



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

Filipa Isabel Figueira Simão

**Ancestralidade materna em populações da América do Sul: contrastando  
dados genéticos e históricos**

Rio de Janeiro

2019

Filipa Isabel Figueira Simão

**Ancestralidade materna em populações da América do Sul: contrastando dados  
genéticos e históricos**

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

Orientadora: Prof.<sup>a</sup> Dra. Maria Leonor Rodrigues de Sousa Botelho de Gusmão

Coorientador: Prof. Dr. Elizeu Fagundes de Carvalho

Rio de Janeiro

2019

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

S588 Simão, Filipa Isabel Figura.  
Ancestralidade materna em populações da América do Sul: contrastando dados genéticos e históricos. / Filipa Isabel Figura Simão. – 2019.  
129 f.

Orientadora: Maria Leonor Rodrigues de Sousa Botelho de Gusmão.  
Coorientador: Elizeu Fagundes de Carvalho.

Tese (Doutorado) - 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. População – Aspectos genéticos - Teses. I. Gusmão, Maria Leonor Rodrigues de Sousa Botelho de. II. Carvalho, Elizeu Fagundes. 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 tese, desde que citada a fonte.

---

Assinatura

---

Data

Filipa Isabel Figueira Simão

**Ancestralidade materna em populações da América do Sul: contrastando dados genéticos e históricos**

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

Aprovada em 15 de janeiro de 2019.

Coorientador: Prof. Dr. Elizeu Fagundes de Carvalho

Instituto de Biologia Roberto Alcântara Gomes – UERJ

Banca Examinadora:

---

Prof.<sup>a</sup> Dra. Maria Leonor Rodrigues de Sousa Botelho de Gusmão  
(Orientadora)

Instituto de Biologia Roberto Alcântara Gomes - UERJ

---

Prof.<sup>a</sup> Dra. Tatiana Simão

Instituto de Biologia Roberto Alcântara Gomes – UERJ

---

Prof. Dr. Rodrigo Moura-Neto

Universidade Federal do Rio de Janeiro

---

Prof. Dr. Rodrigo Grazinoli Garrido

Universidade Federal do Rio de Janeiro

Rio de Janeiro

2019

## **AGRADECIMENTOS**

À minha orientadora Leonor Gusmão por ter aceite a orientação da presente Dissertação, pela forma generosa com que sempre me ajudou e pelo rigoroso nível científico exigido. Grata pela liberdade que me deu ao longo destes quatro anos, crucial para que este trabalho contribuísse também para o meu desenvolvimento pessoal.

Ao meu co-orientador, Professor Elizeu Fagundes, por permitir a realização do presente projeto nas instalações do Laboratório de Diagnóstico por DNA, garantindo sempre as condições necessárias à sua execução. Obrigada pelo apoio prestado durante a realização deste trabalho.

Ao Instituto Nacional de Medicina Legal de Innsbruck, Áustria, representado pelo professor Walther Parson, por me receber durante o período de doutorado Sanduíche, permitindo a aquisição de conhecimentos que, de outra forma, não teria sido possível.

Aos amigos e colegas do Laboratório de Diagnóstico por DNA por tornarem as horas de trabalho mais leves.

Aos amigos e colegas do Instituto Nacional de Medicina Legal pela recepção e pela rápida e fácil integração.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela concessão de bolsa de estudos durante parte do curso de doutorado.

À Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ), que no ano de 2016, designou verba para a realização de doutorado Sanduíche, por 9 meses.

Aos meus pais e irmão, pelo apoio incondicional, mesmo nas escolhas mais difíceis.

## RESUMO

SIMÃO, Filipa Isabel Figura. **Ancestralidade materna em populações da América do Sul:** contrastando dados genéticos e históricos. 2019. 129f. Tese (Doutorado em Biociências) – Instituto de Biologia Roberto Alcântara Gomes, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2019

