicas para o estudo da variabilidade genética humana.

O surgimento da técnica de Reação em Cadeia da Polimerase (PCR), também na década de 1980, tornou possível a análise de amostras com baixas concentrações de DNA,

umentando a sensibilidade de detecção das mesmas e impulsionando a descoberta de novos marcadores de interesse forense.

Além dos VNTRs, existem os marcadores STRs (*Short tandem repeats*) ou Microssatélites, que são sequências curtas que possuem de 2 a 6 pares de bases e apresentam uma alta variabilidade no número de repetições entre indivíduos. Este tipo de marcador é abundante no genoma humano e é utilizado para identificação humana em laboratórios forenses em todo o mundo. Além de serem altamente polimórficos, os STRs podem ser analisados a partir de fragmentos de PCR curtos, sendo possível a amplificação de vários marcadores em uma única reação. Devido às diferenças de tamanho dos alelos, os STRs permitem uma análise simplificada, sem a necessidade de sequenciar as amostras (Budowle e Daal, 2008).

Outra classe de marcadores, que podem ser particularmente úteis na análise de amostras degradadas, são os SNPs (*Single nucleotide polymorphisms*) – polimorfismos de um único nucleotídeo. Os SNPs consistem em substituições de bases, inserções ou deleções que ocorrem em uma posição específica no genoma (Budowle e Daal, 2008). Eles são encontrados a cada 100-300 bases ao longo do genoma e captam aproximadamente 80 % da diversidade genética total em seres humanos (Baye e Wilke, 2010).

Inicialmente, os estudos sobre a diversidade genética humana eram baseados principalmente na análise de polimorfismos STR e SNP. Recentemente as Variações no Número de Cópias (CNVs – *Copy Number Variants*) e os polimorfismos de inserção e deleção (Indels) passaram a ser também muito utilizados para esta finalidade (*International SNP Map Working Group*, 2001, Conrad *et al.*, 2010 e Weber *et al.*, 2002).

Indels são polimorfismos de comprimento caracterizados pela inserção ou deleção de um ou mais nucleotídeos (Weber *et al.*, 2002). O comprimento de um Indel varia de 1 a 10.000 pares de bases e eles encontram-se amplamente espalhados pelo genoma, sendo responsáveis por parte da variação genética encontrada em seres humanos (Mullaney *et al.*, 2010).

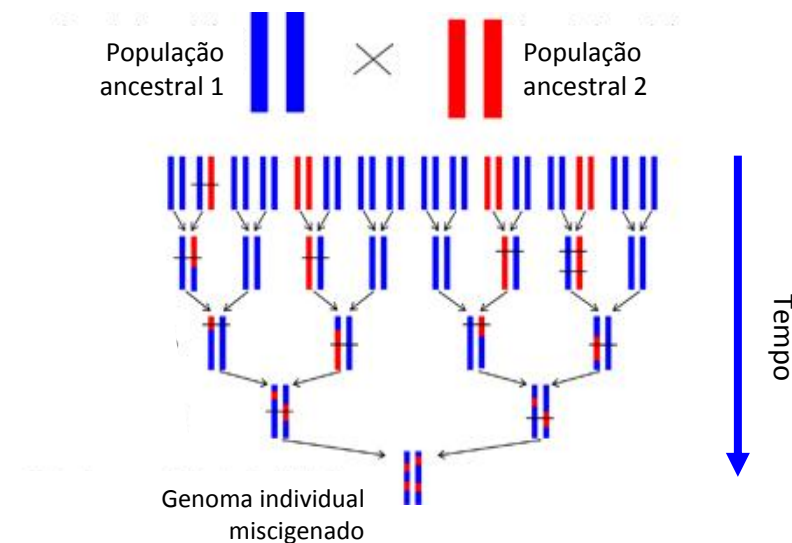
As características dos Indels são semelhantes tanto às dos SNPs quanto às dos STRs. As taxas de mutação são baixas e a amplificação gera fragmentos curtos como nos SNPs. Assim como nos STRs, os Indels são analisados por eletroforese capilar com base no tamanho, sem a necessidade de métodos de sequenciamento direto.

## Marcadores Informativos de Ancestralidade

Marcadores informativos de ancestralidade (AIMs) têm como principal característica uma baixa divergência nas frequências alélicas dentro das populações e uma alta divergência destas frequências em populações separadas geograficamente. Estes marcadores são especialmente úteis para inferir a provável origem ancestral de um indivíduo ou estimar as proporções de ancestralidade em indivíduos ou populações miscigenadas (Pereira *et al.*, 2012)

Miscigenação é uma forma comum de fluxo gênico entre populações e refere-se ao processo pelo qual duas ou mais populações genética e fenotipicamente diversas começam a se casar e formam uma nova população mista ou híbrida. Cada cromossomo de um indivíduo miscigenado se apresenta como um mosaico de segmentos cromossômicos derivados das suas populações ancestrais (**Figura 7**) (Baye e Wilke, 2010).

Figura 7 - Representação esquemática do processo de miscigenação



Fonte: Baye e Wilke, 2010. (Adaptado).

Para avaliar a ancestralidade genética em indivíduos ou populações miscigenadas é necessário utilizar populações de referência, que são as populações ancestrais que contribuíram para a formação da população miscigenada. Portanto, é preciso conhecer minimamente o contexto histórico de formação da população de interesse (Kidd *et al.*, 2014).

A informação sobre a ancestralidade pode ser usada para corrigir efeitos de estruturação populacional em estudos de caso-controle, estimar proporções de ancestralidade

em populações miscigenadas e avaliar a estrutura dessas populações. Além disso, são de grande interesse na genética forense, pois na ausência de outras informações, podem indicar o perfil de ancestralidade de suspeitos e a provável população de origem dos mesmos, podendo direcionar o curso das investigações (Phillips *et al.*, 2009; Kayser e Knijff, 2011).

## 1 OBJETIVO

O presente trabalho teve por objetivo principal aprofundar e comparar diferentes estratégias de inferência da ancestralidade nas populações do Rio de Janeiro (Brasil), Bogotá, Chocó, Tolima, Huila e San Andrés (Colômbia) e apresentar, pela primeira vez, o perfil de ancestralidade genética da população miscigenada da região Oriental do Paraguai.

### 1.1 Objetivos específicos

- a) Avaliar a ancestralidade da população do Rio de Janeiro mediante a análise de 46 AIM-Indels e comparar diferentes estratégias de amostragem em amostras de voluntários oriundos da região metropolitana e amostras de casos de paternidade coletadas em diversas regiões do Estado;
- b) Inferir ancestralidade em famílias da população do Rio de Janeiro utilizando diferentes conjuntos de marcadores: (i) 46 AIM-Indels e (ii) 165 AIM-SNPs (Painel de ancestralidade HID-Ion™ AmpliSeq – *Ion Torrent*);
- c) Avaliar a ancestralidade em 5 amostras da população da Colômbia mediante a análises de 46 AIM-Indels;
- d) Avaliar, pela primeira vez, a ancestralidade da região oriental do Paraguai mediante a análise de 46 AIM-Indels e criar uma base de dados para estes marcadores;
- e) Comparar os perfis genéticos e padrões de miscigenação das populações do Brasil (Rio de Janeiro), Colômbia e Paraguai.

## 2 MATERIAL E MÉTODOS

### 2.1 Amostras

Neste trabalho foram utilizadas 1726 amostras de indivíduos não relacionados de três países da América do Sul: Brasil (N = 1016), Colômbia (N = 188) e Paraguai (N = 522). Além destas, também foram utilizadas 121 famílias compostas por pai, mãe e dois filhos, da população do Rio de Janeiro, que correspondem a 484 amostras.

#### 2.1.1 Brasil

Para a realização do presente trabalho foram selecionadas 360 amostras de indivíduos não relacionados oriundos do Estado do Rio de Janeiro, cadastradas no Registro Nacional de Doadores de Medula Óssea (REDOME). Estas amostras foram obtidas em parceria com o Laboratório de Histocompatibilidade e Criopreservação da UERJ (CAAE: 56628116.3.0000.5259). Somado a estas, foram utilizadas 414 amostras que fazem parte de um projeto que visa o estudo de marcadores relacionados com a pigmentação e ancestralidade na população brasileira (projeto aprovado pelo comitê de ética do Hospital Universitário Pedro Ernesto/UERJ: CAAE 34301714.7.0000.5259). Além destas, também foram selecionadas 484 amostras de indivíduos, que correspondem a 121 famílias compostas por pai, mãe e dois filhos da população do Estado do Rio de Janeiro, onde 110 famílias fazem parte do banco de amostras do Laboratório de Diagnóstico por DNA da UERJ (CAAE 0067.0.228.00009) e 11 são de famílias voluntárias. Os voluntários receberam todas as informações referentes ao presente trabalho e assinaram Termo de Consentimento Livre e Esclarecido (**Anexo 1**).

### 2.1.2 Colômbia

Neste trabalho, foram utilizadas 188 amostras coletadas de cinco regiões da Colômbia: Bogotá (N = 49), Tolima (N = 24), Huila (N = 31), Chocó (N = 70) e San Andrés (N = 14). As amostras foram coletadas em locais isolados em cada região e chegaram ao LDD através de uma colaboração com o Doutor William Usaquém Martinez, pesquisador do Instituto de Genética da Universidade Nacional da Colômbia.

### 2.1.3 Paraguai

Na análise do perfil genético do Paraguai, foram utilizadas 522 amostras da região oriental, que é a mais populosa. Essas amostras foram cedidas pelo Dr. Carlos Vullo, diretor do *Laboratorio de Genética Forense del Equipo Argentino de Antropología Forense - Forensic Genetic Laboratory* – EAAF.

## 2.2 **Métodos de extração de DNA**

Para isolar o DNA contido no núcleo das células foram utilizadas diferentes técnicas de extração, de acordo com as características observadas em cada grupo de amostras (tempo entre a coleta e a extração, condições de armazenamento e quantidade de amostra). As amostras do Rio de Janeiro foram extraídas pelo método de extração orgânica (Fenol-Clorofórmio), as amostras do Paraguai e da Colômbia foram extraídas utilizando o reagente Chelex®, exceto as amostras de San Andrés que foram extraídas com SDS 0,05 %.

Todas as amostras utilizadas neste trabalho foram coletadas em papel FTA, ou seja, tratam-se de manchas de sangue.

### 2.2.1 Extração orgânica de DNA

As amostras (manchas de sangue em papel FTA) foram cortadas em pedaços de aproximadamente 0,5 mm<sup>2</sup>, inseridas em tubos de 1,5 ml com 1000 µl de *Saline-Sodium*

*Citrate* (SSC) e maceradas com auxílio de ponteiras. Em seguida, as amostras foram centrifugadas à 13000 rotações por minuto (12000 g) por 5 minutos (centrífuga Centrimicro 242 – Fanem). O sobrenadante foi desprezado com o auxílio de uma pipeta de 1000 µl, restando aproximadamente 30 µl. Foram acrescentados em cada tubo 400 µl de Tampão de Extração (Tris 0,01M; EDTA Na<sub>2</sub> 0,01 M; NaCl 0,1 M; SDS 20 % e H<sub>2</sub>O MiliQ), 25 µl de Proteinase K e 20 µl de Ditiotreitól (DTT). As amostras foram incubadas em banho-maria à 37 °C por 24 horas. Após a digestão, foram acrescentados 200 µl de Fenol/Clorofórmio/Álcool Isoamílico (25:24:1) – Invitrogen® (*Thermo Fisher Scientific*), seguida de centrifugação à 13000 rpm (12000 g) por 3 minutos. A fase aquosa superior foi recuperada em novo tubo no qual foram acrescentados 200 µl de Clorofórmio/Álcool Isoamílico (24:1), seguida de nova centrifugação nas mesmas condições citadas acima. Após esta nova centrifugação, a fase superior foi novamente recuperada. A cada amostra foram acrescentados 1000 µl de Etanol Absoluto armazenado no frio (-20 °C) e 35 µl de Acetato de Sódio à 3 molar (M). As amostras foram colocadas no *freezer* à -20°C por 24 horas. Em seguida foram centrifugadas à 12000 g, 4 °C por 15 minutos. O sobrenadante foi cuidadosamente desprezado e em cada tudo foram acrescentados 1000 µl de etanol 70 %, seguido de nova centrifugação à 12000 g por 5 minutos. O sobrenadante foi mais uma vez descartado e as amostras foram colocadas em temperatura ambiente por 24 horas para a evaporação do etanol restante. Por fim, foram acrescentados 40 µl de TE (Tris-EDTA), pH 8, em cada tudo. As amostras foram armazenadas em *freezer* à -20 °C.

### 2.2.2 Extração de DNA com Chelex®

Chelex® 100 (100-200 *Mesh*) é uma resina que atua como quelante para íons metálicos polivalentes. Durante o processo de extração as células são lisadas com água MiliQ expondo o DNA. Os grupos quelantes presentes no Chelex® modificam o equilíbrio iônico do meio fazendo com que as proteínas que estão ligadas ao DNA liberem a molécula, em seguida, o ato de ferver o meio desnatura as proteínas impedindo-as de voltarem a se ligar ao DNA. A extração com Chelex® é um bom método, pois é rápido, não requer múltiplas transferências de tubos e não utiliza solventes orgânicos tóxicos como o Fenol-Clorofórmio,

porém não consegue remover inibidores (como a Hemoglobina) que podem ser prejudiciais em etapas posteriores (Phillips *et al.*, 2012).

As amostras em papel FTA foram cortadas com cerca de 0,5 mm<sup>2</sup>, colocadas em tubo de 1,5 ml contendo 1 ml de água MilliQ e deixadas em temperatura ambiente por 30 minutos. Após este período foram centrifugadas a 12000 g por 10 minutos. O sobrenadante foi descartado e foram adicionados 200 µl de Chelex® 100 (100-200 Mesh, solução à 5 %) em agitação manual, utilizando uma ponteira com a extremidade cortada. Em seguida foram realizadas duas incubações sucessivas: a primeira à 56 °C por 20 minutos e a segunda à 100 °C por 8 minutos. Por último as amostras foram novamente centrifugadas à 12000 g por 10 minutos e armazenadas em *freezer* à -20 °C. Este procedimento resulta de adaptações ao método descrito por Lareu e colaboradores (1994).

### 2.2.3 Extração de DNA com Dodecil Sulfato de Sódio (SDS) à 0,05%

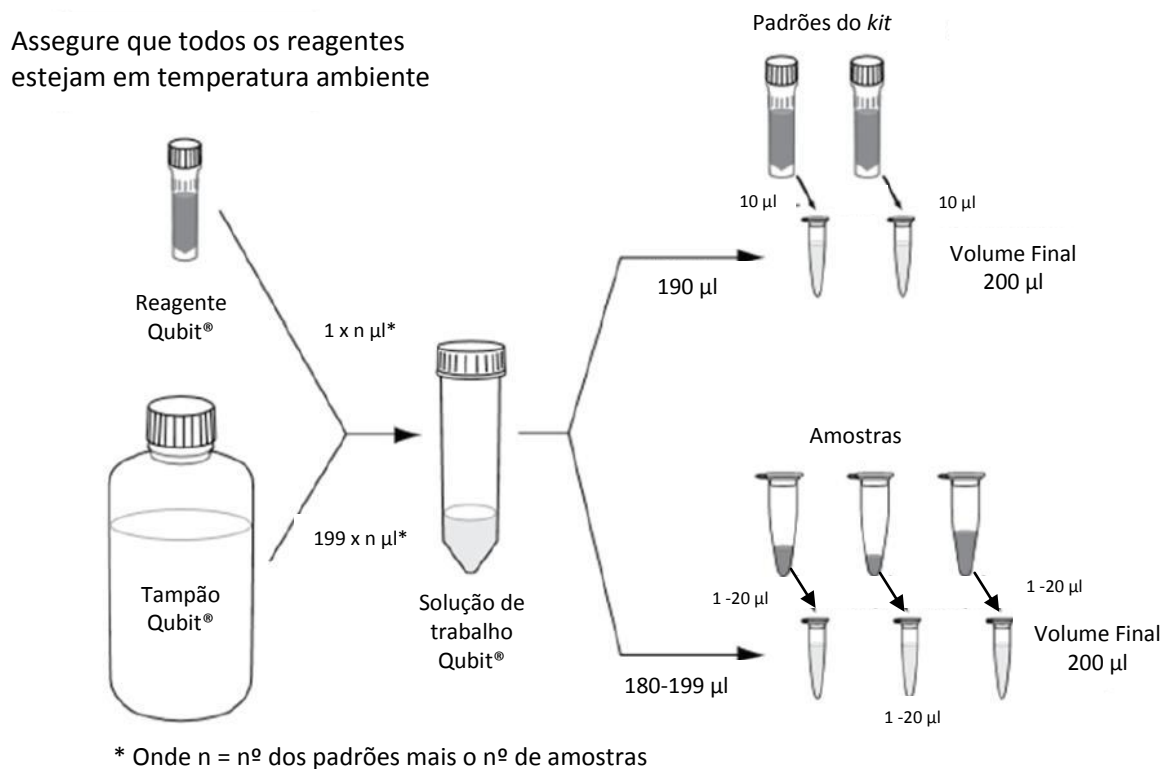
As amostras em papel FTA de aproximadamente 2 mm<sup>2</sup> foram colocadas em temperatura ambiente por 10 minutos com 200 µl de SDS à 0,05 %. Após este tempo foram agitadas em um agitador automático VX 38® (Warmnest) e em seguida o sobrenadante foi removido com o auxílio de uma pipeta de 1000 µl. Este processo foi realizado duas vezes seguido de mais duas lavagens com 200 µl de água MiliQ autoclavada. Por fim, as amostras foram incubadas a 56 °C por 25 minutos com a tampa do tubo aberta para que a água evaporasse. Logo em seguida foi feita a PCR para AIM-Indels com um volume final de 10 µl.

## 2.3 **Quantificação de DNA**

As amostras foram quantificadas utilizando o *kit* comercial Qubit® dsDNA High Sensitivity (HS) da Invitrogen (*Life Technologies*), seguindo as instruções do fabricante. A faixa de detecção deste *kit* varia de 0,2-100 ng/µl. Para calibrar o instrumento, dois padrões, *Standard # 1* (0 ng / µl) e *Standard # 2* (10 ng / µl), foram preparados e medidos no instrumento, de acordo com o protocolo.

A solução de trabalho foi preparada por combinação de 1  $\mu\text{l}$  de reagente Qubit® dsDNA HS e 199  $\mu\text{l}$  de tampão Qubit® dsDNA HS, multiplicado pelo número de amostras a serem medidas, incluindo os dois padrões. Cada solução padrão foi preparada por adição de 190  $\mu\text{l}$  de solução de trabalho e 10  $\mu\text{l}$  da solução de estoque adequada. As amostras de ensaio foram preparadas por combinação de 195  $\mu\text{l}$  de solução de trabalho e 5  $\mu\text{l}$  de DNA (**Figura 8**).