A heterogeneidade cultural e étnica presente em populações da América do Sul é resultado de extensos processos de miscigenação, ocorridos ao longo dos últimos séculos, entre indivíduos nativo-americanos, colonizadores europeus (maioritariamente portugueses e espanhóis) e escravos africanos. Por causa do padrão de herança uniparental, os polimorfismos localizados no genoma mitocondrial contam histórias independentes sobre os ancestrais maternos das populações, funcionando como um registro molecular da história genealógica e das migrações das mulheres ao longo de várias gerações. O presente trabalho teve como objetivo analisar a composição genética materna de três grupos populacionais da América do Sul: populações miscigenadas do Rio de Janeiro (Brasil) e do Paraguai e população nativa do Peru, pertencente ao grupo Ashaninka (família linguística Arawak). Pretendeu-se investigar de que forma os diferentes eventos históricos contribuíram para a composição genética atual destas populações e suas diferenças inter- e intra-populacionais. Assim, a região controle do mtDNA de 675 indivíduos residentes na América do Sul foi sequenciada através da metodologia de Sanger e, com base nos polimorfismos identificados, foi determinada a origem continental das linhagens maternas obtidas. Adicionalmente 105 amostras de indivíduos residentes no departamento de Alto Paraná, Paraguai, e 20 amostras de indivíduos Ashaninka foram estudadas por metodologia de sequenciamento massivo paralelo, com obtenção de haplótipos da molécula completa de mtDNA. A população do Rio de Janeiro apresentou uma elevada proporção de haplogrupos africanos (42 % dos haplótipos pertencem ao macrohaplogrupo L), a maioria com possível origem geográfica na região Sudoeste de África. Para os haplogrupos euroasiáticos e nativo-americanos encontrados no Rio de Janeiro (representando 32% e 26% das amostras, respectivamente), não foi possível determinar uma origem geográfica clara ou uma afiliação linguística. A comparação das proporções de ancestralidade, obtidas para mtDNA e marcadores autossômicos, apontou para a presença de subestrutura populacional no Rio de Janeiro. Cerca de 88 % dos haplótipos observados na população do Paraguai pertence a haplogrupos nativo-americanos. A comparação entre a origem continental das linhagens maternas do Paraguai e os resultados obtidos para marcadores autossômicos informativos de ancestralidade, em 39 amostras, suportaram a existência de processos de mistura entre paraguaios e indivíduos de outras populações, provavelmente já miscigenadas, onde os homens terão contribuído com elevada ancestralidade europeia. Além do mais, a comparação dos resultados dos mitogenomas de indivíduos residentes no Alto Paraná com outras populações da América do Sul (dados disponíveis na literatura) indicou a influência, não só de indígenas Tupi-Guarani, como também de outros grupos linguísticos. Na população Ashaninka, residente na floresta tropical do Peru, não foi observada qualquer contribuição não nativa na sua composição genética materna, confirmando o isolamento geográfico deste grupo indígena. Os resultados do sequenciamento da molécula completa de 20 amostras Ashaninka, pertencentes ao haplogrupo B2+16051G+16129A+152C, para além de apoiarem as teorias sobre a origem da família linguística Arawak, revelaram uma possível rota de dispersão no sentido sul do subcontinente. Com os resultados obtidos no presente trabalho, é evidente a contribuição diferencial dos componentes maternos nas três populações estudadas, com uma associação aos eventos históricos particulares de cada uma delas.

Palavras-chave: mtDNA. Ancestralidade. América do Sul.

## ABSTRACT

SIMÃO, Filipa Isabel Figueira. **Maternal ancestry in South American populations:** contrasting genetic and historical data. 2019. 129f. Tese (Doutorado em Biociências) – Instituto de Biologia Roberto Alcântara Gomes, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2019.

The cultural and ethnic heterogeneity present in South American populations is the result of extensive admixture processes, over the past centuries, among native Americans, European settlers (mainly Portuguese and Spanish) and African slaves. Because of its uniparental inheritance, polymorphisms located in the mitochondrial genome tell independent histories about maternal ancestors of the populations, working as a molecular record of the genealogical history and women migrations over several generations. The aim of the present study was to analyse the maternal genetic composition of three population groups of South America: admixed populations of Rio de Janeiro (Brazil) and Paraguay and native population of Peru, belonging to the Ashaninka group (Arawak language family). It was intended to investigate how the different historical events contributed to the current genetic composition of these populations as well as to evaluate their inter- and intra-population differences. Thus, the mtDNA control region of 675 individuals from South America was sequenced using the Sanger methodology. The continental origin of the maternal lineages was determined, based on the polymorphisms identified. In addition, 105 samples from individuals living in the department of Alto Paraná, Paraguay, and 20 samples from Ashaninka individuals were studied by massively parallel sequencing methodology, with the analysis of complete mtDNA molecule haplotypes. The Rio de Janeiro population presented a high proportion of African haplogroups (42% of the haplotypes belong to the macrohaplogroup L), the majority with a possible geographic origin in the Southwest region of Africa. For the Eurasian and Native American haplogroups found in Rio de Janeiro (representing 32% and 26% of the samples, respectively), it was not possible to determine a clear geographic origin or linguistic affiliation. The comparison of ancestry proportions, obtained for mtDNA and autosomal markers, pointed to the presence of population substructure in Rio de Janeiro. About 88% of the haplotypes observed in the Paraguayan population belong to Native American haplogroups. The comparison between the continental origin of the Paraguayan maternal lineages and the results obtained for autosomal ancestry markers, in 39 samples, supported the existence of admixture processes between Paraguayans and individuals from other populations, probably already admixed, where the men will have contributed with a high European ancestry. Moreover, the results obtained from the comparison of Alto Paraná mitogenomes with data available in the literature for other South American populations, indicated the influence of Tupi-Guarani natives, and of other linguistic groups. For the Ashaninka population, living in the Peruvian rainforest, no non-native maternal ancestry was observed, confirming the geographic isolation of this native group. The sequencing results of mitogenome for 20 Ashaninka samples, belonging to haplogroup B2 + 16051G + 16129A + 152C, not only support the theories about the origin of the Arawak linguistic family, but also revealed a possible south dispersal route in the subcontinent. With the results obtained in the present study, it is evident the differential contribution of the maternal components in the three populations studied, with an association with their particular historical events.

Keywords: mtDNA. Ancestry. South America

## LISTA DE FIGURAS

Figura 1 –	Estrutura da mitocôndria .....	15
Figura 2 –	Representação da estrutura circular do genoma mitocondrial.....	16
Figura 3 –	Representação da herança uniparental materna associada ao genoma mitocondrial.....	19
Figura 4 –	Esquema representativo do método de sequenciamento de Sanger.....	22
Figura 5 –	Organização dos haplogrupos e suas ramificações.....	26
Figura 6 –	Distribuição dos haplogrupos no Mundo.....	27
Figura 7 –	Distribuição dos haplogrupos de mtDNA nas 5 regiões de África.....	29
Figura 8 –	Exemplos de sequências de mtDNA com a) heteroplasmia de comprimento e b) heteroplasmia de posição.....	34
Figura 9 –	Entrada e dispersão do Homem moderno no continente americano.....	36
Figura 10 –	Mapa do Brasil, destacando as 5 regiões geográficas.....	37
Figura 11 –	Principais portos de tráfico de escravos em África.....	39
Figura 12 –	Distribuição dos principais grupos Arawak (cinza escuro), à chegada dos colonizadores europeus.....	43
Figura 13 –	Locais de coleta das amostras.....	48
Figura 14 –	Mapa dos departamentos do Paraguai.....	49
Figura 15 –	Etapas do sequenciamento massivo paralelo no aparelho Ion S5.....	53

## **LISTA DE TABELAS**

Tabela 1 – Regras de anotação de polimorfismos na molécula de mtDNA.....	25
Tabela 2 – Códigos IUPAC.....	34

## LISTA DE ABREVIATURAS E SIGLAS

A	Adenina
AIMs	Marcadores informativos de ancestralidade ( <i>Ancestry Informative Markers</i> )
ATP	Adenosina Trifosfato ( <i>Adenosine TriPhosphate</i> )
C	Citosina
CRS	Sequência de referência de Cambridge ( <i>Cambridge Reference Sequence</i> )
ddNTPs	Dideoxinucleóideos ( <i>Dideoxynucleotides</i> )
Del	Deleção
del	Deleção
D-loop	<i>Displacement loop</i>
DNA	Ácido desoxiribonucleico ( <i>Deoxyribonucleic Acid</i> )
EDNAP	<i>The European DNA Profiling group</i>
EMPOP	<i>EDNAP forensic MtDNA POPulation database</i>
G	Guanina
HVSI	Segmento Hipervariável I ( <i>HiperVariable Segment I</i> )
HVSII	Segmento Hipervariável II ( <i>HiperVariable Segment II</i> )
HVSIII	Segmento Hipervariável III ( <i>HiperVariable Segment III</i> )
ISFG	Sociedade Internacional de Genética Forense ( <i>International Society for Forensic Genetics</i> )
LD	Laboratório de Diagnóstico por DNA
MPS	Sequenciamento paralelo massivo ( <i>Massively Parallel Sequencing</i> )
mtDNA	DNA mitocondrial
pbs	Pares de bases
PCR	Reação em cadeia da polimerase ( <i>Polymerase Chain Reaction</i> )
rCRS	Sequência de Referência de Cambridge revista ( <i>revised Cambridge Reference Sequence</i> )
RFLPs	Polimorfismos de comprimento de fragmento de restrição ( <i>Restriction Fragment Length Polymorphism</i> )
RNA	ácido ribonucleico ( <i>Ribonucleic Acid</i> )
rRNA	RNA ribossômico
SNP	Polimorfismo de nucleotídeo único ( <i>Single Nucleotide Polymorphism</i> )