Figura 8 - Preparo da reação de quantificação com Qubit® dsDNA High Sensitivity (HS), de acordo com o fabricante



Fonte: <https://www.thermofisher.com/br/en/home/industrial/spectroscopy-elemental-isotope-analysis/molecular-spectroscopy/fluorometers/qubit/qubit-assays.html> (Adaptado)

As amostras apresentaram uma concentração média de 1,82 ng/ $\mu\text{l}$  e mesmo as que ficaram abaixo da faixa de detecção do *kit* apresentaram resultados satisfatórios nas etapas posteriores.

#### 2.4 Inferência de ancestralidade utilizando 46 AIM-Indels

O sistema utilizado neste trabalho reúne 46 AIM-Indels (**Tabela 2**) que apresentam diferenças nas frequências alélicas entre os grupos ancestrais africano, europeu, nativo

americano e asiático, sendo altamente divergentes entre pelo menos dois destes grupos. Como populações de referência para a análise dos dados foram utilizados somente os grupos ancestrais africano, europeu e ameríndio, do painel de diversidade humana HGDP-CEPH (<http://www.hagsc.org/hgdp/>).

A amplificação dos 46 marcadores informativos de ancestralidade foi realizada em uma única reação conforme protocolo descrito por Pereira e colaboradores, em 2012, e adaptado no LDD para uma reação com volume final de 5  $\mu$ l. Os primers utilizados apresentavam concentração inicial de 100  $\mu$ M e foram diluídos para uma concentração 1  $\mu$ M, adicionando 10  $\mu$ l de cada um dos 46 pares de primers (*forward* e *reverse*) e 80  $\mu$ l de H<sub>2</sub>O livre de nucleases, completando um volume final de 1 ml. As amostras foram amplificadas no termociclador Verit® (*Thermo Fisher Scientific*) utilizando o *kit* QIAGEN Multiplex PCR® (1x). As condições termocíclicas usadas encontram-se na **Tabela 1**.

Tabela 1 - Programação utilizada no termociclador Verit® para a amplificação dos marcadores 46 AIM-Indels reunidos no multiplex descrito por Pereira *et al.* (2012)

<b>Programa para PCR AIM-Indels</b>		
<b>Ciclo</b>	<b>Temperatura</b>	<b>Tempo</b>
Desnaturação inicial	95°C	15 min
28 x	94°C	30 s
	60°C	90 s
	72°C	45 s
Extensão final	72°C	60 min

Tabela 2 - Sistema multiplex – 46 AIM-Indels, localização no genoma (de acordo com dbSNP build 138) e seus alelos. O código MID indica a nomenclatura dos marcadores de acordo com a base de dados *Marshfield Diallelic Insertion/Deletion Polymorphisms*. Tabela adaptada de Pereira *et al.* (2012)

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

MID-1734	rs2307930	6	84476378	-/CCAT
MID-406	rs25630	6	14734341	-/AG
MID-1386	rs2307582	1	247768775	-/AACTATTCATTTTTCACCT
MID-1726	rs2307922	1	39896964	-/CAAGAACTATAAT/CACTATCTATTAT
MID-3626	rs11267926	15	45526069	-/AATATAATTTCTCCA
MID-360	rs25584	12	112145217	-/AA
MID-1603	rs2307799	5	70828427	-/TTGT
MID-2719	rs34541393	20	30701405	-/AACT

Legenda: As cores do código MID representam os diferentes fluorocromos (indicados pelas cores azul – 6-FAM, verde – VIC, amarelo – NED e vermelho - PET) com que os *primers* foram marcados para possibilitar a identificação de marcadores com o mesmo tamanho (pares de bases).

## 2.5 Análise de fragmentos por Eletroforese capilar

Para a realização da eletroforese capilar foi preparada uma mistura contendo 8,8 µl de Formamida Hi-Di® mais 0,2 µl de LIZ 500® por amostra. Em cada poço de uma placa de 96 poços foram inseridos 9 µl dessa mistura e 1 µl de produto de PCR. Os fragmentos amplificados foram separados por eletroforese capilar utilizando um ABI 3500 *Genetic Analyzer* (Life Technologies). Os *primers* utilizados são marcados com diferentes fluorocromos que tornam possível a amplificação de regiões com o mesmo tamanho (pares de bases). Durante a separação dos fragmentos por eletroforese capilar, os fluorocromos são identificados pelo laser do equipamento. A determinação dos alelos foi realizada posteriormente utilizando o *software GeneMapper v.4.1* (Life Technologies).

## 2.6 Análise de dados

Para estimar as proporções de ancestralidade utilizou-se o *software STRUCTURE v2.3.3* (Pritchard, 2000). Foi realizada uma análise supervisionada, utilizando informações prévias sobre a origem geográfica das amostras de referência da África, da Europa e dos Nativos americanos (Pereira *et al.*, 2012). As corridas de *STRUCTURE* compreenderam repetições de 100.000 etapas seguidas por 100.000 interações de MCMC (*Markov Chain Monte Carlo*). Foi assumida uma contribuição tri-híbrida de ameríndios, europeus e africanos

( $K = 3$ ). Foi utilizado o modelo *Admixture (Use population Information to test for migrants)*. As frequências de alelos foram correlacionadas e atualizadas utilizando apenas indivíduos com  $POPFLAG = 1$  (neste caso, as amostras de HGDP-CEPH utilizadas como referência).

As análises de distância genética foram realizadas utilizando o *software Arlequin v3.5.2.2*. Neste *software* a distância genética é dada pelos valores de *Fst* e probabilidade de não diferenciação. O nível de significância utilizado foi de 0,05.

## 2.7 Inferência de ancestralidade utilizando 165 AIM-SNPs

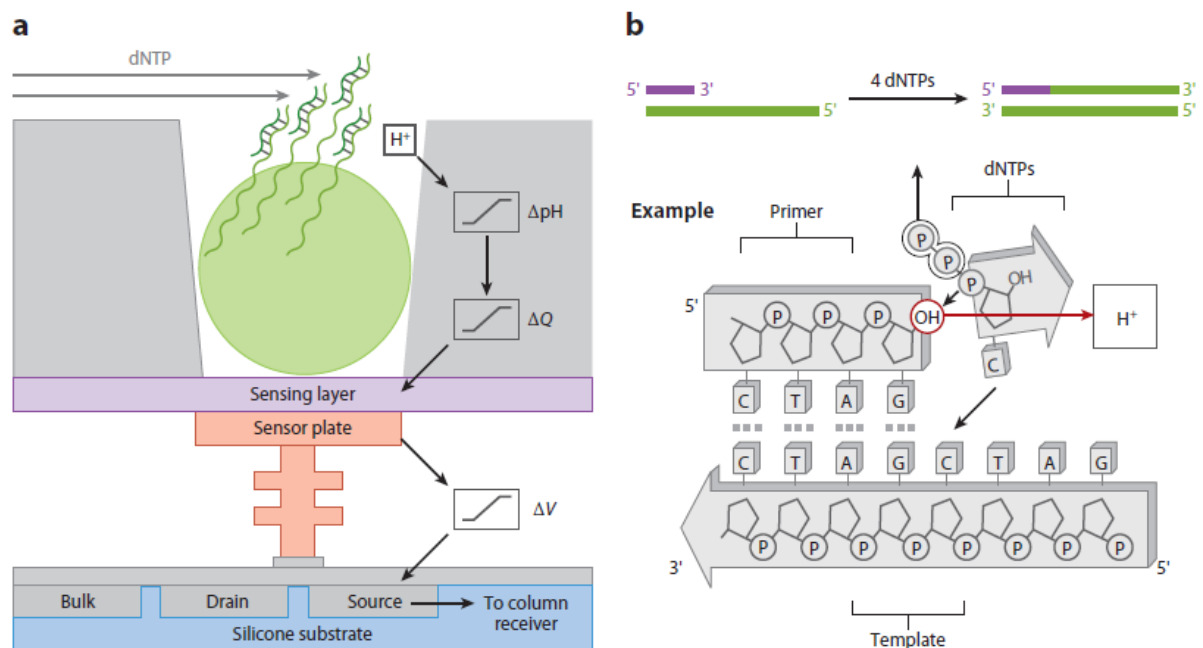
Neste trabalho foi utilizada a tecnologia de sequenciamento *Ion Torrent* que tem como base a detecção na variação de pH quando ocorre a incorporação das bases nitrogenadas no processo de sequenciamento (**Figura 9b**). Para a inferência da ancestralidade das amostras foi utilizado o Painel de Ancestralidade HID-Ion™ AmpliSeq – *Ion Torrent* (Thermo Fisher Scientific) que contém 165 marcadores autossômicos. Cinquenta e cinco destes marcadores foram selecionados com base numa publicação do Dr. Kenneth Kidd (Kidd *et al.*, 2014) e 123 marcadores foram selecionados com base numa publicação do laboratório do Dr. Michael Seldin (Kosoy *et al.*, 2009). Há 13 SNPs entre os dois conjuntos que estão sobrepostos. O protocolo foi otimizado para trabalhar com 1 ng de DNA. Portanto, as amostras que tinham uma concentração maior foram diluídas em água isenta de nucleases, a fim de obter uma concentração final de 1 ng / ul.

O fluxo de trabalho do sequenciamento *Ion Torrent* consiste, em primeiro lugar, na amplificação por PCR das regiões alvo do DNA, seguido de construção da biblioteca que inclui a remoção da sequência dos primers e ligação dos adaptadores aos fragmentos de DNA. Os fragmentos de DNA ligados ao adaptador são preparados para o sequenciamento utilizando o equipamento Ion Chef™ (Thermo Fisher Scientific, Waltham, EUA). As bibliotecas de diferentes indivíduos são adicionadas em quantidades equimolares e os fragmentos são amplificados por clonagem por PCR em emulsão (emPCR). Na emPCR, o *pool* da biblioteca é adicionado às pequenas esferas (partículas chamadas Ion Sphere™ - ISPs), aos reagentes da PCR e as moléculas de DNA polimerase (Mardis, 2013). As esferas têm seus adaptadores de sequenciamento anexados na superfície e são complementares aos adaptadores *Ion Torrent* ligados aos produtos de PCR. No final da emPCR, cada esfera é

coberta por milhares de moléculas de DNA amplificadas por clonagem de fragmentos (**Figura 9a**).

Durante o processo de sequenciamento os nucleotídeos são adicionados, um de cada vez. Se o nucleotídeo for complementar a base do DNA molde, a DNA polimerase catalisa a reação e um íon de hidrogênio é liberado, diminuindo o pH da solução contida no micropoço. Um chip semiconductor detecta a variação de pH de cada poço e converte em uma mudança de tensão que é então traduzida para o nucleotídeo adicionado. Se houver duas bases idênticas na cadeia de DNA molde, a tensão registrada será o dobro e o *software* irá identificar duas bases idênticas. Se o nucleotídeo não for complementar ao DNA molde, ele não será adicionado, não ocorrendo nenhuma variação de tensão. O processo continua até que todas as bases sejam sequenciadas (Santangelo, 2016).

Figura 9 - (a) Esquema de um único poço em um chip de sequenciamento *Ion Torrent*. (b) Detecção da variação do pH quando ocorre a incorporação do nucleotídeo



Fonte: Mardis, 2013.

No trabalho do grupo do Dr. Michael Seldin (*Kosoy et al.*, 2009), foi apresentado um painel com 128 SNPs capazes de distinguir entre quatro grupos continentais (Europeus, Ameríndios, da África Ocidental e do Leste Asiático). Estes marcadores mostraram grandes

diferenças entre as populações continentais e diferenças relativamente pequenas em indivíduos dentro de cada um dos grupos continentais.

Já no trabalho de Kidd e colaboradores (2011), foi apresentado um painel com 55 SNPs altamente informativos, que são capazes de distinguir de sete a oito populações biogeográficas utilizando 73 populações de referência. É ressaltado neste trabalho, que a ancestralidade pode não ser estimada com precisão em populações miscigenadas, a menos que as populações ancestrais sejam representadas entre as populações de referência.

Para a análise dos dados juntamente com o conjunto de marcadores 46 AIM-Indels, as sequências geradas com este painel foram convertidas em genótipos, onde A=1, T=2, C=3 e G=4. Em seguida foram realizadas as análises utilizando o *software STRUCTURE v2.3.3* (Pritchard, 2000) com os mesmos parâmetros de análise utilizados para os 46 AIM-Indels. As populações de referência utilizadas neste caso foram as mesmas dos trabalhos que descreveram estes marcadores (Kosoy *et al.*, 2009; Kidd *et al.*, 2011).

### 3 RESULTADOS

Primeiramente serão apresentados os resultados referentes ao estudo dos indivíduos não relacionados da população do Rio de Janeiro, bem como as análises feitas utilizando diferentes painéis para inferência de ancestralidade em famílias desta população. Em seguida serão apresentados os resultados obtidos para as populações da Colômbia e Paraguai.

#### 3.1 Rio de Janeiro

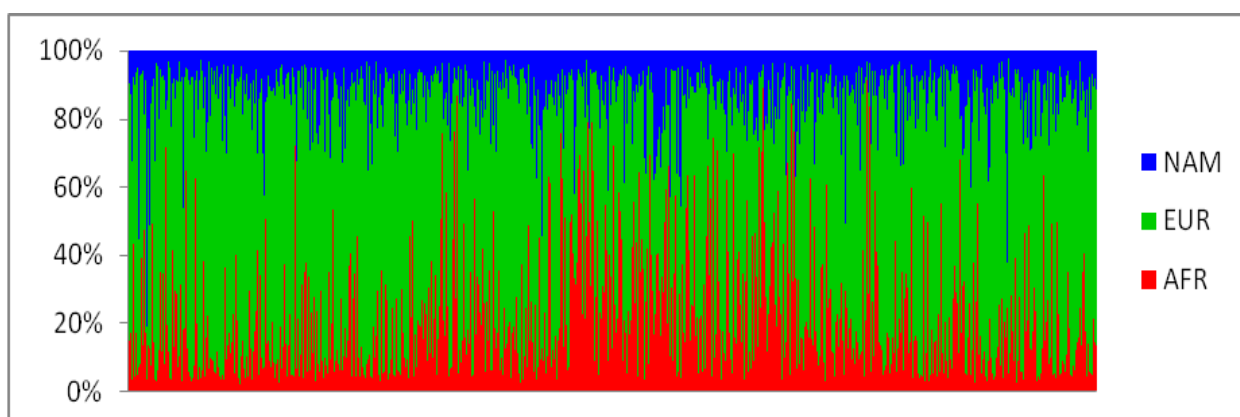
Para as amostras da população do Rio de Janeiro, foram feitos diferentes tipos de análises, nomeadamente: (i) foram estimadas as proporções de ancestralidade utilizando somente indivíduos não aparentados; (ii) comparou-se os resultados obtidos para diferentes estratégias amostrais e (iii) para diferentes painéis de AIMS, previamente descritos.

##### 3.1.1 Perfil de ancestralidade do Rio de Janeiro

O perfil de ancestralidade geral da população do Rio de Janeiro foi estimado numa amostra incluindo todos os indivíduos não aparentados, analisados ao longo deste trabalho.

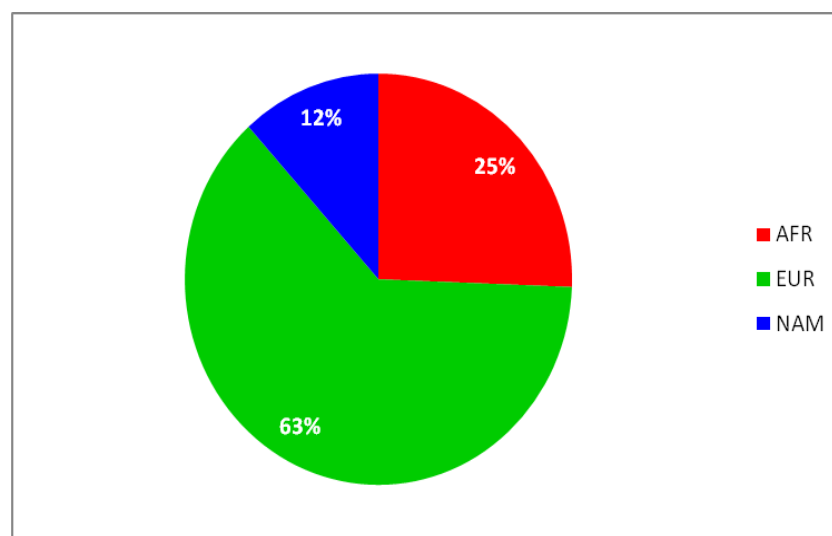
Os resultados obtidos na análise de 957 indivíduos demonstraram um predomínio da ancestralidade europeia, seguida da ancestralidade africana e em menor proporção a ancestralidade ameríndia (**Figura 10 e 11**).

Figura 10 - Perfil de ancestralidade individual em 957 indivíduos não relacionados da população do Rio de Janeiro utilizando 46 AIM-Indels



Legenda: Ancestralidade ameríndia (NAM), europeia (EUR) e africana (AFR).

Figura 11 - Média das proporções de ancestralidade para a população do Rio de Janeiro utilizando 46 AIM-Indels

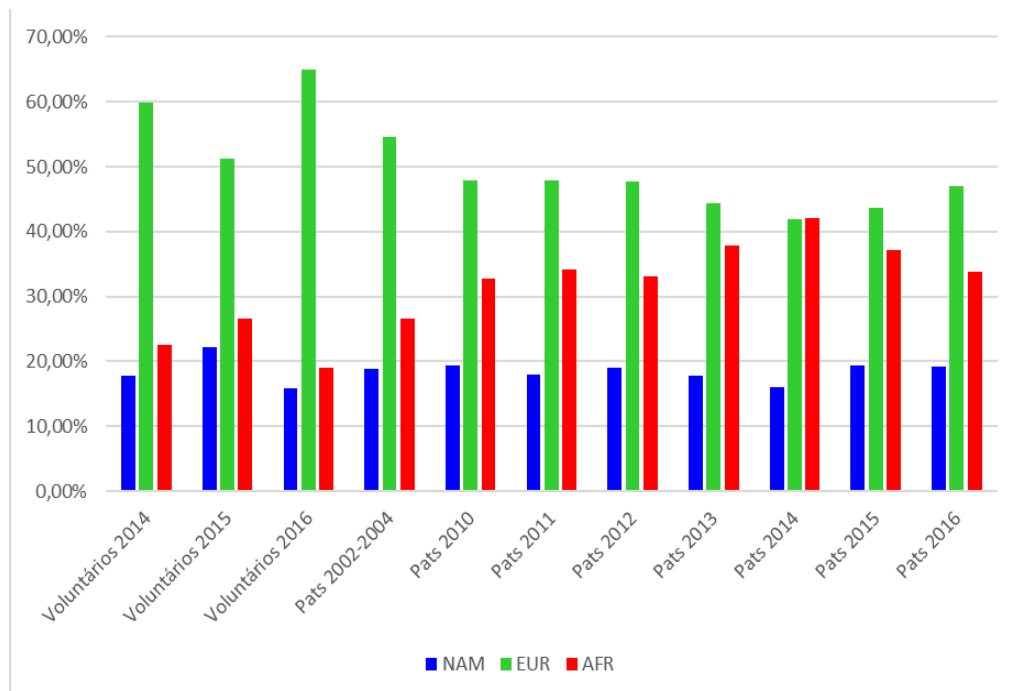


Legenda: Ancestralidade ameríndia (NAM), europeia (EUR) e africana (AFR).

Para determinar a influência do método de amostragem nas estimativas de ancestralidade da população do Rio de Janeiro, as amostras utilizadas neste trabalho foram divididas em: i) amostras que estavam armazenadas no LDD e que foram coletadas para a realização de testes de paternidade e ii) amostras que foram coletadas de voluntários que frequentam a UERJ. Além disso, as amostras foram ainda agrupadas por ano de coleta. Os resultados expostos na **Figura 12** demonstram a prevalência da ancestralidade europeia, seguida da africana e em menor proporção a ancestralidade ameríndia, nas amostras dos voluntários nos períodos de 2014 e 2015. O mesmo perfil foi observado para as amostras dos casos de paternidade nos

períodos de 2002-2004 e 2010-2012. No entanto, no ano de 2013 o percentual de ancestralidade africana já estava próximo ao percentual europeu e em 2014 estes percentuais igualaram-se. Os percentuais de ancestralidade ameríndia não apresentaram variação importante entre grupos.

Figura 12 - Médias das proporções de ancestralidade em indivíduos não relacionados da população do Rio de Janeiro, obtidas através da análise de 46 AIM-Indels (Pereira *et al.*, 2012).



Legenda: Ancestralidade ameríndia (NAM), europeia (EUR) e africana (AFR). Pats: Paternidades.

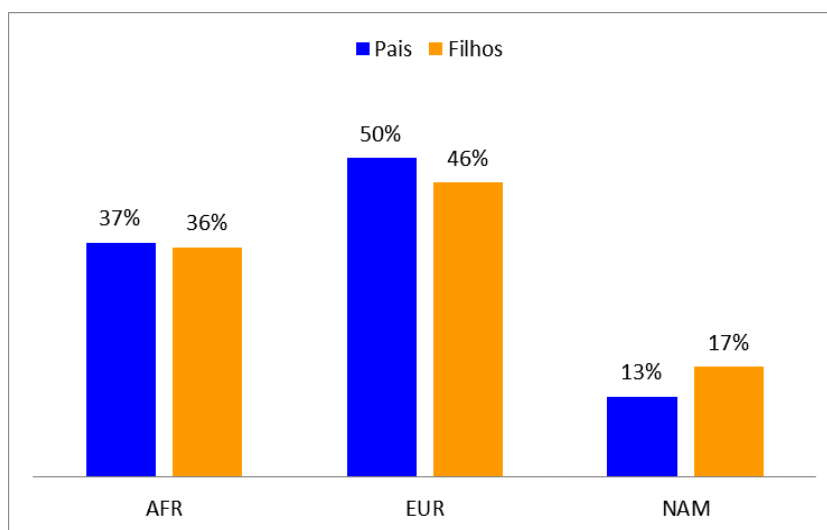
### 3.1.2 Comparação da eficiência de diferentes painéis de AIMs

Neste trabalho, foi comparada a eficiência de diferentes painéis de AIMs: (i) um multiplex incluindo 46 AIM-Indels descrito por Pereira e colaboradores, em 2012, e (ii) um segundo multiplex contendo 165 AIM-SNPs. Destes, 123 foram descritos no trabalho de Kosoy e colaboradores, em 2009, e 55 marcadores foram descritos por Kidd e colaboradores, em 2014 (13 SNPs encontram-se sobrepostos nos painéis). Pretendemos com esta análise avaliar a eficiência dos painéis em estimar as proporções de ancestralidade em indivíduos miscigenados. Para tal, foram analisadas famílias constituídas por pais e dois filhos. Nesta

análise assumiu-se que (i) a valor de ancestralidade observado nos filhos deve aproximar-se da média encontrada nos pais e, sendo assim, (ii) dois irmãos deverão apresentar valores próximos de ancestralidade.

Uma primeira análise foi feita em 121 famílias utilizando somente os 46 AIM-Indels. Nesta análise, foi observado que as diferenças entre a média da ancestralidade dos pais e a ancestralidade dos filhos variaram de 0 – 27 %. As diferenças das proporções de ancestralidade média dos pais - para os componentes africano, europeu e nativo - e as encontradas nos filhos estão expostas na **Figura 13**. Foi observado também que a ancestralidade africana varia pouco entre pais e filhos, já a ancestralidade europeia apresentou uma redução nos filhos, enquanto que a ancestralidade ameríndia apresentou um aumento. Esta tendência já havia sido observada nos resultados parciais apresentados do 62º Congresso de Genética, quando analisadas 100 famílias (**Anexo 2**).

Figura 13 - Proporções de ancestralidade média dos pais (azul) *versus* ancestralidade observada nos filhos (laranja)



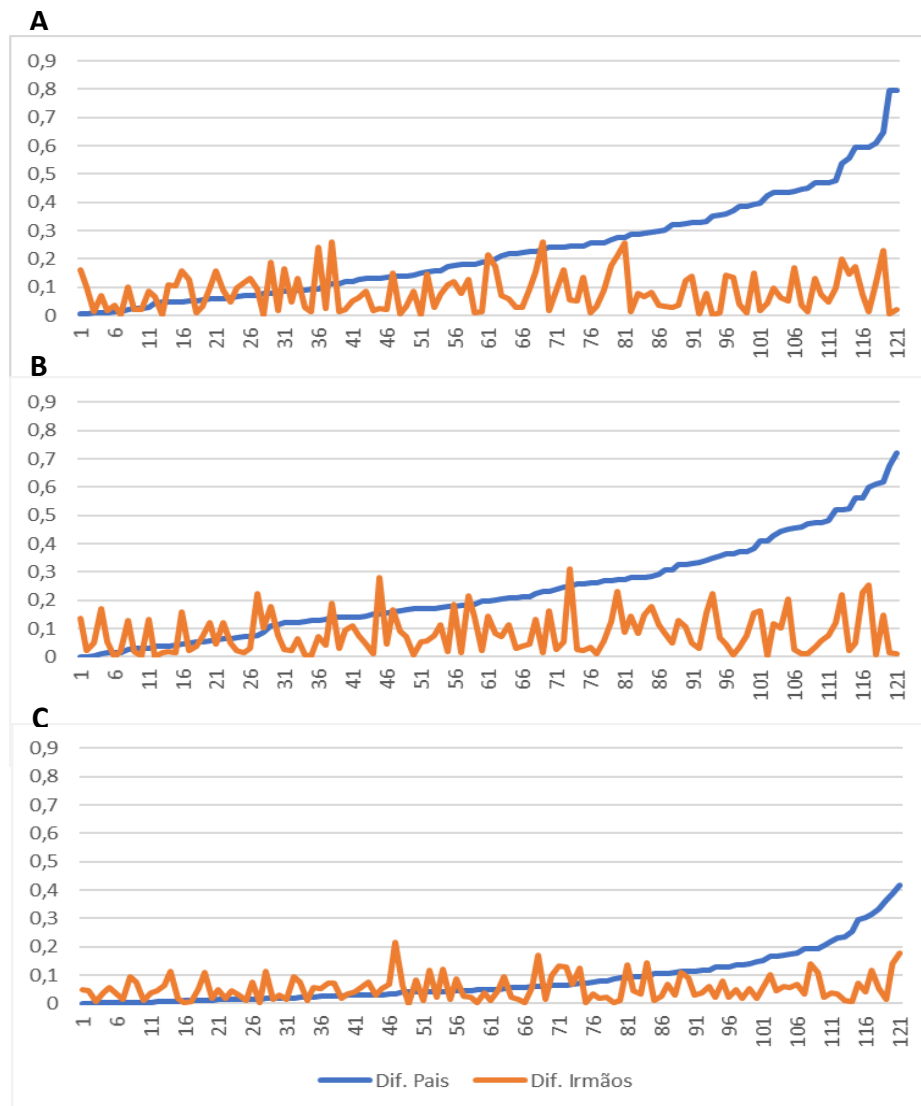
Legenda: Ancestralidade africana (AFR), europeia (EUR) e ameríndia (NAM).

Na comparação dos valores de ancestralidade entre os 121 pares de irmão, foram observadas diferenças que variaram de 0 – 28 %.

Outra análise foi feita para testar se as diferenças observadas entre irmãos estariam relacionadas com uma contribuição diferencial por parte dos progenitores, para o conjunto de marcadores analisados. Neste caso, seria de esperar que quanto mais diferentes fossem os pais, em termos de proporções de ancestralidade, maiores seriam as diferenças encontradas entre irmãos.

Como exposto na **Figura 14**, não foi observada correlação entre as diferenças de ancestralidade dos pais e as diferenças observadas entre os irmãos, para este grupo de marcadores. Ou seja, há uma variabilidade entre os irmãos que parece não ter relação com o aumento nas diferenças de ancestralidade dos pais.

Figura 14 - Comparação das diferenças observadas entre pais versus diferenças observadas entre os irmãos, para os perfis de ancestralidade africano, europeu e nativo americano, utilizando 46 AIM-Indels

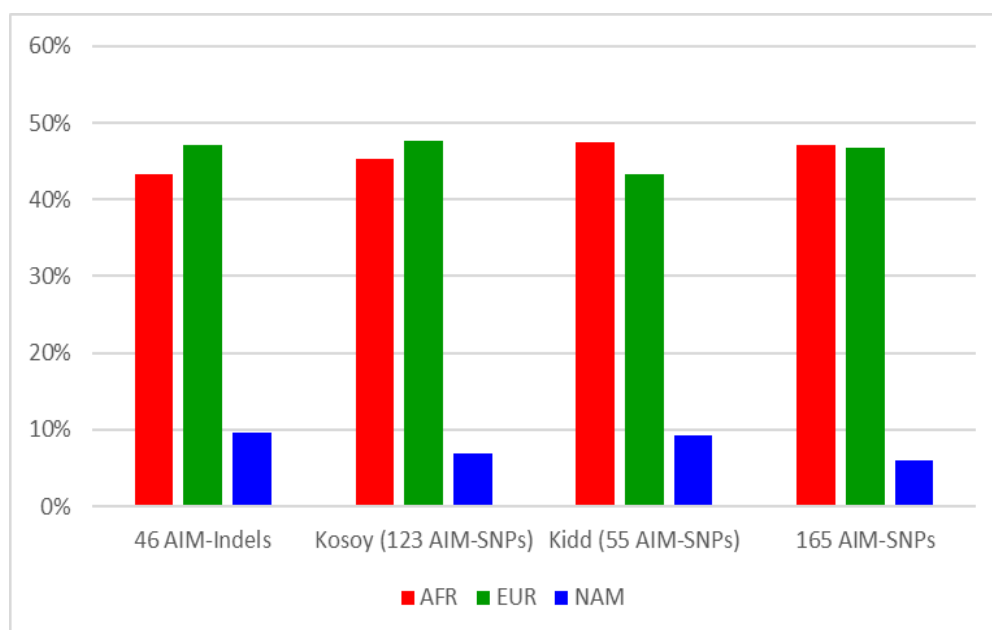


Legenda: A) Perfil de ancestralidade africano, B) Perfil de ancestralidade europeu e C) Perfil de ancestralidade Nativo americano. Dif. Pais: diferenças entre os pais. Dif. Irmãos: diferenças entre os irmãos. Eixo X: famílias e Eixo Y: percentual de ancestralidade.

Em uma segunda análise, foram selecionadas 20 famílias para serem analisadas utilizando 165 AIM-SNPs presentes no painel de ancestralidade HID-Ion™ AmpliSeq da Ion

*Torrent* e os resultados foram comparados com os obtidos para os 46 AIM-Indels. Primeiramente, foram comparadas as médias das proporções de ancestralidade somente dos indivíduos não relacionados (pais e mães) dessas 20 famílias. Podemos observar (**Figura 15**) que os valores encontrados foram bastante próximos, quando utilizados os vários grupos de marcadores. No entanto, para o componente africano, a média obtida com os 46 AIM-Indels está mais próxima da obtida com os 123 AIM-SNPs; já com os 55 AIM-SNPs, a média é igual à obtida para o conjunto total de 165 marcadores SNP. Para o componente europeu a média dos 55 AIM-SNPs ficou abaixo da média dos outros grupos de marcadores, com uma diferença de 5 % em comparação com a média dos 123 AIM-SNPs. Para o componente nativo americano a média dos 46 AIM-Indels ficou próxima da média dos 55 AIM-SNPs e o conjunto total com 165 AIM-SNPs apresentou o menor valor médio.

Figura 15 - Média das proporções de ancestralidade em indivíduos não relacionados da população do Rio de Janeiro (pais e mães). Comparação entre 46 AIM-Indels e ID-Ion™ AmpliSeq – *Ion Torrent*

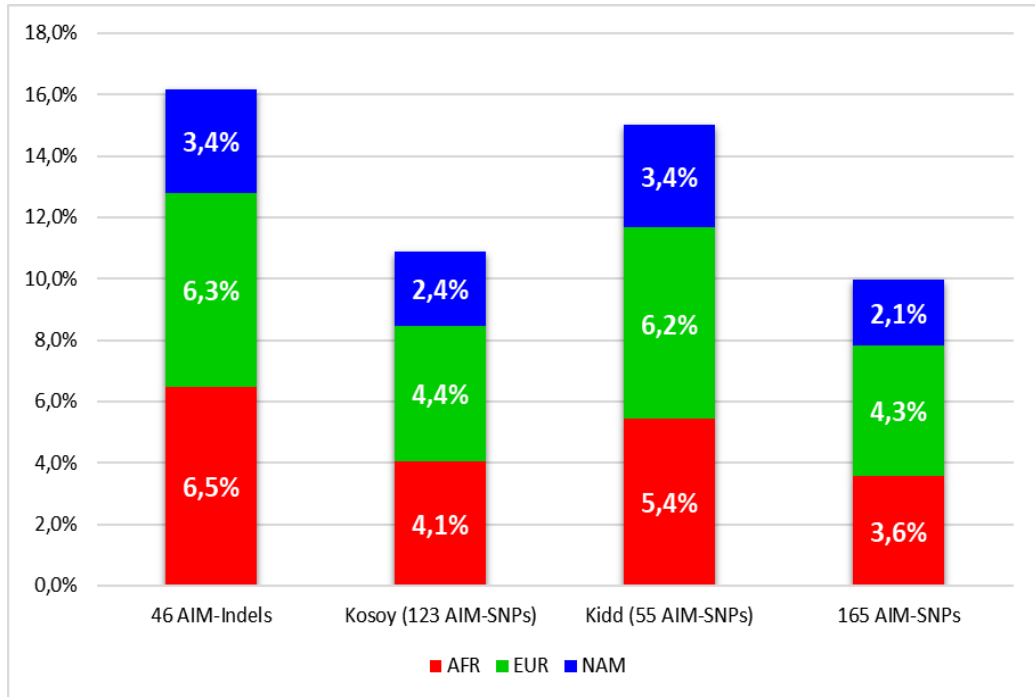


Legenda: Comparação entre 46 AIM-Indels (Pereira *et al.*, 2012), 123 AIM-SNPs (Kosoy *et al.*, 2009), 55 AIM-SNPs (Kidd *et al.*, 2014) e 165 AIM-SNPs (HID-Ion™ AmpliSeq - Kosoy *et al.*, 2009 + Kidd *et al.*, 2014).

Para avaliar qual dos painéis estudados apresenta uma maior correlação entre os valores de ancestralidade em pais e filhos, foi comparada a diferença entre a média de ancestralidade encontrada nos pais e a ancestralidade observada de cada filho (**Figura16**). Foi observado que o painel com 165 AIM-SNPs apresentou as menores diferenças, seguido pelo

painel com 123 AIM-SNPs. Os painéis com 46 AIM-Indels e com 55 AIM-SNPs apresentaram perfil semelhante, com diferenças maiores que os outros painéis.

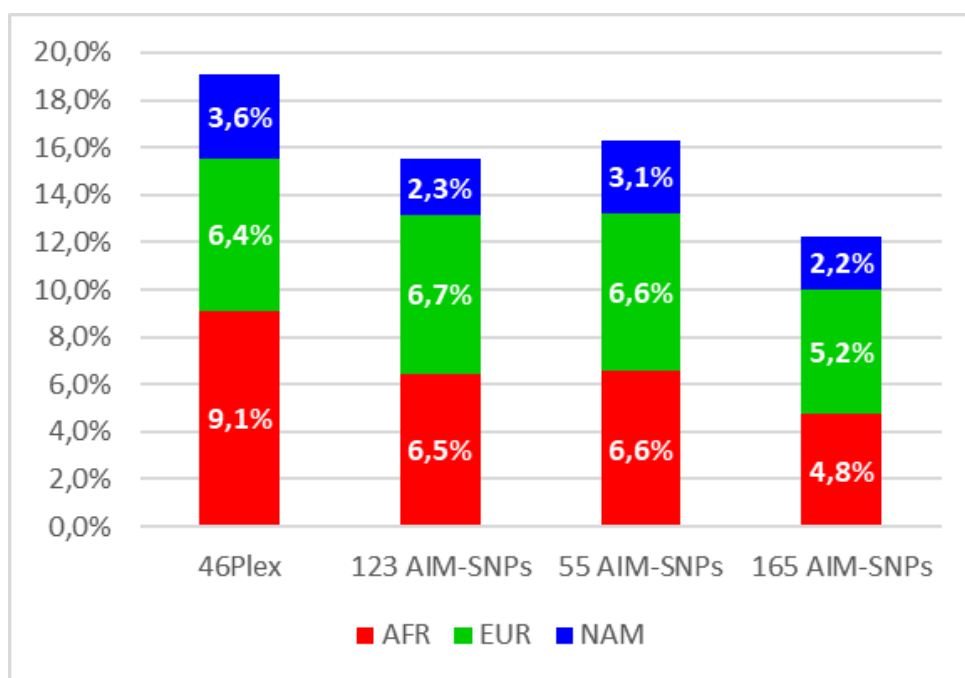
Figura 16 - Média de ancestralidade dos pais versus ancestralidade dos filhos com diferentes painéis de ancestralidade



Legenda: 46 AIM-Indels (Pereira *et al.*, 2012), 123 AIM-SNPs (Kosoy *et al.*, 2009), 55 AIM-SNPs (Kidd *et al.*, 2014) e 165 AIM-SNPs (HID-Ion™ AmpliSeq - Kosoy *et al.*, 2009 + Kidd *et al.*, 2014).

Também foram comparadas as diferenças médias entre os irmãos, para cada conjunto de marcadores (**Figura 17**). Foi observado que o painel com 165 AIM-SNPs apresentou as menores diferenças entre os irmãos. Já o painel com 46 AIM-Indels apresentou as maiores diferenças, principalmente para o componente africano.