STR      *Short Tandem Repeat*

T      Timina

tRNA    RNA transferência

## SUMÁRIO

<b>INTRODUÇÃO .....</b>	12
<b>1      OBJETIVOS .....</b>	44
<b>1.1    Geral .....</b>	44
<b>1.2    Específicos .....</b>	44
<b>2      METODOLOGIA .....</b>	46
<b>2.1    Seleção de amostras .....</b>	46
<b>2.1.1    Rio de Janeiro .....</b>	46
<b>2.1.2    Paraguai .....</b>	46
<b>2.1.3    Peru .....</b>	47
<b>2.2    Sequenciamento do mtDNA .....</b>	49
<b>2.2.1    Sequennciamento de Sanger .....</b>	49
<b>2.2.1.1    Rio de Janeiro .....</b>	49
<b>2.2.1.2    Paraguai e Peru .....</b>	51
<b>2.2.2    Sequenciamento massivo paralelo.....</b>	52
<b>2.3    Classificação de haplogrupos e análise estatística .....</b>	53
<b>3      RESULTADOS E DISCUSSÃO.....</b>	55
<b>3.1    Caracterização do patrimônio genético de origem materna de indivíduos residentes na cidade do Rio de Janeiro, Brasil, através da análise do mtDNA .....</b>	55
<b>3.2    Caracterização do patrimônio genético de origem materna de indivíduos residentes no Paraguai, através da análise do mtDNA.....</b>	65
<b>3.3    Caracterização do patrimônio genético de origem materna de indivíduos pertencentes ao grupo linguístico Ashaninka do Peru através da análise do mtDNA .....</b>	78
<b>DISCUSSÃO FINAL E CONCLUSÕES .....</b>	88
<b>REFERÊNCIAS.....</b>	92
<b>ANEXO - Extração orgânica de sangue total - mancha em papel fta.....</b>	100
<b>APÊNDICE A – Artigo “<i>Contrasting admixture estimates in Rio de Janeiro obtained by different sampling strategies</i>”.....</b>	102
<b>APÊNDICE B – Material suplementar Artigo 1 “<i>Defining mtDNA origins and</i></b>	

<i>population stratification in Rio de Janeiro”.....</i>	105
<b>APÊNDICE C – Haplótipos do Artigo 2 “Paraguay: Unveiling migration patterns with ancestry genetic markers” .....</b>	113
<b>APÊNDICE D – Material suplementar do Artigo 3 “Defining mtDNA origins and population stratification in Rio de Janeiro”.....</b>	118
<b>APÊNDICE E - Material suplementar do Artigo 4 “The maternal inheritance of the Ashaninka native group from Peru”.....</b>	125
<b>APÊNDICE F Pôster “Tracing Arawakan Southern Dispersal: Clues from Mitogenome Sequencing of Southern Amerinds”.....</b>	129

## INTRODUÇÃO

A origem do Homem moderno, a sua dispersão pelo mundo e seus contínuos movimentos migratórios têm sido alvo de investigação por diversos séculos. No estudo da história das populações humanas são frequentemente usadas ferramentas linguísticas ou arqueológicas. No entanto, estas disciplinas apresentam algumas limitações que podem levar a teorias contraditórias e a explicações díspares para os mesmos eventos. Nos anos 90, o desenvolvimento tecnológico na área da biologia molecular estimulou o estudo genético das populações, numa tentativa de complementar os dados históricos existentes. Aplicando os princípios da hereditariedade genética mendeliana, a genética populacional foca-se no estudo da composição genética das populações, com o objetivo de elucidar a origem e distribuição da variação encontrada em diferentes grupos populacionais. Desta forma, através da análise da composição genética das populações humanas tem sido possível rastrear a sua ancestralidade e mapear padrões de migração ocorridos no passado. Além disso, o conhecimento da estrutura genética das populações é também vantajoso em outras áreas da genética, como em estudos clínicos e na casuística forense. Em genética clínica é importante o mapeamento da heterogeneidade associada a populações miscigenadas, uma vez que diferentes padrões de mistura genética podem resultar em diferentes graus e tipos de susceptibilidade a doenças. Já na área forense, a elaboração de bases de dados de perfis genéticos, com a descrição detalhada das frequências genotípicas, permite atribuir um peso estatístico às evidências genéticas, o qual é formulado com base em hipóteses alternativas, mutuamente exclusivas.

### Marcadores genéticos

É no genoma humano que se encontra toda a informação hereditária de um indivíduo. O DNA está localizado no núcleo das células, organizado em 23 pares de cromossomos (22 pares autossômicos e 1 par sexual), compostos por duas cadeias polinucleotídicas antiparalelas lineares, e no interior das mitocôndrias, com estrutura circular. A informação genética contida nos autossomos apresenta herança biparental,

sendo transmitida por ambos os progenitores para os descendentes. Já a mitocôndria e o cromossomo Y têm transmissão uniparental (o mtDNA é transmitido pelas mães e o cromossomo Y é transmitido de e para os homens).