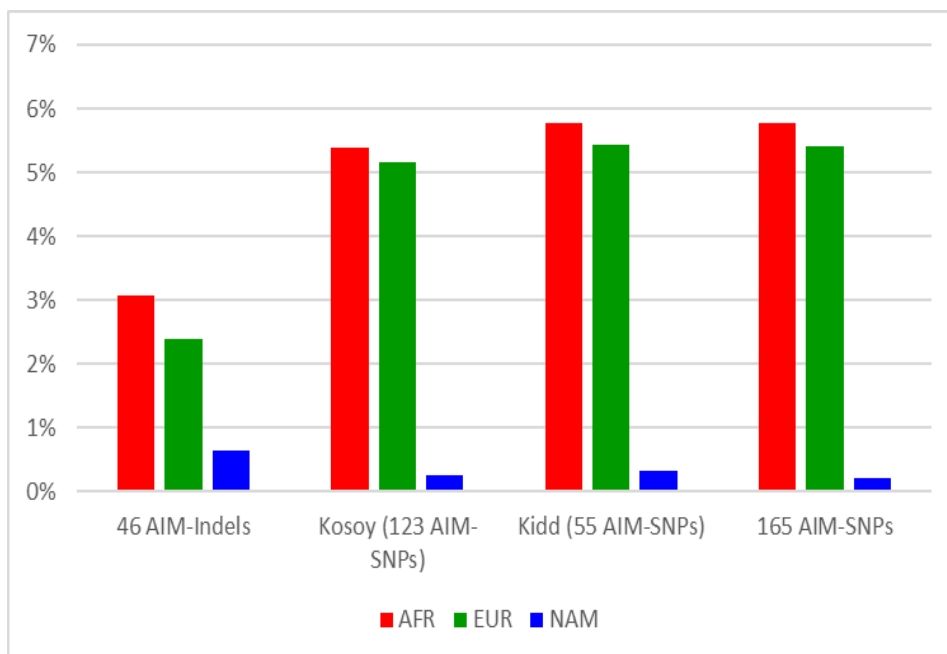
Figura 17 - Diferenças médias entre os irmãos com os diferentes painéis de ancestralidade



Legenda: 46 AIM-Indels (Pereira *et al.*, 2012), 123 AIM-SNPs (Kosoy *et al.*, 2009), 55 AIM-SNPs (Kidd *et al.*, 2014) e 165 AIM-SNPs (HID-Ion™ AmpliSeq - Kosoy *et al.*, 2009 + Kidd *et al.*, 2014).

As menores diferenças encontradas para o grupo de 165 AIM-SNPs poderão estar relacionadas com uma maior eficiência destes marcadores para determinar ancestralidade. No entanto, uma aproximação dos valores de ancestralidade entre indivíduos poderá, também, resultar de uma menor capacidade dos marcadores para separar as populações de origem. Neste caso, espera-se que uma menor variância intrapopulacional esteja associada a grupos de marcadores com menor eficiência na determinação da ancestralidade. Sendo assim, para os indivíduos não relacionados destas 20 famílias foi feita também uma análise de variância, partindo do pressuposto que, um painel de ancestralidade que é capaz de detectar a variabilidade individual em uma população ou conjunto de amostras, deve apresentar variância elevada. Foi observado que os sub-painéis que compoem o painel de ancestralidade HID-Ion™ AmpliSeq apresentaram os maiores valores de variância em comparação com os 46 AIM-Indels, exceto para a ancestralidade ameríndia, onde o painel 46 AIM-Indels apresentou o maior valor (**Figura 18**).

Figura 18 - Análise de variância entre diferentes painéis de ancestralidade, para indivíduos não relacionados (pais e mães) da população do Rio de Janeiro



Legenda: 46 AIM-Indels (Pereira *et al.*, 2012), 123 AIM-SNPs (Kosoy *et al.*, 2009), 55 AIM-SNPs (Kidd *et al.*, 2014) e 165 AIM-SNPs (HID-Ion™ AmpliSeq - Kosoy *et al.*, 2009 + Kidd *et al.*, 2014).

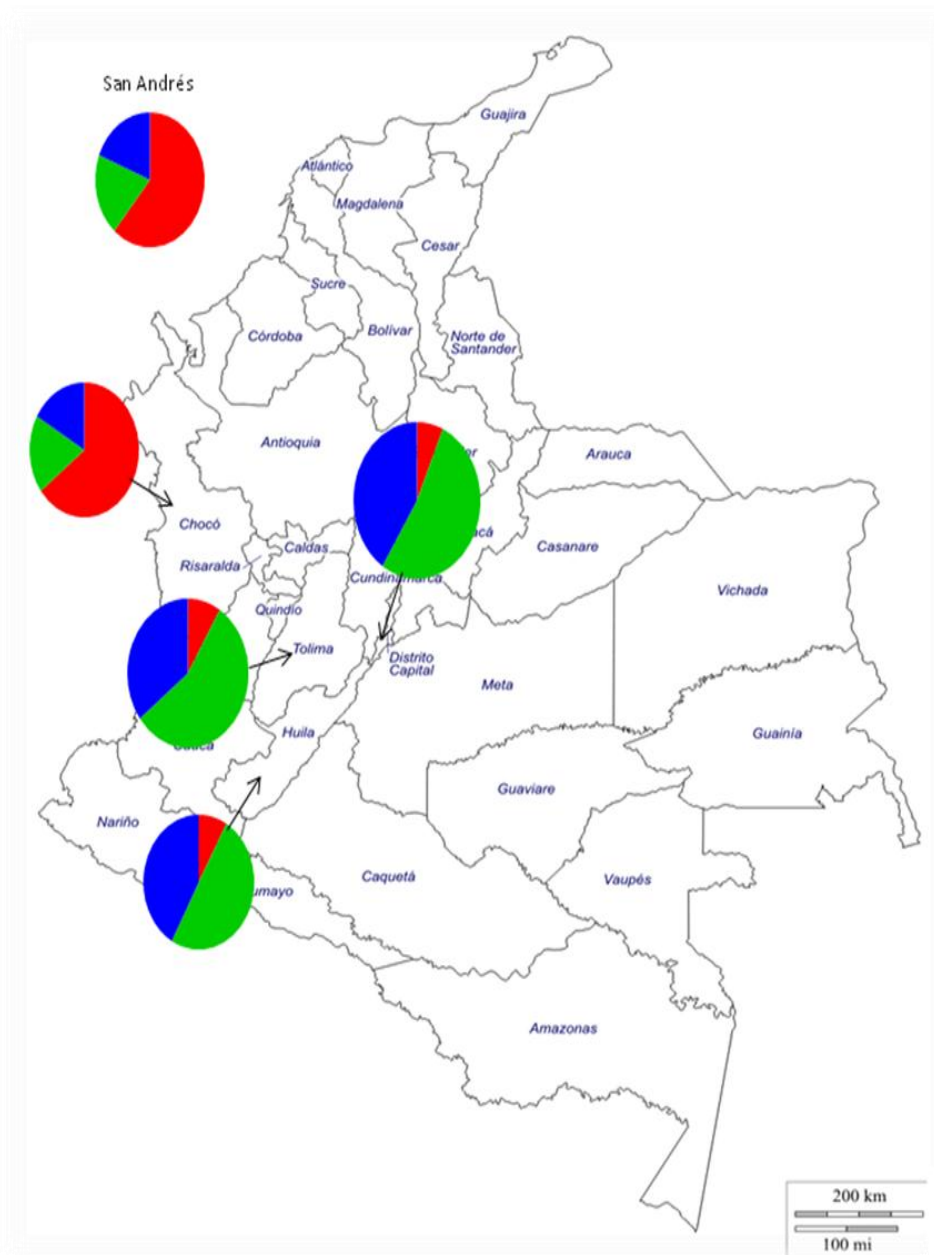
### 3.2 Colômbia

Neste trabalho, foram analisadas 188 amostras de cinco regiões da Colômbia: Bogotá, Chocó, Tolima, Huila e San Andrés. Os resultados obtidos estão expostos na **Figura 19**. Bogotá, Tolima e Huila apresentam uma maior proporção de ancestralidade europeia, seguida pela proporção de ancestralidade nativa americana e em menor proporção a ancestralidade africana. Já Chocó e San Andrés apresentam maior proporção de ancestralidade africana, seguida pelas proporções de ancestralidade europeia e nativa.

Após esta primeira análise, foi feita uma comparação entre as amostras presentes neste trabalho e as amostras do trabalho de Ossa e colaboradores (2016), para determinar se existiam diferenças significativas em termos de distância genética ( $F_{st}$ ), utilizando o *software Arlequin v3.5.2.2* (**Tabela 3**). As amostras de Chocó estão mais próximas da população africana enquanto que as outras populações estão mais próximas das populações ancestrais europeia e ameríndia. Quando comparadas as amostras dos dois trabalhos por regiões, observou-se que os valores de  $F_{st}$  variaram de 0,2 % a 1,8 %, sem diferenças significativas

(probabilidades de não diferenciação,  $p > 0,05$ ). Estes resultados demonstram que o grupo de amostras utilizadas neste trabalho não difere significativamente das amostras analisadas por Ossa e colaboradores, em 2016. A amostra de San Andrés apresenta resultados diferentes do esperado para a região do Caribe, apresentando perfil de ancestralidade parecido com a região do Pacífico.

Figura 19 - Proporções de ancestralidade africana, europeia e nativa americana em Bogotá, Chocó, Tolima, Huila e San Andrés



Departamento	N	AFR	EUR	NAM
Bogotá	49	0,067	0,525	0,408
Chocó	70	0,646	0,187	0,167

Tolima	24	0,088	0,558	0,354
Huila	31	0,084	0,498	0,418
San Andrés	14	0,613	0,196	0,190

Legenda: N- número de indivíduos, AFR – componente africano, EUR – componente europeu e NAM – componente nativo americano.

Tabela 3 - Análise de distâncias genéticas (Fst) incluindo as três populações ancestrais (África, Europa e Nativa Americana), utilizando o *software Arlequin v3.5.2*.

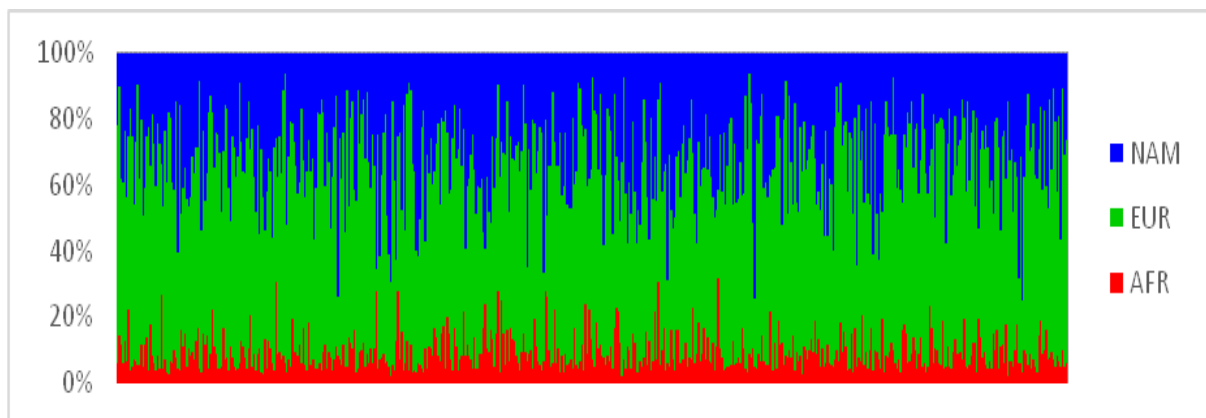
	África	Europa	Nativa Americana	Bogotá_1	Bogotá_2	Chocó_1	Chocó_2	Huila_1	Huila_2	Tolima_1	Tolima_2
África	*	<5E-06	<5E-06	<5E-06	<5E-06	<5E-06	<5E-06	<5E-06	<5E-06	<5E-06	<5E-06
Europa	0,366	*	<5E-06	<5E-06	<5E-06	<5E-06	<5E-06	<5E-06	<5E-06	<5E-06	<5E-06
Nativa americana	0,453	0,318	*	<5E-06	<5E-06	<5E-06	<5E-06	<5E-06	<5E-06	<5E-06	<5E-06
Bogotá_1	0,321	0,051	0,151	*	0,223	<5E-06	<5E-06	0,108	0,287	0,281	0,451
Bogotá_2	0,325	0,074	0,134	0,002	*	<5E-06	<5E-06	0,007	0,812	0,403	0,090
Chocó_1	0,078	0,305	0,409	0,225	0,217	*	0,250	<5E-06	<5E-06	<5E-06	0,001
Chocó_2	0,057	0,217	0,288	0,155	0,149	0,018	*	<5E-06	<5E-06	<5E-06	<5E-06
Huila_1	0,326	0,067	0,150	0,002	0,007	0,225	0,150	*	0,066	0,164	0,016
Huila_2	0,316	0,078	0,126	0,001	0,003	0,200	0,135	0,005	*	0,679	0,249
Tolima_1	0,303	0,064	0,140	0,001	1,8E-04	0,207	0,135	0,002	0,001	*	0,132
Tolima_2	0,333	0,062	0,165	4,4E-04	0,005	0,203	0,146	0,010	0,004	0,004	*

Legenda: As amostras Bogotá\_1, Chocó\_1, Huila\_1 e Tolima\_1 pertencem ao trabalho de Ossa *et al.*, (2016). Na diagonal inferior estão os valores de Fst e na diagonal superior estão os valores de probabilidade de não diferenciação. O nível de significância utilizado foi de 0,05. Em verde estão indicados os resultados obtidos entre amostras de uma mesma região.

### 3.3 Paraguai

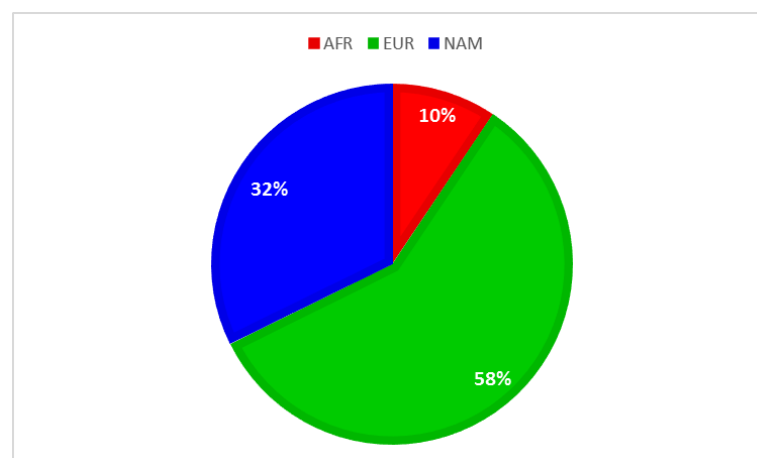
No presente trabalho foram analisadas 522 amostras da região oriental do Paraguai. Os resultados para estas amostras estão expostos nas **Figuras 20 e 21**, tendo-se observado um maior percentual de ancestralidade europeia, seguida pela ancestralidade ameríndia e em menor proporção a ancestralidade africana.

Figura 20 - Perfil de ancestralidade individual em 522 indivíduos não relacionados da população do Paraguai, utilizando 46 AIM-Indels (Pereira *et al.*, 2012)



Legenda: Ancestralidade africana (AFR), europeia (EUR) e ameríndia (NAM).

Figura 21 - Média das proporções de ancestralidade para a população do Paraguai utilizando 46 AIM-Indels (Pereira *et al.*, 2012).

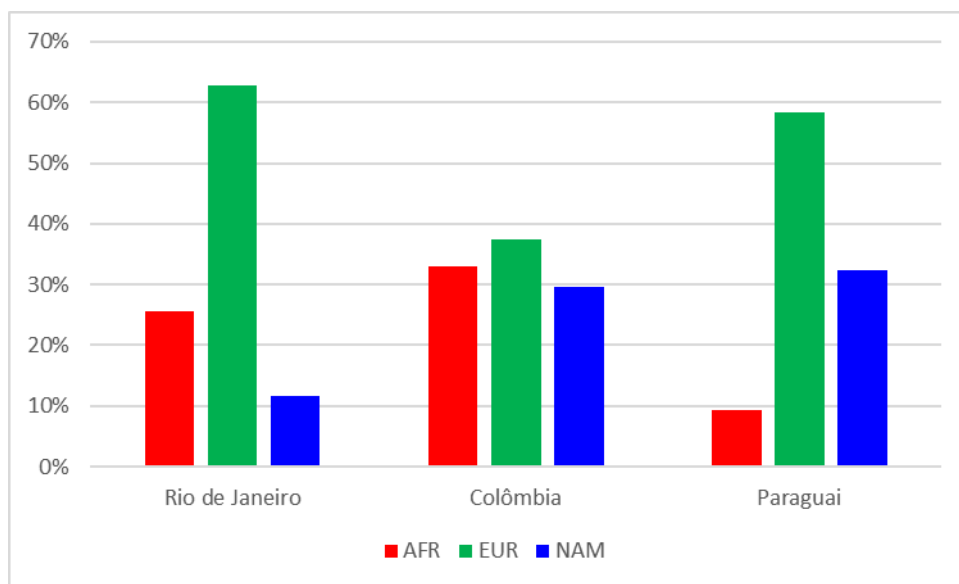


Legenda: Ancestralidade africana (AFR), europeia (EUR) e ameríndia (NAM).

### 3.4 Comparação dos perfis de ancestralidade

Foram comparados os diferentes perfis de ancestralidade para os indivíduos não relacionados do Rio de Janeiro, Colômbia e Paraguai. As amostras do Rio de Janeiro e do Paraguai apresentaram os maiores valores de ancestralidade europeia. Para o componente africano, a Colômbia apresentou o maior valor, seguido pelo Rio de Janeiro. Já para o componente nativo americano, Colômbia e Paraguai apresentaram perfis semelhantes, enquanto que o Rio de Janeiro apresentou o valor mais baixo.

Figura 22 - Perfil de ancestralidade do Rio de Janeiro, Colômbia e Paraguai utilizando 46 AIM-Indels (Pereira *et al.*, 2012)



Legenda: Rio de Janeiro - n=1016, Colômbia - n=188 e Paraguai - n=522. Ancestralidade africana (AFR), europeia (EUR) e ameríndia (NAM).

## 4 DISCUSSÃO

A história de formação das populações da América do Sul tem sido caracterizada pela extensa mistura de nativos americanos e imigrantes de todo o mundo, um processo iniciado pela expansão colonial espanhola a partir de 1492 (Burkholder *et al.*, 2003; Adhikari *et al.*, 2016). As populações da América Latina apresentam padrões variáveis de mistura entre os três principais grupos ancestrais que a formaram, além de apresentarem variações importantes em cada região, que são resultado de diferentes dinâmicas de ocupação do território, que ocorreram a partir do século XVI. Essas características devem ser levadas em consideração nos estudos populacionais, e particularmente nos estudos de associação onde a estruturação populacional pode produzir resultados espúrios. O principal efeito da estratificação populacional, particularmente em populações que são resultado da mistura recente entre grupos continentais, é o aumento da taxa de falsos positivos em estudos de caso-controle ou traços quantitativos (Hoggart *et al.*, 2003).