As diferenças genéticas entre os humanos representam apenas cerca de 0,3 % da sequência de DNA, sendo que mais de 99,7% do genoma é partilhado por todos os indivíduos (BUTLER, 2009). Essa pequena porcentagem de variabilidade pode ser detectada tanto na região codificante como na região não codificante do genoma. Em casos de identificação forense recorre-se comumente ao estudo de regiões não codificantes, uma vez que os polimorfismos localizados nessas regiões são menos propensos a revelar informações sobre doenças ou suscetibilidades genéticas. No entanto, na última década, com o surgimento de novas tecnologias de sequenciamento, a análise da variabilidade nas regiões codificantes do genoma começou a ser aplicada na casuística forense, para previsão de traços físicos, na ausência de suspeitos. Já em estudos populacionais, o uso de marcadores localizados nas regiões codificantes poderá ser vantajoso, uma vez que, estando sujeitos ao efeito de seleção, poderão uma maior diferenciação entre populações (AMORIM; BUDOWLE, 2016).

A variabilidade genética existente nas populações humanas pode ser detectada pela presença de polimorfismos de comprimento ou de sequência.

Os polimorfismos de comprimento são variações no tamanho das sequências de DNA e resultam, frequentemente, de repetições de sequência em *tandem* cujo número varia de indivíduo para indivíduo. Se as repetições tiverem um comprimento entre 6 pbs a mais de 100 pbs são chamadas de minisatélites (ou VNTRs, do inglês *Variable Number of Tandem Repeats*). Por outro lado, os polimorfismos com unidades de repetição que variarem entre 1-5 pbs são denominados de microssatélites (ou STRs, do inglês *Short Tandem Repeats*). Os polimorfismos de comprimento podem também estar associados à presença ou ausência de um ou mais nucleotídeos, alteração denominada de inserção/deleção.

Os polimorfismos de sequência resultam da substituição de uma base por outra. Quando a mutação é de uma só base, denomina-se por polimorfismo de base única (ou SNP, do inglês *Single Nucleotide Polymorphism*).

Em estudos de genética populacional e forense, a escolha do tipo de marcador e a sua localização no genoma depende dos objetivos e das finalidades propostos. Por exemplo, em estudos de parentesco são comumente utilizados marcadores STR

localizados nos autossomos. Por outro lado, na determinação da ancestralidade paterna de um indivíduo, recorre-se a marcadores do tipo SNP localizados no cromossomo Y.

O genoma mitocondrial, apesar de representar apenas cerca de 0,25% do conteúdo genômico de uma célula, apresenta particularidades, que fazem deste marcador genético um recurso importante em diversas investigações (AMORIM; BUDOWLE, 2016).

## A mitocôndria e o DNA mitocondrial

As mitocôndrias são organelas celulares localizadas no citoplasma das células eucariotas e estão envolvidas em diversos processos metabólicos. São delimitadas por duas membranas de composição e função distintas - membrana externa e membrana interna. Estas membranas encerram dois compartimentos denominados por matriz e espaço intermembranas (Figura 1).

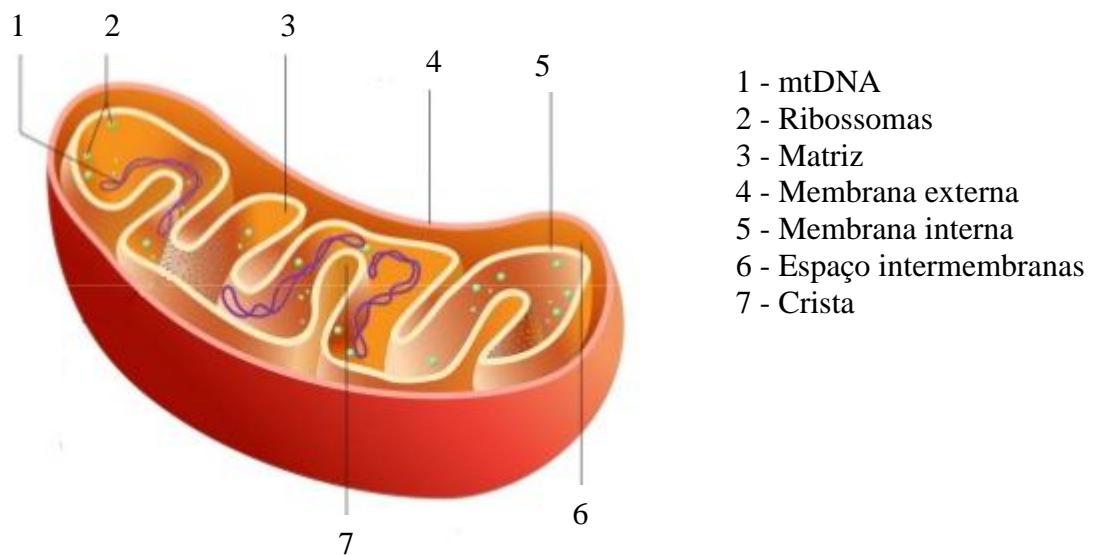
A produção de energia sob a forma de ATP (Adenosina TriFosfato) através da fosforilação oxidativa é a principal função metabólica da mitocôndria. De forma simplificada, a produção de ATP acontece com o fluxo de prótons do espaço intermembranas para a matriz da organela (NELSON; COX, 2004).

A incorporação das mitocôndrias no interior das células eucariotas é explicada pela teoria endossimbiótica. Sendo atualmente aceita de forma consensual pela comunidade científica, esta hipótese postula que as mitocôndrias teriam evoluído a partir de bactérias aeróbias, integradas no seio de células eucariotas primitivas. A relação de simbiose terá garantido à bactéria a sua sobrevivência e, através de processos evolutivos, o ser endossimbionte teria perdido a capacidade de sobreviver fora das células eucariotas, evoluindo e dando origem à mitocôndria atual. As mitocôndrias tornaram-se organelas semiautônomas que mantêm uma relação de dependência com o hospedeiro; células com mitocôndrias dependem destas organelas para realizarem a fosforilação oxidativa e estas, por sua vez, dependem da célula para garantir a sua própria existência (LODISH et al., 2000; NELSON; COX, 2004).

O DNA mitocondrial (mtDNA) humano está localizado na matriz das mitocôndrias (Figura 1) e é constituído por duas cadeias polinucleotídicas complementares, de composição nucleotídica distinta. A cadeia pesada (do inglês

*Heavy*, representada por H) tem uma maior porcentagem de Guaninas (G) e Adeninas (A), que contribuem para um peso molecular superior, enquanto a cadeia leve (do inglês *Light*, representada por L) é rica em Citosinas (C) e Timinas (T) apresentando, consequentemente, um menor peso molecular. Estas cadeias organizam-se de forma circular e têm uma extensão de aproximadamente 16.569 pbs, número de nucleotídeos que pode variar com inserções e/ou deleções na sequência (Figura 2) (LODISH et al., 2000; VOET; VOET; PRATT, 2012).

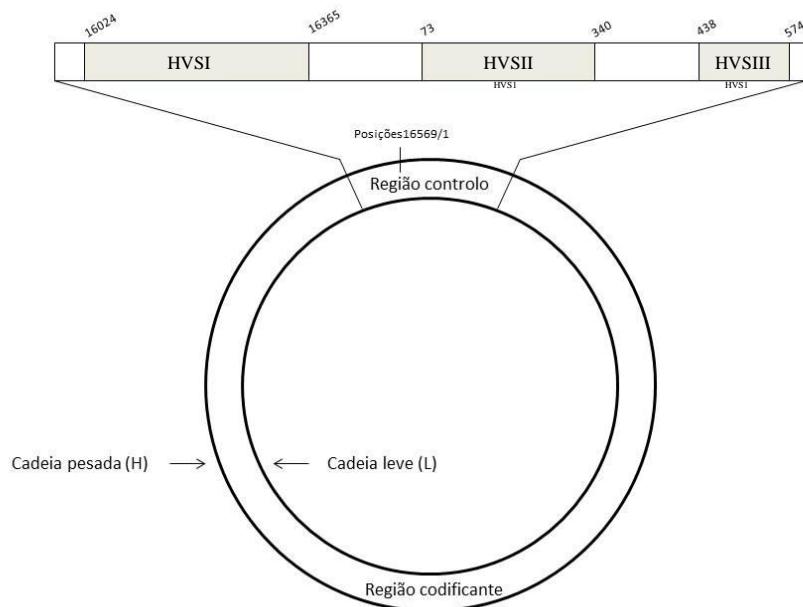
Figura 1- Estrutura da mitocôndria



A molécula de mtDNA é composta por duas regiões - a região codificante e a região controle (Figura 2) - classificadas de acordo com a função que desempenham na molécula. A região codificante representa cerca de 90% do genoma mitocondrial e comprehende 37 genes que se traduzem em 22 tRNAs, 2 rRNAs e 13 polipeptídios. Contudo, grande parte dos polipeptídios envolvidos em funções mitocondriais é codificada por genes nucleares, traduzidos nos ribossomos citoplasmáticos e, posteriormente, transportados por enzimas especializadas até à mitocôndria (NELSON; COX, 2004). O genoma mitocondrial é extremamente compacto, com ausência de íntrons entre os genes e com sequências de genes sobrepostas, contíguas ou separadas por apenas uma ou duas bases não codificantes (CUMMINS, 1998).