Neste trabalho foram ampliados os dados sobre ancestralidade genética da população do Rio de Janeiro utilizando 46 AIM-Indels (Pereira *et al.*, 2012), onde foi observado que há um predomínio da ancestralidade europeia, seguida da ancestralidade africana e em menor proporção, a ancestralidade nativa americana (63 %, 25 % e 12 % respectivamente). Apesar das diferenças nos valores percentuais, estes resultados estão de acordo com a literatura (Pena *et al.*, 2011; Manta *et al.*, 2013; Salzano e Sans, 2014).

As diferenças observadas entre os resultados deste trabalho e os dados encontrados na literatura parecem ser reflexo da utilização de diferentes painéis de marcadores genéticos que variam no tipo de polimorfismo (por exemplo: SNPs e Indels) e no número de marcadores utilizados (Lins *et al.*, 2010; Salzano e Sans, 2014). Ainda assim, em todos os trabalhos há um predomínio da ancestralidade europeia, que é mais elevada na região sul do Brasil, seguida da ancestralidade africana, mais elevada no nordeste e em menor proporção a ancestralidade nativa americana, que apresenta valores mais expressivos no norte do país, podendo superar a ancestralidade africana como segunda mais prevalente nesta região (Pena *et al.*, 2011; Manta *et al.*, 2013; Salzano e Sans, 2014).

Diferentes estratégias de amostragem também podem interferir nos valores de ancestralidade média, como foi observado quando foram avaliadas amostras obtidas de voluntários que participaram de projetos de pesquisa no LDD no período de 2014-2016 e

amostras dos casos de paternidade realizados nos períodos de 2002-2004 e 2010-2016. Para estes grupos de amostras foi observado que os voluntários apresentam um maior percentual de ancestralidade europeia, seguida da ancestralidade africana e nativa americana. Perfil semelhante foi observado para os casos de paternidade no período 2002-2004. No entanto, a partir de 2010 houve uma redução da ancestralidade europeia e um aumento da ancestralidade africana neste grupo de amostras. Manta e colaboradores (2013) afirmam que em qualquer estudo genético as amostras representam um grupo de referência e não a população real, a menos que sejam coletadas aleatoriamente. No entanto, as amostras coletadas aleatoriamente na população não necessariamente serão mais úteis, dependendo da finalidade do estudo. Portanto, as estratégias de amostragem devem refletir o objetivo do projeto ou aplicação da pesquisa. Lins e colaboradores (2011) observaram que amostras de indivíduos oriundos de locais com diferentes níveis de urbanização, podem apresentar diferentes perfis de ancestralidade genética. Portanto, acredita-se que as diferenças observadas no presente trabalho parecem ser reflexo da origem dos dois grupos avaliados, onde os voluntários são indivíduos residentes na região metropolitana do Rio de Janeiro, enquanto que as amostras de casos de paternidade são de diversas regiões do Estado.

Após a análise dos indivíduos não relacionados, foram avaliadas 121 famílias compostas por pai, mãe e dois filhos utilizando os 46 AIM-Indels (Pereira *et al.*, 2012). Do total de famílias, 110 foram obtidas dos casos de paternidade do LDD e 11 famílias foram voluntárias. Como discutido anteriormente, os indivíduos que fazem parte dos casos de paternidade apresentaram ancestralidade africana média (37%) superior a observada para os indivíduos não relacionados da região metropolitana do Rio de Janeiro, ainda assim o componente europeu prevaleceu seguido do componente africano e em menor proporção o componente ameríndio. Na comparação da média de ancestralidade entre pais e filhos foi observado que para o componente africano as médias ficaram próximas, em torno de 37%, já para o componente europeu e para o componente nativo, observamos diferenças de 4% e 5%, respectivamente. As diferenças encontradas podem ser causadas por um erro na estimativa gerado pelo conjunto de marcadores utilizados. Os valores que ficam próximos dos 33% tendem a apresentar menos variação do que os valores que estão muito abaixo ou muito acima deste valor, como demonstrado por Aquino (2016). Ainda assim, fizemos uma comparação entre as diferenças observadas nos pais e as diferenças entre irmãos, com o objetivo de avaliar se existia correlação entre estes valores, assumindo que as maiores diferenças entre irmãos seriam observadas quando os pais apresentassem perfis de ancestralidade discrepantes. Apesar de haver uma variabilidade nas proporções de ancestralidade entre irmãos, não observamos

correlação com as diferenças observadas entre os pais, ou seja, o aumento das diferenças de ancestralidade nos pais parece não estar relacionado com a variabilidade observada entre os irmãos.

Leite e colaboradores (2011) analisaram 21 AIMs em 86 pares de irmãos e observaram que, para a maioria dos irmãos, a média das diferenças entre as estimativas de ancestralidade foi inferior a 10%, superior a média de 7,3 % encontrada neste trabalho para o 46 AIM-Indels, em amostras do Rio de Janeiro. No entanto eles ressaltam que o número de marcadores utilizados não foi suficientemente alto para obter estimativas precisas a nível individual. O que ressalta a importância da utilização de painéis com um número maior de marcadores.

Foram selecionadas ainda 20 famílias para avaliar a eficiência de dois painéis contendo marcadores informativos de ancestralidade (AIMs) para estimar as proporções de ancestralidade em indivíduos miscigenados. Estes painéis são: 46 AIM-Indels descritos por Pereira e colaboradores, em 2012, e 165 AIM-SNPs do painel de ancestralidade HID-Ion™ AmpliSeq da *Ion Torrent*, que é a junção de dois conjuntos de marcadores descritos previamente (123 AIM-SNPs descritos por Kosoy *et al.*, 2009; 55 AIM-SNPs descritos por Kidd *et al.*, 2012).

Primeiramente foram analisados somente os indivíduos não relacionados (pais e mães) com os diferentes conjuntos de painéis. Para o painel *Ion Torrent* os resultados foram avaliados tanto para o total de marcadores (165 AIM-SNPs) quanto para os sub-painéis deste *kit*, pois como descrito anteriormente, os dois conjuntos de marcadores que o compõem apresentam diferenças que devem ser consideradas na análise. Foi observado que o painel 46 AIM-Indels e o painel 123 AIM-SNPs apresentaram o mesmo perfil para as proporções de ancestralidade europeia e africana. O painel com 55 AIM-SNPs parece superestimado a proporção de ancestralidade africana, enquanto que no conjunto com 165 AIM-SNPs as proporções de ancestralidade destes componentes estavam equilibradas. O componente nativo americano se mostrou reduzido no painel com 123 AIM-SNPs e no total. Já os painéis 46 AIM-Indels e 55 AIM-SNPs apresentaram valores semelhantes. As populações de referência utilizadas para inferir a ancestralidade ameríndia, parecem estar mais bem representadas nestes dois últimos painéis.

Foi feita uma análise de variância para as 20 famílias, partindo do pressuposto que um painel de marcadores que é capaz de detectar maior variabilidade individual em uma população ou conjunto de amostras, deverá ser mais eficiente na estimativa dos valores de ancestralidade. Foi observado que os sub-painéis que compoem o painel de ancestralidade HID-Ion™ AmpliSeq apresentam os maiores valores de variância em comparação com os 46

AIM-Indels, o que indica que este painel é mais eficiente para detectar valores de ancestralidade individual mais extremos que o painel com 46 AIM-Indels, exceto para a ancestralidade nativa americana, onde os 46 AIM-Indels parece ser mais eficiente.

Quando se comparou a média de ancestralidade dos pais com a ancestralidade observada em cada um dos filhos, foi observado que o painel 46 AIM-Indels apresenta perfil semelhante aos 55 AIM-SNPs. Este resultado parece estar relacionado ao número de marcadores utilizados, que é bastante inferior ao do painel total. Já o painel com 123 AIM-SNPs apresenta um perfil semelhante ao painel total (165 AIM-SNPs), pois, obviamente, existe sobreposição para um maior número de marcadores (123 AIM-SNPs). O painel com 165 AIM-SNPs foi o mais eficiente, de acordo com os nossos critérios, em reduzir as diferenças observadas entre pais e filhos, com um aumento da variância nos valores de ancestralidade da amostra estudada.

Em seguida, foi comparada a média das diferenças entre os irmãos e foi observado que as menores diferenças foram obtidas com o painel com 165 AIM-SNPs e as maiores diferenças foram observadas com o painel 46 AIM-Indels. Em teoria, irmãos deveriam apresentar poucas diferenças entre si (Leite *et al.*, 2011) e suas proporções de ancestralidade deveriam estar próximas a média de ancestralidade dos pais, neste contexto, o painel de ancestralidade HID-Ion™ AmpliSeq da *Ion Torrent* se mostrou capaz de inferir a ancestralidade com maior eficiência para este grupo de amostras. Vale ressaltar que o painel HID-Ion™ AmpliSeq da *Ion* foi desenvolvido para fornecer a provável origem geográfica de indivíduos em um contexto forense. Portanto, não existem dados na literatura utilizando este painel para inferir ancestralidade em populações miscigenadas. Ainda assim, ele se mostrou eficiente para esta finalidade.

A Colômbia é a porta de entrada (via terrestre) para a América do Sul, pois está ligada ao Istmo do Panamá. Devido a sua localização o país tem sido sujeito, desde os tempos pré-históricos, a um intenso fluxo genético e cultural que se reflete na diversidade de grupos étnicos que são encontrados no país (Arango *et al.*, 2004). Neste trabalho foram avaliadas 188 amostras de cinco regiões da Colômbia: Bogotá, Tolima, Huila (Região dos Andes), Chocó (Região do Pacífico) e San Andrés (Região do Caribe). Foi observado que as regiões localizadas nos Andes apresentam perfil de ancestralidade predominantemente europeu, seguida da ancestralidade nativa americana e em menor proporção a ancestralidade africana. Já a amostra da região do Pacífico (Chocó) apresentam perfil predominantemente africano, seguido da ancestralidade europeia e nativa americana. Estes resultados estão de acordo com o observado na literatura (Ruiz-Linares *et al.*, 2014; Salzano e Sans, 2014; Adhikari *et al.*,

2016; Ossa *et al.*, 2016). As amostras de San Andrés (Região do Caribe) apresentam perfil de ancestralidade predominantemente africano, seguido pelas ancestralidades europeia e nativa americana. Apesar do pequeno número de amostras analisadas para esta região (14 amostras), este resultado corrobora os resultados obtidos por Alonso e Usaqué (2012) que analisaram marcadores do cromossomo Y e observaram um predomínio de haplótipos de origem africana em indivíduos das ilhas de San Andrés e Providencia. Estes resultados mostram que existem diferenças entre as regiões continentais descritas por Ossa e colaboradores (2016) e as regiões insulares do Caribe (Alonso e Usaqué, 2012). Nas regiões continentais do Caribe há um predomínio da ancestralidade europeia, seguida das ancestralidades africana e nativa americana. Já na região insular, parece haver um predomínio da ancestralidade africana, no entanto, é necessário aumentar o número de amostras analisadas para esta região da Colômbia.

Foi feita análise de distância genética ( $F_{st}$ ) para avaliar se o grupo de amostras utilizados neste trabalho era significativamente diferente dos grupos avaliados anteriormente, por nosso grupo, no trabalho de Ossa e colaboradores (2016). Nesta análise, excluimos as amostras de San Andrés, pois esta região não foi analisada no trabalho anterior. Quando se comparou todos os grupos de amostras com as populações de referência africana, europeia e nativa americana, foi observado que as amostras de Chocó eram mais próximas da população africana que das outras populações referência. Já na análise comparando par a par as amostras incluídas neste trabalho com as descritas por Ossa *et al.* (2016) para uma mesma região (exemplo: as amostras de Bogotá do trabalho de e as amostras de Bogotá deste trabalho), foi observado que os valores de  $F_{st}$  variaram de 0,002 a 0,018 e as probabilidades de não diferenciação variaram de 0,066 a 0,250. Este resultado demonstra que as amostras utilizadas não diferem significativamente das utilizadas por Ossa *et al.* (2016), para um nível de significância de  $p$  de 0,05. As amostras de Chocó são significativamente mais distantes das amostras das outras regiões analisadas, tanto neste trabalho quanto no trabalho de Ossa *et al.* (2016), devido a uma maior proximidade com a amostra da África.

O Paraguai é um país localizado no centro da América do Sul, fazendo fronteira com Brasil, Argentina e Bolívia. Está dividido em duas macrorregiões: Ocidental e Oriental, separadas pelo Rio Paraguai (Ferreira e de Herreros, 2014). A região oriental é a mais populosa, concentrando cerca de 97 % da população.

A população do Paraguai resulta da mistura entre europeus (espanhóis), que chegaram ao país no século XVI, e indígenas guaranis, etnia que predominava a região na chegada dos europeus. Embora não haja dados oficiais sobre a composição étnica do país, pois o DGEEC

(Dirección General de Estadística, Encuestas y Censos de La Secretaría Técnica de Planificación de La Presidencia de La República de Paraguay) não inclui em seus levantamentos dados sobre etnia, estima-se que 95 % da população paraguaia seja miscigenada (Benitez *et al.*, 1995). Dados sobre ancestralidade nesta população são escassos e não utilizam marcadores informativos de ancestralidade (Benitez *et al.*, 1995; Benitez *et al.*, 2011; Dipierre *et al.*, 2011). Neste trabalho foi apresentado, pela primeira vez, o perfil de ancestralidade genética da população miscigenada do Paraguai, utilizando 46 AIM-Indels (Pereira *et al.*, 2012).

Foram avaliados 522 indivíduos da região oriental do Paraguai e observou-se que há um predomínio de ancestralidade europeia, seguida pela ancestralidade nativa americana e em menor proporção a ancestralidade africana (58 %, 32 % e 10 %, respectivamente). Os resultados obtidos para a ancestralidade europeia e nativa americana estão de acordo com o contexto histórico de formação desta população (Paoli, 2010). Já o perfil encontrado para a ancestralidade africana foi maior do que o esperado, considerando a dificuldade em se obter informações sobre indivíduos de origem africana que tenham chegado ao Paraguai no início da colonização do país. Segundo Correa (2012), os primeiros negros chegaram ao Paraguai ainda no século XVI vindos, muitas vezes, do Brasil através de contrabando de escravos e, embora a escravidão no Paraguai não tenha atingido a mesma magnitude ou importância que em outras regiões da América do Sul, não foi nada desprezível o papel que ela teve no desenvolvimento do país tanto no período colonial quanto após a independência. Dados históricos apontam também para a entrada recente de imigrantes vindos de diversas regiões da Europa e da América do Sul (inclusive do Brasil), sendo que estes últimos poderão ter contribuído também para o perfil de ancestralidade africana observada. Portanto os resultados encontrados no presente trabalho parecem refletir os dados históricos relativos ao processo de miscigenação que levou a formação da atual população do Paraguai.

Por fim, comparou-se o perfil de ancestralidade encontrado nas amostras do Rio de Janeiro, Colômbia e Paraguai. Foi observado que a ancestralidade europeia é predominante principalmente no Rio de Janeiro e no Paraguai. A ancestralidade africana foi maior na Colômbia seguida do Rio de Janeiro, enquanto que a ancestralidade nativa americana foi maior no Paraguai, seguido da Colômbia. É importante ressaltar que esta análise foi realizada somente como um painel geral dos grupos de amostras utilizadas neste trabalho e para uma avaliação criteriosa de cada uma das regiões é necessário considerar a extensa estrutura genética existente dentro e entre os países sul-americanos (Ruiz-Linares *et al.*, 2014). Nomeadamente, as amostras da Colômbia incluíram duas regiões de forte contribuição

africana, sendo que nas restantes regiões do país o componente africano é geralmente menor que o observado no Rio de Janeiro (Ossa *et al.*, 2016).

## CONCLUSÃO

- a) Foi ampliada a base de dados sobre ancestralidade da população do Rio de Janeiro mediante a análise de 46 AIM-Indels para 957 indivíduos não relacionados e as frequências para os marcadores utilizados foram atualizadas. Observou-se que a ancestralidade europeia é predominante, seguida da ancestralidade africana e em menor proporção, a ancestralidade nativa americana, estando estes resultados de acordo com a literatura;
- b) Foram comparadas diferentes estratégias de amostragem e observou-se que elas podem provocar diferenças nos valores de ancestralidade que, em nossas amostras, parecem ser causadas pelo local de origem das mesmas, onde amostras coletadas de voluntários residentes na região metropolitana do Rio de Janeiro apresentam ancestralidade europeia superior às amostras coletadas em exames de paternidade incluindo outras regiões do Estado;
- c) Observou-se que o painel HID-Ion™ AmpliSeq – Ion Torrent apresentou os maiores valores de variância e reduziu as diferenças de ancestralidade entre irmãos sendo, portanto, mais eficiente para estimar a ancestralidade em indivíduos miscigenados da população do Rio de Janeiro;
- d) Foi ampliada a base de dados sobre ancestralidade em 5 amostras da população da Colômbia (Bogotá, Tolima, Huila, Chocó e San Andrés), analisando 188 amostras com 46 AIM-Indels, sendo que San Andrés foi estudada pela primeira vez neste trabalho. Os resultados encontrados estão de acordo com a literatura. Também foram comparadas as diferentes amostras da Colômbia estudadas neste trabalho, entre si e com outras já publicadas para os mesmos marcadores, e observou-se que as amostras utilizadas neste trabalho não diferem significativamente das amostras estudadas por Ossa et al. (2016) para as mesmas regiões;
- e) Foi apresentado, pela primeira vez, o perfil de ancestralidade genética da população miscigenada do Paraguai utilizando 46 AIM-Indels (Pereira et al., 2012). Observou-se um predomínio da ancestralidade europeia, seguida da ancestralidade nativa americana e, em menor proporção, a ancestralidade africana. Estes resultados estão de acordo com os dados históricos encontrados para esta população;

- f) Foram comparados os perfis genéticos e padrões de miscigenação das populações do Brasil (Rio de Janeiro), Colômbia e Paraguai. Observou-se a heterogeneidade já esperada entre estes grupos populacionais, que refletem o complexo processo de miscigenação que originou as populações atuais da América Latina.

## REFERÊNCIAS

- ALONSO, L. A.; USAQUÉM, W. *Y-chromosome and surname analysis of the native islanders of San Andrés and Providencia (Colombia)*. - J. Comp. Hum. Biol. 2012. 64(1):71-84p.
- AQUINO, J. G. *Avaliação de diferentes estratégias na determinação da ancestralidade genética e sua aplicação no estudo de populações nativas e miscigenadas da América do Sul*. 2016. 198f. Tese (Doutorado em Biociências) – Instituto de Biologia Roberto Alcântara Gomes, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2016.
- ARANGO, O. R. *et al. Los pueblos indígenas de Colombia en el umbral del nuevo milenio: población, cultura y territorio: bases para el fortalecimiento social y económico de los pueblos indígenas*. Bogotá: Departamento Nacional de Planeación; 2004.
- BAMSHAD, M. *et al. Deconstructing the relationship between genetics and race*. Nat. Rev. Genetics. 2004. v. 5, 598-609p.
- BENITEZ, O. *et al. Métissage hispano-indien en Amérique du Sud: essai de compréhension grâce à l'analyse sanguine du système HLA au Paraguay*. Hématologie. 1995. 1:437–439p.
- BENITEZ, O. *et al. HLA polymorphism in a Guaraní-Indian population from Paraguay and its usefulness for the Hispano-Indian admixture study in Paraguay*. International Journal of Immunogenetics. 2011. v. 38, 7-11p.
- BETHELL, L. (Ed.). *The Cambridge History of Latin America*. Cambridge: Cambridge University Press. 1984. 632p.
- BOURGEON, L.; BURKE, A.; HIGHAM, T. *Earliest Human Presence in North America Dated to the Last Glacial Maximum: New Radiocarbon Dates from Bluefish Caves, Canadá*. PLoS ONE 12 (1): e0169486. 2017.
- BROWN, M. D. *et al. MtDNA Haplogroup X: an Ancient Link between Europe/Western Asia and North America*. Am. J. Hum. Genet. 1998. v. 63, 1852–1861p.
- BURKHOLDER, M. A.; JOHNSON, L. L. *Colonial Latin America*. Oxford University Press; 2003.
- CARVAJAL-CARMONA, L. G. *et al. Strong Amerind/White Sex Bias and a Possible Sephardic Contribution among the Founders of a Population in Northwest Colombia*. Am. J. Hum. Genet. 2000. v. 67, 1287–1295p.
- COLOMBO, F. *História do Paraguai*. Disponível em: <http://f1colombohistoriando.blogspot.com.br/2012/11/historia-do-paraguai.html>. Acesso em: 16 Mai. 2017.
- COMISSÃO ECONÔMICA PARA AMÉRICA LATINA E O CARIBE – CEPAL. *“Los pueblos indígenas en América Latina. Avances en el último decenio y retos pendientes para la garantía de sus derechos”*. 2014. 128f. Disponível em:

<http://www.cepal.org/es/publicaciones/37050-pueblos-indigenas-america-latina-avances-ultimo-decenio-retos-pendientes-la>. Acesso em: 12 Abr. 2017.

CONRAD, D. F. *et al.* *Origins and functional impact of copy number variation in the human genome*. *Nature*. 2010. 464: 704–712p.

CORREA, P. M. A. *Seria o paraguaio o mestiço de apenas duas cores? Uma ponte entre a histórica mistura étnica do Paraguai e alguns dos seus desafios contemporâneos*. 73f. Universidade Anhembi Morumbi. São Paulo. 2012.

DENEVAN, W. M. *The Native Population of the Americas in 1492*. Madison, University of Wisconsin Press. 1976.

DEPARTAMENTO ADMINISTRATIVO NACIONAL DE ESTADÍSTICA – DANE. Censo geral 2005. Disponível em: <http://www.dane.gov.co>. Acesso em: 19 Abr. 2017.

DIPIERRI J. *et al.* *A Study of the Population of Paraguay through Isonymy*. *Annals of Human Genetics*. 2011. v.75, 678–687p.

DIRECCIÓN GENERAL DE ESTADÍSTICA, ENCUESTAS Y CENSES DE LA SECRETARÍA TÉCNICA DE PLANIFICACIÓN DE LA PRESIDENCIA DE LA REPÚBLICA DE PARAGUAY - DGEEC. Disponível em: <http://www.dgeec.gov.py/Publicaciones/evoluciontotal.php>. Acesso em: 13 Fev. 2017.

FAGUNDES, N. J. R.; KANITZ, R.; BONATTO, S. L. *A Reevaluation of the Native American MtDNA Genome Diversity and Its Bearing on the Models of Early Colonization of Beringia*. *PLoS ONE* 3(9): e3157. 2008.

FERREIRA, C. R.; de HERREROS, M. B. *Medical Genetics in Paraguay*. *Molecular Genetics & Genomic Medicine*. 2014. 2(6):458-466p.

GREENBERG, J. H.; TURNER, C. G.; ZEGURA, S. L. *The settlement of the Americas: a comparison of the linguistic, dental and genetic evidence*. *Curr Anthropol*. 1986. v.27, 477–495p.

GRIFFITHS, A. J. F. *et al.* *Introdução à genética*. Editora: Guanabara Koogan, – 10ª edição. 2013.

HOGGART, C. J. *et al.* *Control of confounding of genetic associations in stratified populations*. *Am J Hum Genet*. 2003. 72:1492–1504p.

HOMBURGER, J. R. *et al.* *Genomic Insights into the Ancestry and Demographic History of South America*. Ed. Eduardo Tarazona-Santos. *PLoS Genetics* 11.12. 2015.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA – IBGE. Disponível em: [http://www.ibge.gov.br/home/estatistica/populacao/estimativa2016/estimativa\\_tcu.shtm](http://www.ibge.gov.br/home/estatistica/populacao/estimativa2016/estimativa_tcu.shtm). Acesso em: 06 Abr. 2017.

- INTERNATIONAL SNP MAP WORKING GROUP. *A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms*. Nature. 2001. 409: 928–933p.
- JEFFREYS, A. J.; WILSON, V.; THEIN, S. L. *Individual-specific 'fingerprints' of human DNA*. Nature. 1985. v 316, 76–79p.
- JOBLING, M. A.; GILL, P. *Encoded evidence: DNA in forensic analysis*. Nature Reviews Genetics. 2004.v. 5, 739-751p.
- KIDD, J. R. *et al. Analyses of a set of 128 ancestry informative single-nucleotide polymorphisms in a global set of 119 population samples*. Investig. Geneti. 2011.
- KIDD, K. K. *et al. Progress toward an efficient panel of SNPs for ancestry inference*. Forensic Sci. Int. Genet. 2014. v.10, 23-32p.
- KOSOY, R. *et al. Ancestry Informative Marker Sets for Determining Continental Origin and Admixture Proportions in Common Populations in America*. Human mutation 30.1. 2009. 69–78p.
- LANDER, E. S. *Initial impact of the sequencing of the human genome*. Review. Nature, 2011, v.470, 187-197p.
- LANDSTEINER, K. *Zur Kenntnis der antifermentativen, lytischen und agglutinierenden Wirkungen des Blutserums und der Lymphe*. Zentralbl. Bakteriol. 1900. v. 27, 357–362p.
- LAREU, M. V. *et al. Investigation of the STR locus HUMTH01 using PCR and two electrophoresis formats: UK and Galician Caucasian population surveys and usefulness in paternity investigations*. Forensic Sci. Int., 1994, v.66, 41-52p.
- LEITE, T. K. M. *et al. Genomic Ancestry, Self-Reported "Color" and Quantitative Measures of Skin Pigmentation in Brazilian Admixed Siblings*. Gaunt TR, ed. PLoS ONE.6(11):e27162. 2011.
- LEVY, S. *et al. The Diploid Genome Sequence of an Individual Human*. PLoS Biology. 2007.
- LINS, T. C. *et al. Genetic composition of Brazilian population samples based on a set of twenty-eight ancestry informative SNPs*. Am J Hum Biol. 2010. 22: 187-192p.
- LINS, T. C. *et al. Genetic heterogeneity of self-reported ancestry groups in an admixed Brazilian population*. J Epidemiol. 2011. 21: 240-245p.
- MARDIS, E. R. *Next-generation sequencing platforms*. Annu. Rev. Anal. Chem. 2013. v. 6, 287-303p.
- MARIAN, A. J. *Molecular Genetic Studies of Complex Phenotypes*. Translational Research. 2012. v.159(2), 64–79p.
- MISHRA, P.; CHAN, D. C. *Mitochondrial dynamics and inheritance during cell division, development and disease*. Nature reviews Molecular cell biology. 2014. 15(10):634-646p.

MULLANEY, J. M. *et al.* *Small insertions and deletions (INDELs) in human genomes.* Hum Mol Genet. 2010. 19(R2): 131-136p.

NASSIR, R. *et al.* *An ancestry informative marker set for determining continental origin: validation and extension using human genome diversity panels.* B.M.C. Genet. 2009.

O'ROURKE D. H.; RAFF, J. A. *The Human Genetic History of the Americas: The Final Frontier,* Current Biology, Volume 20, Issue 4,23 February. 2010. R202-R207p.

OSSA, H. *et al.* *Outlining the Ancestry Landscape of Colombian Admixed Populations.* Ed. Tzen-Yuh Chiang. *PLoS ONE* 11.10: e0164414. 2016.

PAOLI, J. B. R. *La Colonización del Paraguay (1537 - 1680).* Portal Guaraní. Disponível em: [http://www.portalguarani.com/1067\\_juan\\_bautista\\_rivarola\\_paoli/11925\\_la\\_colonizacion\\_del\\_paraguay\\_1537\\_\\_1680\\_por\\_juan\\_bautista\\_rivarola\\_paoli.html](http://www.portalguarani.com/1067_juan_bautista_rivarola_paoli/11925_la_colonizacion_del_paraguay_1537__1680_por_juan_bautista_rivarola_paoli.html). 2010. Acesso em: 13 Jun. 2017.

PENA, S. *et al.* *The Genomic Ancestry of Individuals from Different Geographical Regions of Brazil Is More Uniform Than Expected.* Ed. Henry Harpending. *PLoS ONE* 6.2. 2011.

PHILLIPS, C. *et al.* *Inferring ancestral origin using a single multiplex assay of ancestry-informative marker SNPs.* Forensic Sci. Int. Genetics. 2007. v. 1, 273-80p.

PHILLIPS, C. *et al.* *Ancestry analysis in the 11-M Madrid bomb attack investigation.* *PLoS ONE* 4: e6583. 2009.

PHILLIPS, K.; MCCALLUM, N.; WELCH, L. *A comparison of methods for forensic DNA extraction: Chelex-100® and the QIAGEN DNA Investigator Kit (manual and automated),* Forensic Science International: Genetics. 2012. v. 6, 282-285p.

KAYSER, M.; KNIJFF, P. *Improving human forensics through advances in genetics, genomics and molecular biology.* Nat Rev Genet 12: 179–192. 2011.

PEREIRA, R. *et al.*; *Straightforward Inference of Ancestry and Admixture Proportions through Ancestry-Informative Insertion Deletion Multiplexing.* *PLoS One* , v. 2 (1), p e29684; 2012.

PEREÑA, L. *Genocidio en América.* Mapfre, Madrid, 401 pp. 1992.

PRITCHARD, J. K.; DONNELLY, P. *Case-control studies of association in structured or admixed populations.* Theor. Popul. Biol. 2001.

PRITCHARD, J. K. *et al.*; *Inference of Population Structure Using Multilocus Genotype Data;* v. 155; p. 945-59, 2000.

RODAS, C.; GELVEZ, N.; KEYEUX, G. *Mitochondrial DNA Studies Show Asymmetrical Amerindian Admixture in Afro-Colombian and Mestizo Populations.* Hum Biol. 2003;75: 13–30p.

- ROJAS, W. *et al.* Genetic make up and structure of Colombian populations by means of uniparental and biparental DNA markers. *Am J Phys Anthropol.* 2010;143: 13–20p.
- ROTHHAMMER, F.; DILLEHAY, T. D. *The late Pleistocene colonization of South America: an interdisciplinary perspective.* *Ann Hum Genet.* 2009. 73: 540–9p.
- RUIZ-LINARES, A. *et al.* Admixture in Latin America: Geographic Structure, Phenotypic Diversity and Self-Perception of Ancestry Based on 7,342 Individuals. *PLoS Genet* 10(9): e1004572. 2014.
- SALOUM, F. M. N., *et al.* Revisiting the Genetic Ancestry of Brazilians Using Autosomal AIM-Indels. *PLoS ONE* 8(9): e75145. 2013.
- SALZANO, F. M.; BORTOLINI, M. C. *The Evolution and Genetics of Latin American Populations.* Cambridge University Press, Cambridge. 2002. 512 pp.
- SALZANO, F. M.; SANS, M. *Interethnic admixture and the evolution of Latin American populations.* Review Article. *Genetics and Molecular Biology*, 37, 1 (suppl). 2014. 151-170p.
- SANTANGELO, R. *Typing of ancestry informative markers in population samples from Ecuador with next generation sequencing methods.* (Tese) Laurea Magistralis in Molecular Biology and Genetics. Università Degli Studi di Pavia. 2016. 108p.
- SANS, M. *Admixture studies in Latin America: From the 20th to the 21st century.* *Hum Biol* 72:155-177. 2000.
- SCHURR, T. G. *et al.* Amerindian mitochondrial DNAs have rare Asian mutations at high frequencies, suggesting they derived from four primary maternal lineages. *Am. J. Hum Genet.* 1990. 46: 613–623p.
- THOMAS, H. *The Slave Trade.* Simon and Schuster. 1997.
- TURNER, C. G. *Late Pleistocene and Holocene population history of East Asia based on dental variation.* *Am J Phys Anthropol.* 1987. 73, 305–321p.
- VERANO, V. W.; UBELAKER, D. H. (Orgs.). *Disease and demography in the Americas.* Washington: Smithsonian Institution. 1992.
- VULLO, C. *et al.*; Genetic polymorphism of 22 autosomal STR markers in Paraguay. *Forensic Science International: Genetics*, Volume 25, November. 2016. e16-e17p.
- WEBER, J. L. *et al.* Human Diallelic Insertion/Deletion Polymorphisms. *American Journal of Human Genetics.* 2002. 71(4), 854–862p.
- XAVIER, C. *et al.* Admixture and genetic diversity distribution patterns of non-recombining lineages of Native American ancestry in Colombian populations. *PloS one.* 2015.
- ZEGURA, S. L. *et al.* High-resolution SNPs and microsatellite haplotypes point to a single, recent entry of Native American Y chromosomes into the Americas. *Mol Biol Evol.* 2004. 21, 164–175p.

## ANEXO A - Termo de Consentimento Livre e Esclarecido



## TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

**LDD** Laboratório de  
Diagnóstico  
por DNA

**Código da  
amostra:** \_\_\_\_\_

O (A) Senhor (a) está sendo convidado (a) a participar do projeto “Análise de ancestralidade da população brasileira através de metodologias de sequenciamento de segunda geração (*Next Generation Sequence*)”, cujo objetivo é avaliar a ancestralidade da população brasileira através das novas tecnologias de sequenciamento de DNA (NGS) e comparar estes resultados com aqueles que são gerados pelos métodos atualmente utilizados em nosso laboratório (Análise de fragmentos e/ou sequenciamento de Sanger). Pretendemos com este trabalho ampliar os conhecimentos sobre a composição ancestral de nossa população e desta forma contribuir para um melhor entendimento de como as frequências alélicas se apresentam em uma população miscigenada formada a partir da interação de três principais populações ancestrais (Européia, Africana e Ameríndia). Este tipo de estudo pode ser aplicado a várias áreas do conhecimento como a genética clínica e a genética forense além de poder ser utilizado para revelar aspectos evolutivos e históricos das populações.

### **Como irá ser efetuado o projeto**

Pretendemos selecionar 100 famílias compostas por pai, mãe e dois filhos, formando um total de 400 indivíduos.

Para realizar este estudo, precisamos de DNA que será obtido a partir de uma gota de sangue. Portanto, pedimos-lhes se podemos coletar uma gota do seu sangue usando uma lanceta esterilizada.

### **Efeitos secundários e riscos**

Apesar da coleta de sangue ser feita usando uma lanceta para extrair uma gota de sangue, poderá haver o mínimo de desconforto associado a recolha de sangue. Salientamos que todas as amostras de sangue serão recolhidas com material estéril.

### **Registro e uso da informação no projeto**

Os pesquisadores envolvidos no estudo irão processar os seus dados pessoais de forma estritamente confidencial. A informação não pode ser associada aos voluntários do estudo. Os

resultados serão mais tarde incluídos em publicações científicas. Todos os voluntários poderão ter acesso aos resultados dessa investigação. Se este for o caso, deverá ser indicada esta opção na declaração de consentimento informado. O material biológico será armazenado no Laboratório de Diagnóstico por DNA da UERJ (Universidade do Estado do Rio de Janeiro) e será usado apenas para fins científicos ligados a este projeto.

### **Desvinculação ao estudo**

Ressaltamos que a sua participação neste projeto é voluntária e pode, a qualquer momento, decidir desvincular-se ao estudo. Se assim o desejar, sua amostra de sangue e todos os dados obtidos até o momento serão descartados.

### **Declaração do participante**

Declaro que recebi informação oral e escrita sobre o projeto e tenho conhecimento dos objetivos, métodos e vantagens/desvantagens para aceitar participar. Apesar de não existir nenhum risco de saúde ou segurança associado a este estudo, poderá haver o mínimo de desconforto associado à coleta de sangue, que será feita usando uma lanceta para extrair uma gota de sangue. Salientamos que todas as amostras de sangue serão recolhidas com material estéril. Tenho conhecimento que a participação é voluntária e que posso pedir a minha exclusão do projeto a qualquer momento. Aceito a participação no projeto e o armazenamento da minha amostra biológica no Laboratório de Diagnóstico por DNA da UERJ (Universidade do Estado do Rio de Janeiro). A amostra será anonimizada e será atribuído um código para garantir a confidencialidade dos meus dados. A análise dos dados será efetuada utilizando os códigos individuais. O consentimento informado será guardado num armário fechado apenas acessível aos membros do projeto. A privacidade dos participantes será preservada e a informação não poderá ser associada a cada participante. Recebi uma cópia desta declaração e uma cópia do Termo de Consentimento Livre e Esclarecido meu uso pessoal.

Nome do participante: \_\_\_\_\_

Idade: \_\_\_\_\_

Data: \_\_\_\_\_ Assinatura: \_\_\_\_\_

Gostaria de ser informado (a) sobre os resultados gerais do projeto quando os dados forem publicados em revistas científicas? Sim  Não

Gostaria de ser informado sobre os resultados da análise com marcadores de ancestralidade?