A porção não codificante do mtDNA tem uma extensão aproximada de 1.120 pbs e é denominada região controle por conter nela a origem de replicação da cadeia pesada. Esta região é também designada por D-loop (*Displacement loop*), uma vez que, durante o início da transcrição e replicação, o DNA forma uma estrutura de cadeia tripla nesta região. Por não incluir informação necessária à codificação de nenhuma substância essencial ao funcionamento celular, as mutações acumuladas nesta região são fixadas com maior frequência. Por este motivo, fazem parte da região controle três segmentos de elevada variabilidade genética denominados por segmento hipervariável I (HVS1 do inglês *HiperVariable Segment I*), segmento hipervariável II (HVSII do inglês *HiperVariable Segment II*) e segmento hipervariável III (HVSIII do inglês *HiperVariable Segment III*).

Figura 2 - Representação da estrutura circular do genoma mitocondrial



Legenda: A cadeia pesada (H) está representada pela linha externa e a cadeia leve (L) pela linha interna. A região controle apresenta os segmentos de elevada variabilidade genética HVS1, HVSII e HVSIII e a região codificante inclui os 37 genes codificados pela molécula.

Em 1981, a sequência completa do DNA mitocondrial Humano foi descrita pela primeira vez, na sua totalidade, por Anderson e colaboradores (ANDERSON et al.,

1981), no laboratório de Frederick Sanger, em Cambridge, motivo pelo qual ficou conhecida como Sequência de referência de Cambridge (CRS, do inglês *Cambridge Reference Sequence*). Cada posição nucleotídica foi numerada, com início na origem de replicação da cadeia pesada (posição 1) e término na posição 16.569. O primeiro segmento hipervariável está localizado entre as posições 16.024 e 16.365, o segundo entre as 73 e 340 e o terceiro segmento entre as posições 438 e 574 da região controle.

Mais tarde, o genoma mitocondrial estudado por Anderson foi reanalisado pelo grupo de Andrews e colaboradores (ANDREWS et al., 1999) que identificaram 11 erros na sequência descrita em 1981. As correções foram publicadas e a sequência passou a ser denominada por Sequência de Referência de Cambridge revista (rCRS, do inglês *revised Cambridge Reference Sequence*). A rCRS é considerada a sequência padrão de comparação em genética populacional e forense.

Em 2012, Behar e colaboradores (BEHAR et al., 2012) propuseram a implementação de uma nova sequência de referência, hipotética, denominada por *Reconstructed Sapiens Reference Sequence* (RSRS). Esta sequência manteve a mesma anotação numérica que a rCRS, mas representa o genoma ancestral da Eva mitocondrial, mulher da qual todos os genomas humanos mitocôndrias atuais terão descendido. Contudo, a sua aplicação em futuros estudos da molécula de mtDNA implicaria uma alteração da nomenclatura de todas as sequências descritas até à época e a alteração de milhares de dados de genomas mitocondriais. Em forense é importante dispor de uma sequência de referência estável e a implementação da RSRS implicaria uma atualização constante, à medida que mais sequências de mtDNA fossem estudadas. Desta forma, a rCRS continua a ser utilizada como sequência de referência e a RSRS está disponível apenas para consulta (SALAS et al., 2012).

## **Características do genoma mitocondrial**

O uso de marcadores de linhagem está limitado a alguns casos de investigação forense, pelo seu reduzido poder de discriminação comparativamente aos marcadores autossomais. Quando usado em estudos populacionais, o genoma mitocondrial apenas nos fornece a visão materna da história. No entanto, este apresenta outras

particularidades que o tornam um marcador genético valioso, tanto na área forense como em genética de populações.

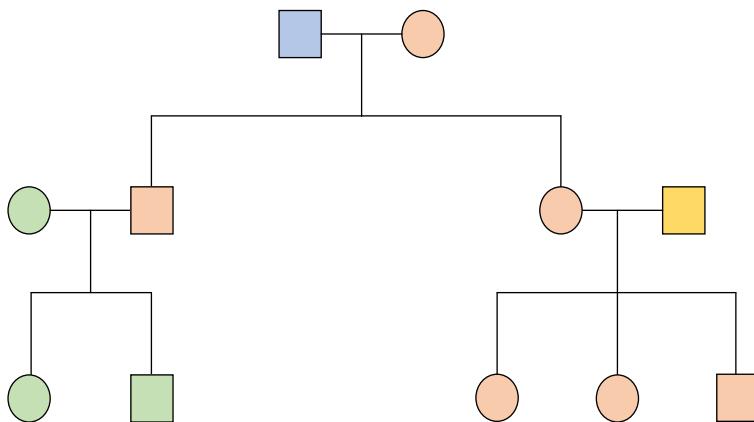
### Herança uniparental materna

O mtDNA apresenta uma transmissão uniparental materna. Durante a fecundação, apenas o núcleo do espermatozoide entra no oócito, fundindo-se diretamente com o núcleo materno, não contribuindo com outros elementos celulares. Mais ainda, um sistema de destruição seletiva por marcação com ubiquitina elimina possíveis mitocôndrias paternas (CUMMINS, 1998; PAKENDORF; STONEKING, 2005a; SUTOVSKY et al., 2000). Desta forma, a transmissão uniparental, apesar de não garantir a identificação individual, permite rastrear linhagens maternas ao longo do tempo, sem os efeitos associados à herança biparental e recombinação (Figura 3).