Sim  Não

Se respondeu 'Sim' a alguma destas perguntas, por favor, indique o seu endereço de email para que possa ser contactado posteriormente:

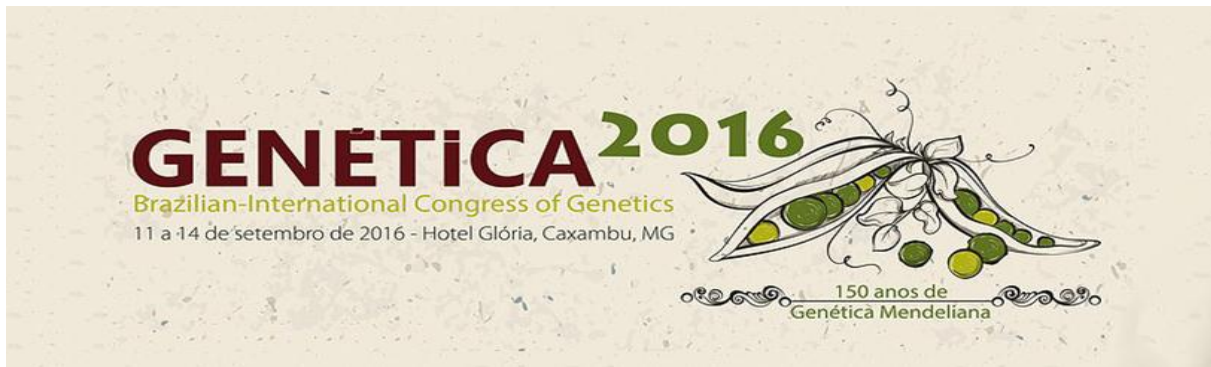
\_\_\_\_\_

**Declaração do pesquisador que prestou a informação:**

Declaro que o participante recebeu a informação oral e escrita necessária para formular uma decisão acerca da participação no projeto.

Nome: \_\_\_\_\_

Data: \_\_\_\_\_ Assinatura: \_\_\_\_\_

**ANEXO B - Pôster apresentado no 62º Congresso Brasileiro de Genética****Resumo traduzido:**

Existem diferentes tipos de polimorfismo no genoma humano que podem ser utilizados em estudos genéticos populacionais. Os Marcadores Informativos de Ancestralidade (AIMs) caracterizam-se por apresentar uma variação moderada entre indivíduos da mesma população e uma alta divergência nas frequências alélicas entre os principais grupos étnicos. Os AIMs são frequentemente usados para estudar a diversidade em populações humanas, especialmente nas miscigenadas. A análise da ascendência em populações miscigenadas é complexa porque a dinâmica da mistura deve ser interpretada num contexto histórico. Os AIMs podem ser utilizados para determinar a contribuição de cada grupo populacional ancestral em indivíduos e, portanto, na população total. Historicamente, sabe-se que a população brasileira é o resultado de cruzamentos entre três grandes grupos ancestrais: europeus, ameríndios e africanos. A variação nas distribuições de genótipos em populações ancestrais, juntamente com o padrão de mistura diferencial que está por trás de cada população, criou uma população brasileira caracterizada por diferenças regionais. Portanto, a constituição atual da população brasileira reflete o processo de colonização, associado a migrações de outros continentes e migrações internas subsequentes. Com este estudo, pretendemos descrever a diversidade em termos da ancestralidade do Rio de Janeiro e avaliar possíveis níveis de subestrutura populacional. Além disso, pretendemos determinar se há uma relação entre a ancestralidade dos filhos e o valor médio da ancestralidade dos pais, bem como entre os irmãos. Para tanto, utilizou-se, no presente estudo, um multiplex contendo 46 AIM - Indels para analisar as proporções de ancestralidade em indivíduos da população do Rio de Janeiro: (i) 667 amostras de indivíduos não relacionados; e (ii) 400 amostras correspondentes a 100 famílias constituídas por pai, mãe e dois filhos. Os resultados obtidos até agora estão de acordo com o que foi descrito anteriormente sobre a população do Rio de Janeiro. Os resultados mostraram

um perfil de ancestralidade com uma contribuição européia principal, seguida por um afluxo africano, e a ancestralidade nativa americana foi reduzida. Ao comparar o valor médio da ancestralidade dos pais com a ancestralidade dos filhos, as diferenças encontradas variaram entre 0 e 20%. Uma variação maior foi observada ao comparar os valores de ancestralidade observados entre irmãos, variando de 0,1% a 30%. As diferenças obtidas entre os membros da família apontam para a necessidade de marcadores adicionais para aumentar a exatidão das estimativas de ancestralidade em populações miscigenadas da América do Sul.

**Título do trabalho:** Ancestry estimates in individuals of the Rio de Janeiro population: evaluation of the results obtained based on 46 AIM – Indels.

**Almeida, APF<sup>1</sup>**; Simão, F<sup>1</sup>; Sardenberg, LC<sup>2</sup>; Jannuzzi, J<sup>1</sup>; Chaves, AC<sup>1</sup>; Motta, J<sup>2</sup>; Moraes, MEH<sup>3</sup>; Silva, DA<sup>1</sup>; Porto, LC<sup>2</sup>; Carvalho, EF<sup>1</sup>; Gusmão, L<sup>1</sup>

<sup>1</sup>Laboratório de Diagnóstico por DNA, UERJ, Rio de Janeiro, RJ.

<sup>2</sup>Laboratório de Histocompatibilidade e Criopreservação, UERJ, Rio de Janeiro, RJ.

<sup>3</sup>JRM-Investigação Imunológicas, Rio de Janeiro, RJ

## **Introduction**

The Brazilian population is considered one of the most heterogeneous in the world, as the result of admixture between three major ancestral groups: Amerindians, Europeans and Africans [1]. It is believed that the Amerindian population originated in Asia and arrived in America about 15,000 to 18,000 years ago. However, both these dates, as well as the original population in Asia and how this migration occurred, are still discussed [2]. The Portuguese arrived in Brazil in the late fifteenth century, but the effective Portuguese occupation of the coastline only occurred in 1534 [3]. Africans were brought to Brazil as slaves, initially to work in the sugar economy, but soon became the main labor force in virtually all sectors of society. During the sixteenth to mid-nineteenth centuries it is estimated that about four million Africans were brought to Brazil [3]. The country's recent history is also marked by migrations, of individuals from other countries in Europe and Asian, who primarily occupied the southeastern Brazil, in addition to internal migration. The analysis of admixed populations is complex, and ancestry informative markers (AIMs) can be used to determine the contribution of each ancestral population. In this work we used 46 AIM - Indels described by Pereira et al. (2012), which are characterized by presenting a moderate variation between individuals of the same population and a high divergence in allelic frequencies between the

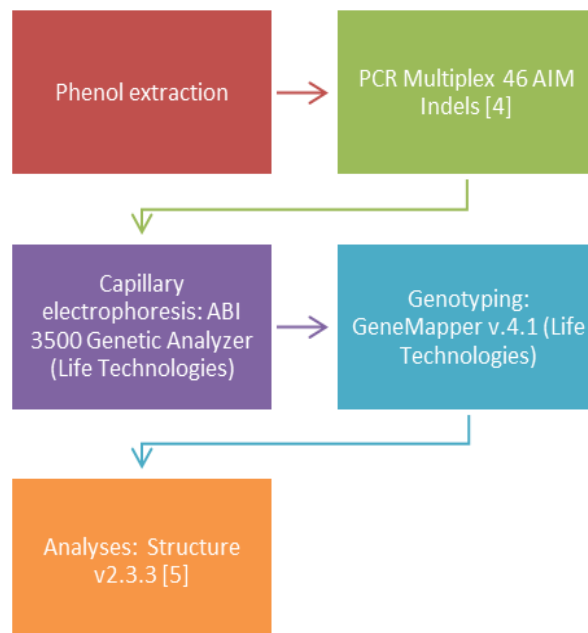
main ethnic groups. Understanding genetic variation at individual and population levels has been a subject of interest of several research groups around the world, since the study of genetic markers is an important tool, capable of revealing historical and evolutionary aspects of the population. The information obtained through this type of analysis can be of great value in several areas as clinical genetics, species conservation, forensics, among others.

### **Objective**

We intend to understand the regional diversity in terms of ancestry and evaluate possible levels of population substructure. In addition, we are evaluating whether there is a relationship between ancestral values observed between parents and children and between siblings. To this end, the present study used a multiplex containing 46 AIM - Indels to analyze the proportions of ancestry in unrelated individuals and families consisting of father, mother and two children, all from the population of Rio de Janeiro.

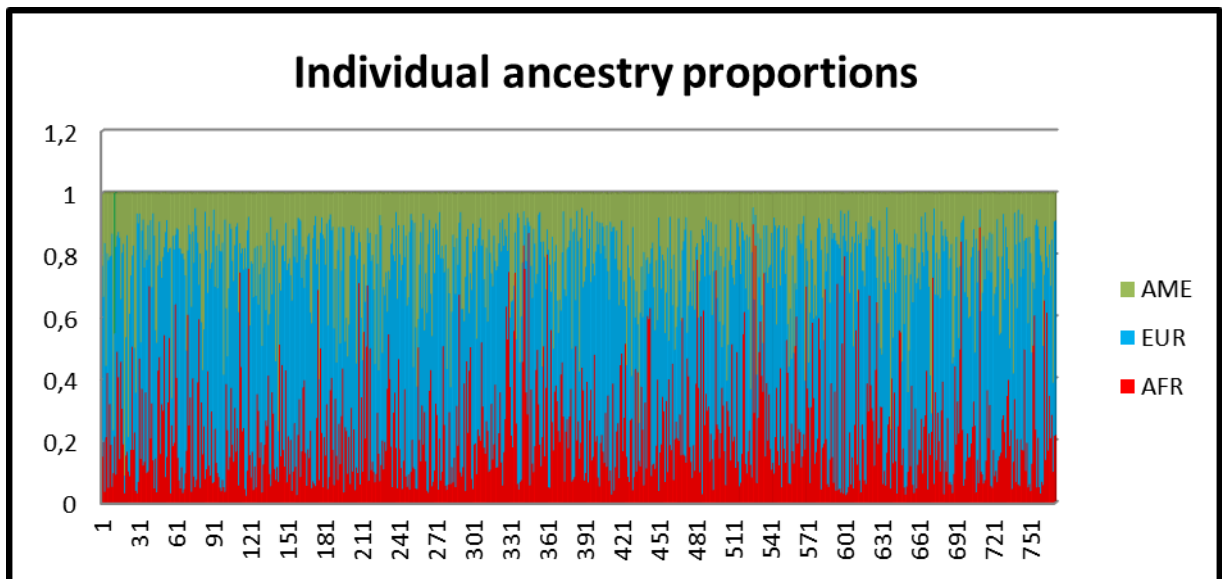
### **Material and Methods**

We analyzed 769 samples from unrelated individuals and 408 samples corresponding to 102 families consisting of father, mother and two children.

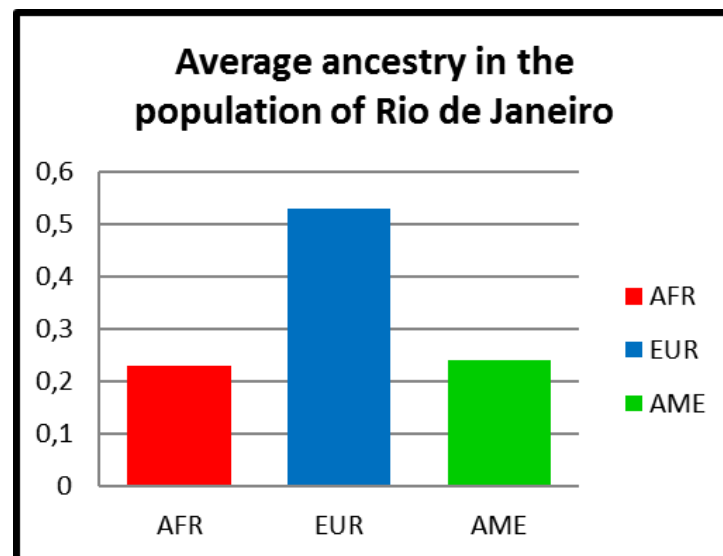


### **Results and Discussion**

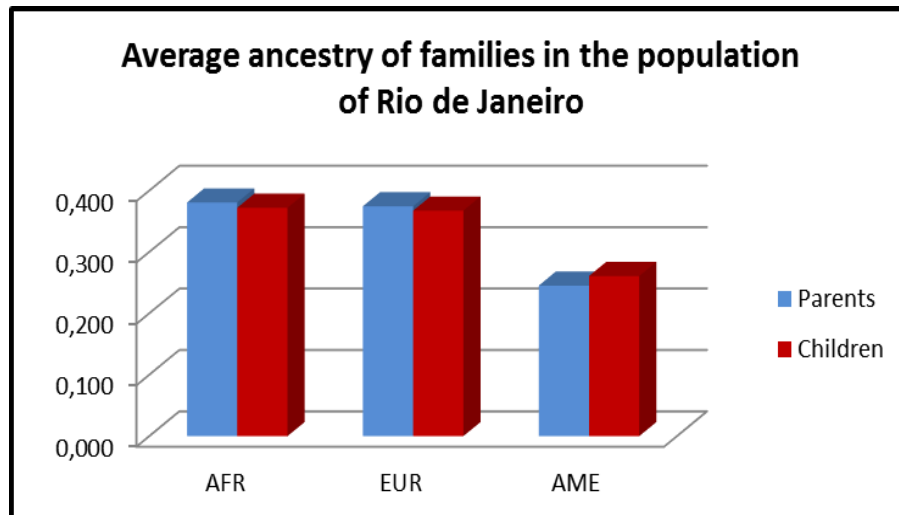
**Figure 1.** Average of individual ancestry proportions in the population of Rio de Janeiro using 46 AIM-Indels. Ancestry estimates were obtained using STRUCTURE, for the following options:  $k=3$ ; 100,000 burnin steps followed by 100,000 MCMC iterations; Admixture model (“Use population Information to test for migrants”); and allele frequencies were correlated and updated using only individuals with POPFLAG=1.



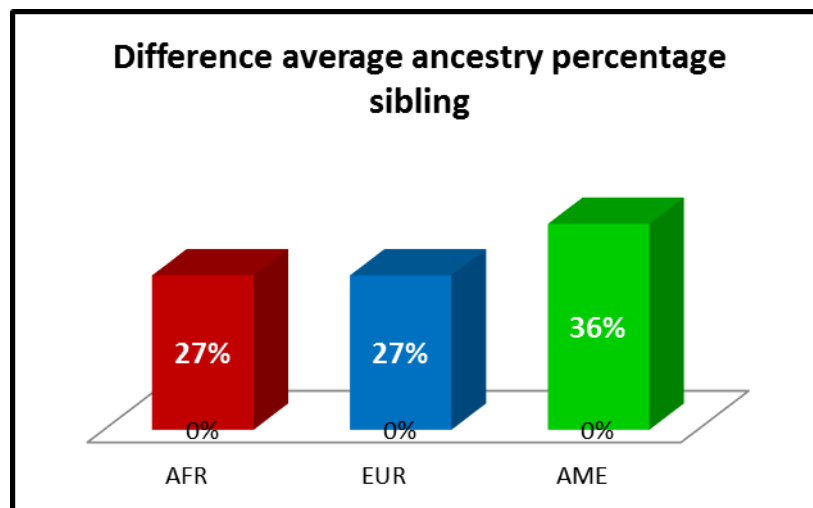
**Figure 2.** Average ancestry in the population of Rio de Janeiro using 46 AIM-Indels



**Figure 3.** Mean proportion of ancestry in families in Rio de Janeiro using 46 AIM-Indels.



**Figure 4.** Mean difference of ancestry in siblings in the population of Rio de Janeiro using 46 AIM-Indels.



## Conclusions

The results obtained so far are in line with what has been previously described concerning the population of Rio de Janeiro [6]. The results have shown an ancestry profile with a main European contribution, followed by an African influx and the Native American ancestry. When comparing the average value of ancestry in parents regarding the ancestry of the children, the differences found varied between 0 and 38%. A higher variation was observed when comparing the observed ancestry values between siblings, varying from 0% to 36%. The differences obtained among family members point to the need for additional markers to increase the accuracy of ancestry estimates in admixed populations from South America.

## Financial support

Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq e Laboratório de Diagnóstico por DNA da UERJ.

## References

- [1] Saloum, N. M. F, Pereira R, Vianna R, Rodolfo Beuttenmüller de Araújo A, Leite Góes Gitaí D, et al.; Revisiting the Genetic Ancestry of Brazilians Using Autosomal AIM-Indels. PLoS ONE 8(9): e75145. doi:10.1371/journal.pone.0075145 (2013).
- [2] Salzano, F. M.; Sans, M.; Interethnic admixture and the evolution of Latin American populations. Review Article. Genetics and Molecular Biology, 37, 1 (suppl), 151-170 (2014).
- [3] Instituto Brasileiro de Geografia e Estatística - IBGE; Brasil: 500 anos de povoamento / IBGE, Centro de Documentação e Disseminação de Informações. 232 p - Rio de Janeiro : IBGE, 2007.
- [4] PEREIRA, R.; et al.; Straightforward Inference of Ancestry and Admixture Proportions through Ancestry-Informative Insertion Deletion Multiplexing. PLoS One , v. 2 (1), p e29684; 2012.
- [5] PRITCHARD, J. K.; et al.; Inference of Population Structure Using Multilocus Genotype Data; v. 155; p. 945-59, 2000.
- [6] Manta, F. S., et al.; Analysis of genetic ancestry in the admixed Brazilian population from Rio de Janeiro using 46 autosomal ancestry-informative indel markers. Ann Hum Biol, 2013 JAN; 40(1):94-8. doi 10.3109/03014460.2012.742138. Epub Nov 15, 2012.

**ANEXO C** - Trabalho que será apresentado no 27º Congresso da Sociedade Internacional de Genética Forense



**Resumo traduzido:**

As populações brasileiras são altamente miscigenadas, com contribuições nativa americana, europeia e africana que variam ao longo do país. Neste estudo, 1042 amostras de indivíduos não relacionados do estado do Rio de Janeiro foram genotipadas para 46 AIM-Indels. Os resultados mostraram uma prevalência da ancestralidade europeia (63%), seguida de uma contribuição africana (25%), e a ancestralidade nativa americana foi menos representada (12%). Esta amostra foi dividida em subconjuntos, incluindo voluntários de trabalhos de pesquisa (2014-2016) e casos de paternidade (2002-2004 e 2010-2016). A ancestralidade média foi a mesma para amostras de voluntários e casos de paternidade no período 2002-2004. No entanto, uma redução da ancestralidade europeia e um aumento da ancestralidade africana foi observada no grupo das paternidades de 2010 em diante. A ancestralidade nativa americana apresentou baixa variação entre os grupos de voluntários e a paternidade, bem como ao longo do período avaliado. Comparando o local de residência dos indivíduos em cada subamostra, observaram-se diferenças na representação de cada município no estado do Rio de Janeiro, e maiores proporções de ancestralidade africana foram observadas em amostras de fora da região metropolitana do Rio de Janeiro, o que é o caso das amostras dos casos de paternidade. Em conclusão, nossos resultados mostram que os valores de ancestralidade obtidos para as amostras globais não representam bem o perfil das amostras de referência dos casos de paternidade e enfatiza a importância das estratégias de amostragem em populações miscigenadas. Essas diferenças devem ser consideradas ao realizar investigações de paternidade envolvendo indivíduos provenientes de diferentes regiões do

Estado do Rio de Janeiro, usando bases de dados regionais ou aplicando uma correção para a subestrutura populacional.

**Título do trabalho:** Contrasting admixture estimates in Rio de Janeiro obtained by different sampling strategies.

**Ana Paula F. Almeida**<sup>1</sup>, Filipa Simão<sup>1</sup>, Juliana G. Aquino<sup>1</sup>, Elizeu F. Carvalho<sup>1</sup>, Leonor Gusmão<sup>1,2</sup>

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

<sup>2</sup> Institute of Pathology and Molecular Immunology from University of Porto (IPATIMUP) / Instituto de Investigação e Inovação em Saúde (I3S), Porto, Portugal

### **Abstract**

Brazilian populations are highly admixed, with Native American, European and African contributions that vary along the country. In this study, 1042 samples of non-related individuals from Rio de Janeiro State were genotyped for 46 AIM-Indels. The results showed a prevalence of the European ancestry (63%), followed by an African contribution (25%), and Native American background was less represented (12%). This sample was divided in subsets including volunteers from research works (2014-2016) and selected from paternity casework (2002-2004 and 2010-2016). The average ancestry was the same for samples from volunteers and paternity casework in 2002-2004. However, lower European and higher African ancestries were observed for paternity groups from 2010 onwards. Native American ancestry presented low variation between the groups of volunteers and paternity, as well as along the evaluated period. Comparing the residence place of the individuals in each subsample, differences were observed in the representation of each municipalities in Rio de Janeiro State, and higher proportions of African ancestry were observed in samples from outside Rio de Janeiro metropolitan region, which is the case of the paternity subsamples. In conclusion, our results show that the ancestry values obtained for the global samples did not represent well the profile of the reference samples from paternity casework and emphasizes the importance of sampling strategies in admixed populations. These differences should be considered when performing paternity investigations involving individuals coming from different regions of Rio de Janeiro State, by using regional databases or by applying a correction for substructure.

**Keywords:** Ancestry informative markers, Population Genetics, Brazil, Admixed populations

## 1. Introduction

Brazilian populations are highly diverse, due to an intense admixture that took place during the last five centuries, between Native Americans, Europeans and Africans [3]. Archeologic and genetic data showed that Native American populations descended from Asian groups that entered the continent by the land bridge of Beringia, 15–18,000 years ago [2]. The Portuguese arrived in Brazil in 1500 but the effective occupation of the Brazilian coast only took place after 1534. The Africans were brought to Brazil as slaves during the 16th to 19th centuries. The recent history of the country is also marked by immigrations of individuals from other countries in Europe and Asia, who occupied mainly the South and Southeast regions of Brazil, in addition to internal migration. Due to the colonization process, and the recent European and internal migrations, Brazil shows varying regional patterns of admixture [1][9]. The ancestry of the Rio de Janeiro population was previously investigated by Saloum *et al.* [1] using 46 AIM indels described by Pereira *et al.* [4], in a sample of 280 unrelated individuals born in the metropolitan region. In the present study 762 samples from Rio de Janeiro were genotyped for the same set of AIM indels in order to update the available database. At the same time, we aim to investigate if differences existed between the samples collected for research studies and those from paternity casework.

## 2. Materials and methods

### 2.1 DNA samples

A total sample of 1042 non-related individuals from Rio de Janeiro was included in this study. These samples have been collected with two different purposes. Three subsets include samples collected from volunteers (mainly employees and students from the State University of Rio de Janeiro - UERJ) to participate in research works, during the years 2014, 2015 and 2016. The remaining samples were selected from paternity casework in the DNA Diagnostic Laboratory from the State University of Rio de Janeiro (LDD-UERJ), for the years 2002-2004 and 2010-2016.

This study, together with the written informed consent signed by sample donors, was approved by the Ethics Committee of the State University of Rio de Janeiro (CAAE: 0067.0.228.000-09).

DNA was extracted from bloodstains using a standard phenol–chloroform protocol.

### 2.2 Genetic markers and genotyping

The samples were genotyped for 46 AIM-InDel markers, by PCR multiplex and capillary electrophoresis, as previously described in Pereira et al. [4]. Separation of the PCR amplified products and detection were performed on a 3500 Genetic Analyser (Applied Biosystems), using POP-7™ polymer (Applied Biosystems). The genotypes were assigned using the software GeneMapper v4.1 (Applied Biosystems).

### *2.3 Statistical data analyses*

The apportionment of genetic ancestral contributions was estimated using the software STRUCTURE v2.3.3 [5,6]. To estimate the ancestral membership proportions, a supervised analysis was performed using prior information on the geographic origin of the reference samples from Africa, Europe and Native America (data from Pereira et al. [4]). The STRUCTURE runs comprised three replicates of 100,000 burning steps followed by 100,000 Markov Chain Monte Carlo (MCMC) iterations. A tri-hybrid contribution from Native Americans, Europeans and Africans was assumed ( $K=3$ ). The “Admixture model” was used, and allele frequencies were correlated and updated using only individuals with POPFLAG=1 (in this case, the HGDP-CEPH samples used as reference).

## **3. Results and discussion**

The results obtained from the analysis of the 46 AIM-Indels showed that, when increasing the size of the sample from Rio the Janeiro from 280 to 1042 individuals, there is prevalence of the European ancestry (63%), followed by an African contribution of 25%, and the Native American background was less represented (12%). The increment on sample size, did not revealed significant differences to the previously estimates [5].

To see if the ancestry estimates for Rio de Janeiro population can vary with sampling collection criteria, samples included in this study were subdivided in eleven different groups. The first three groups included employees and students from UERJ that participated as volunteers in research projects during the years 2014, 2015 and 2016, respectively. The remaining groups include individuals from paternity casework, corresponding to different years of samples collection. The Native American, European and African contributions to each group are represented in Fig. 1.

The average ancestry values calculated for the volunteers participating in research projects were similar to those observed in the group from paternity casework in 2002-2004. However, lower European and higher African ancestries were observed for the paternity groups from 2010 from onwards (Fig. 1). The average values of Native American ancestry presented low

variation between the groups of volunteers and paternity, as well as along the evaluated period, being around 18%.

Comparing the residence place of the individuals in each subsample, we can see that differences exist in the representation of each municipalities in Rio de Janeiro State. Indeed, higher proportions of African ancestry are observed when samples come from outside Rio de Janeiro metropolitan region, which is the case of the paternity subsamples.

#### **4. Conclusion**

In the present study, we have enlarged the population database of Rio de Janeiro State. Although the ancestry values did not change significantly from those previously published by Saloum *et al.* [1], it was found that the ancestry values obtained for the global samples did not represent well the profile of the reference sample from our paternity casework.

The differences observed between the sample from volunteers and paternities emphasize the importance of sampling strategies in admixed populations. Moreover, these differences should also be considered when performing paternity investigations involving individuals coming from different regions of Rio de Janeiro State, by using regional databases or by applying a correction for substructure.

#### **Financial support**

Financial support was granted by Conselho Nacional de Desenvolvimento Científico e Tecnológico (*CNPq*) and DNA Program - State University and Justice Court of Rio de Janeiro, Brazil. IPATIMUP integrates the i3S Research Unit, which is partially supported by FCT, the Portuguese Foundation for Science and Technology.

#### **Conflict of interest**

None.

#### **References**

[1] IBGE - Instituto Brasileiro de Geografia e Estatística, Brasil: 500 anos de povoamento / IBGE, Centro de Documentação e Disseminação de Informações. 232 p - Rio de Janeiro: IBGE, 2007. ISBN 978-85-240-3940-9.

[2] D.H. O'Rourke, J.A. Raff, Review - The Human Genetic History of the Americas: The

Final Frontier, *Curr. Biol.* 20 (2010) R202–R207.

[3] N.M.F Saloum, R. Pereira, R. Vianna, et al., Revisiting the Genetic Ancestry of Brazilians Using Autosomal AIM-Indels. *PLoS One* 8 (2013) e75145.

[4] F.M. Salzano, M. Sans, Interethnic admixture and the evolution of Latin American populations, *Genet. Mol. Biol.* 37 (2014) 151-170.

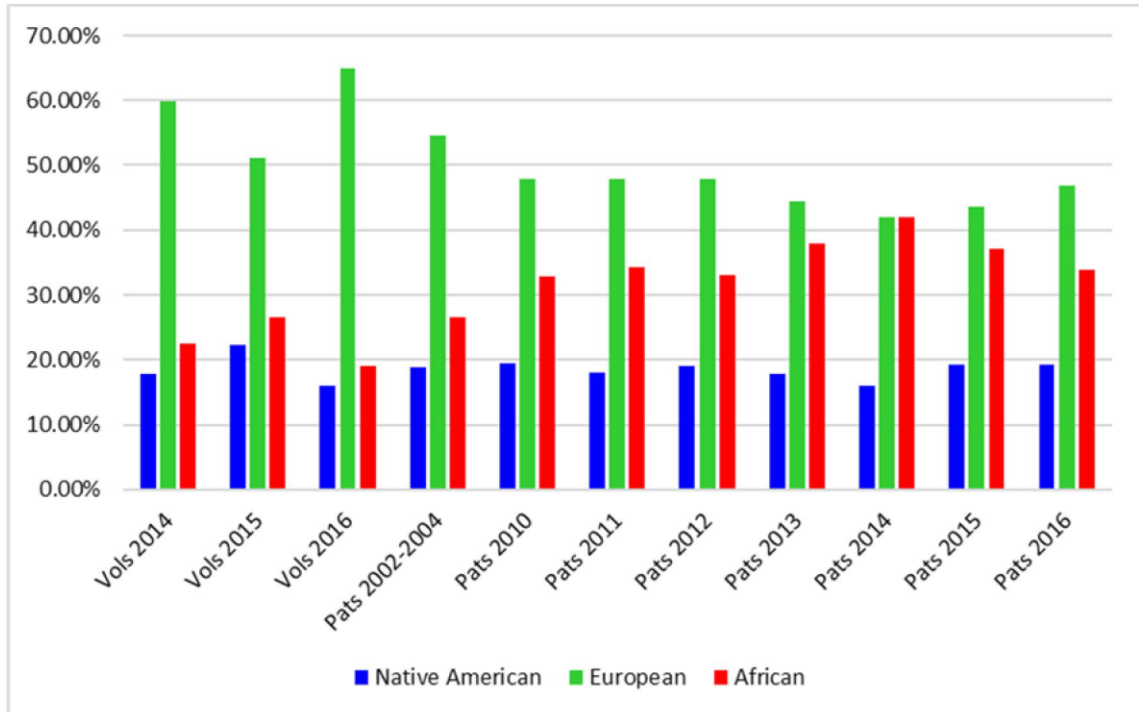
[5] F.S. Manta, R. Pereira, A. Caiafa, et al., Analysis of genetic ancestry in the admixed Brazilian population from Rio de Janeiro using 46 autosomal ancestry-informative indel markers, *Ann. Hum. Biol.* 40 (2013) 94-98.

[6] R. Pereira, C. Phillips, N. Pinto, et al., Straightforward inference of ancestry and admixture proportions through ancestry-informative insertion deletion multiplexing. *PloS One.* 7 (2012) e29684.

[7] J.K. Pritchard, M. Stephens, P. Donnelly, Inference of Population Structure Using Multilocus Genotype Data. *Genetics* 155 (2000) 945-959.

[8] D. Falush, M. Stephens, J.K. Pritchard, Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164 (2003) 1567-1587.

**Figure 1.** Comparison of the average proportion of Native American, European and African ancestry, among non-related individuals within eleven sample sets (based on sampling criteria).



Note: All samples belong to Rio de Janeiro State population, and the results were obtained based on 46 AIM-Indels. Vols: Volunteers, Pats: Paternity cases.

**ANEXO D** - Tabela contendo os 165 AIM-SNPs presentes no painel de ancestralidade Ion Torrent e suas respectivas localizações no genoma. Adaptado de Santangelo, 2016

<b>Código rs</b>	<b>Painel</b>	<b>Crom.</b>	<b>Posição</b>
rs1040404	Seldin	1	168159890
rs12130799	Seldin	1	55663372
rs1325502	Seldin	1	42360270
rs1407434	Seldin	1	186149032
rs2814778	Kidd	1	159174683
rs2986742	Seldin	1	6550376
rs3118378	Seldin	1	68849687
rs316873	Seldin	1	242342504
rs3737576	Both	1	101709563
rs4908343	Seldin	1	27931698
rs4951629	Seldin	1	212786883
rs647325	Seldin	1	18170886
rs6541030	Seldin	1	12608178
rs7554936	Both	1	151122489
rs10496971	Seldin	2	145769943
rs10497191	Kidd	2	158667217
rs13400937	Seldin	2	79864923
rs1569175	Seldin	2	201021954
rs1834619	Kidd	2	17901485
rs1876482	Kidd	2	17362568
rs260690	Both	2	109579738
rs2627037	Seldin	2	179606538
rs3827760	Kidd	2	109513601
rs4666200	Seldin	2	29538411
rs4670767	Seldin	2	37941396
rs6754311	Kidd	2	136707982
rs7421394	Seldin	2	14756349
rs798443	Both	2	7968275
rs12498138	Kidd	3	121459589
rs12629908	Seldin	3	120522716
rs1513181	Seldin	3	188574996
rs2030763	Seldin	3	179964727
rs4955316	Seldin	3	30415612
rs6548616	Seldin	3	79399575
rs734873	Seldin	3	147750355
rs9809104	Seldin	3	39146429
rs9845457	Seldin	3	135914476
rs10007810	Seldin	4	41554364
rs1229984	Kidd	4	100239319
rs1369093	Seldin	4	73245191
rs2702414	Seldin	4	179399523
rs3811801	Kidd	4	100244319
rs385194	Seldin	4	85309078
rs4833103	Kidd	4	38815502
rs7657799	Both	4	105375423
rs9291090	Seldin	4	5390637
rs12657828	Seldin	5	79085726
rs1500127	Seldin	5	165739982
rs16891982	Kidd	5	33951693
rs316598	Seldin	5	2364626
rs37369	Seldin	5	35037115

rs6422347	Seldin	5	177863083
rs6451722	Seldin	5	43711378
<b>Código rs</b>	<b>Painel</b>	<b>Crom.</b>	<b>Posição</b>
rs6556352	Seldin	5	155471714
rs7722456	Kidd	5	170202984
rs870347	Both	5	6845035
rs1040045	Seldin	6	4747159
rs1871428	Seldin	6	168665760
rs192655	Seldin	6	90518278
rs2504853	Seldin	6	12535111
rs3823159	Kidd	6	136482727
rs4458655	Both	6	163221792
rs4463276	Seldin	6	145055331
rs7745461	Seldin	6	21911616
rs10236187	Seldin	7	139447377
rs10954737	Seldin	7	83533047
rs2330442	Seldin	7	42380071
rs32314	Seldin	7	32179124
rs4717865	Seldin	7	73454199
rs6464211	Seldin	7	151873853
rs705308	Seldin	7	97695363
rs731257	Seldin	7	12669251
rs7803075	Seldin	7	130742066
rs917115	Kidd	7	28172586
rs10108270	Seldin	8	4190793
rs12544346	Seldin	8	86424616
rs1462906	Kidd	8	31896592
rs1471939	Seldin	8	28941305
rs1871534	Kidd	8	145639681
rs2001907	Seldin	8	140241181
rs2196051	Kidd	8	122124302
rs3943253	Seldin	8	13359500
rs6990312	Kidd	8	110602317
rs7844723	Seldin	8	122908503
rs10511828	Seldin	9	28628500
rs10513300	Seldin	9	120130206
rs2073821	Seldin	9	135933122
rs2306040	Seldin	9	93641199
rs3793451	Seldin	9	71659280
rs3814134	Kidd	9	127267689
rs3793791	Seldin	10	50841704
rs4746136	Seldin	10	75300994
rs4880436	Seldin	10	134650103
rs4918664	Kidd	10	94921065
rs4918842	Seldin	10	115316812
rs1079597	Kidd	11	113296286
rs10839880	Seldin	11	7850316
rs11227699	Seldin	11	66898492
rs174570	Kidd	11	61597212
rs1837606	Seldin	11	15838137
rs2946788	Seldin	11	24010530
rs948028	Seldin	11	120644447

rs1503767	Seldin	12	118889488
rs1513056	Seldin	12	17407792
rs2070586	Seldin	12	109277720
rs214678	Seldin	12	47676950
<b>Código rs</b>	<b>Painel</b>	<b>Crom.</b>	<b>Posição</b>
rs2238151	Kidd	12	112211833
rs2416791	Seldin	12	11701488
rs671	Kidd	12	112241766
rs772262	Seldin	12	56163734
rs1572018	Kidd	13	41715282
rs2166624	Kidd	13	42579985
rs7326934	Kidd	13	49070512
rs7997709	Both	13	34847737
rs9319336	Seldin	13	27624356
rs9522149	Both	13	111827167
rs9530435	Seldin	13	75993887
rs1760921	Seldin	14	20818131
rs1950993	Seldin	14	58238687
rs200354	Both	14	99375321
rs2357442	Seldin	14	52607967
rs3784230	Seldin	14	105679055
rs8021730	Seldin	14	67886781
rs946918	Seldin	14	83472868
rs12439433	Both	15	36220035
rs12913832	Kidd	15	28365618
rs1426654	Kidd	15	48426484
rs1800414	Kidd	15	28197037
rs2899826	Seldin	15	74734500
rs735480	Kidd	15	45152371
rs8035124	Seldin	15	92105708
rs2966849	Seldin	16	85183682
rs459920	Kidd	16	89730827
rs4781011	Seldin	16	10975311

rs4984913	Seldin	16	740466
rs818386	Seldin	16	65406708
<b>Código rs</b>	<b>Painel</b>	<b>Crom.</b>	<b>Posição</b>
rs10512572	Seldin	17	69512099
rs11652805	Both	17	62987151
rs17642714	Kidd	17	48726132
rs1879488	Seldin	17	1401613
rs2033111	Seldin	17	53788280
rs2125345	Seldin	17	73782191
rs2593595	Kidd	17	41056245
rs4411548	Kidd	17	40658533
rs4471745	Kidd	17	53568884
rs2042762	Kidd	18	35277622
rs3916235	Kidd	18	67578931
rs4798812	Seldin	18	9420504
rs4891825	Both	18	67867663
rs7226659	Kidd	18	40488279
rs7238445	Seldin	18	49781544
rs874299	Seldin	18	75056284
rs881728	Seldin	18	59333108
rs2532060	Seldin	19	55614923
rs3745099	Seldin	19	52901905
rs7251928	Kidd	19	4077096
rs8113143	Seldin	19	33652247
rs310644	Kidd	20	62159504
rs3907047	Seldin	20	54000914
rs6104567	Seldin	20	10195433
rs2835370	Seldin	21	37885625
rs1296819	Seldin	22	18076546
rs2024566	Kidd	22	41697338
rs4821004	Seldin	22	32366359
rs5768007	Seldin	22	48207872