### Ausência de recombinação

Contrariamente ao DNA nuclear, a molécula de mtDNA não sofre recombinação e, por esse motivo, a informação genética contida no mtDNA apenas revela informação do ascendente materno. A ausência de recombinação faz com que a variabilidade existente no genoma mitocondrial resulte apenas de mutações que se acumularam ao longo de gerações.

Figura 3 - Representação da herança uniparental materna associada ao genoma mitocondrial



Legenda: As mulheres estão representadas por círculos e os homens por quadrados. Cada cor representa uma linhagem diferente de mtDNA.

### Elevada taxa de mutação

A taxa de mutação do genoma mitocondrial é superior à do nuclear. A proximidade da molécula com a cadeia de transferência de elétrons, e a consequente exposição a espécies reativas de oxigênio, aumenta a probabilidade de ocorrência de mutação. Por outro lado, a ausência de histonas protetoras faz com que o genoma mitocondrial seja vulnerável ao ataque oxidativo e a baixa fidelidade da polimerase torna o sistema de correção e reparação de erros menos eficiente durante a replicação do mtDNA.

O cálculo da taxa de mutação no genoma mitocondrial começou a ser alvo de estudo assim que a variabilidade genética encontrada na molécula começou a ser aplicada em estudos genéticos. No entanto, ao longo dos anos, têm sido propostas diferentes metodologias de cálculo e, consequentemente, divulgados diferentes valores de taxa de mutação. O valor de taxa de mutação mitocondrial comumente usado é a taxa de substituição da região controle, publicada por Mishmar et al. (MISHMAR et al., 2003) -  $1,26 \times 10^{-8}$  substituição de base, por nucleotídeo, por ano.

Um dos principais motivos de discrepância entre valores publicados de taxa de mutação mitocondrial é o modelo de cálculo aplicado. Existem duas principais abordagens - uma com base em estudos de *pedigree* e outra em estudos filogenéticos. Os ramos mais antigos de mtDNA aparentam ter uma maior proporção de mutações

sinônimas do que os ramos mais recentes. Esta diferença tem sido explicada pela ação de seleção negativa. Por este motivo, a taxa de mutação obtida por métodos baseados na análise de *pedigree* tende a ser 10 vezes mais rápida que a obtida por métodos filogenéticos. Em estudos de *pedigrees* é calculada a taxa germinativa, com a comparação de sequências entre pais e filhos onde são consideradas apenas algumas gerações. Por outro lado, a taxa evolutiva, obtida através de estudos filogenéticos, considera escalas de tempo de milhares ou até mesmo milhões de anos, permitindo a extração de valores de taxa de mutação contabilizando somente aquelas mutações que se mantiveram nas populações e que não foram eliminadas por fenômenos de deriva genética. Em casos de identificação forense, onde se analisam amostras de indivíduos separados por poucas gerações, o valor de frequência das mutações germinativas é de maior relevância. No entanto, em estudos evolutivos, a utilização da taxa germinativa levaria a uma subestima da antiguidade dos haplogrupos presentes numa população, uma vez que, nestes casos, não seriam contabilizadas aquelas mutações que se foram perdendo ao longo do tempo, por deriva (SOARES et al., 2009).

O efetivo populacional pode também influenciar os valores obtidos de taxa de mutação do mtDNA. Em populações com elevado número de indivíduos, mais mutações germinativas são preservadas. Por esse motivo, a taxa de mutação será inferior em populações de efetivo reduzido ou que sofreram mecanismos de afunilamento genético, onde as mutações germinativas foram perdidas.

Além disso, a taxa de mutação não se encontra distribuída uniformemente ao longo do genoma mitocondrial, isto é, a velocidade de mutação na molécula é heterogênea. A região controle é mais suscetível a sofrer alterações que a região codificante porque as mutações desta última são alvo de seleção negativa. Para além disso, a região não codificante apresenta *hotspots* mutacionais que variam quatro a cinco vezes mais rápido que qualquer outro local dessa mesma região (PAKENDORF; STONEKING, 2005b).

Em 2009, Soares e colaboradores (SOARES et al., 2009) propuseram um valor de taxa de mutação baseado num método de reconstrução filogenética usando mais de 2.000 mitogenomas publicados na literatura, considerando a presença de heterogeneidade ao longo do genoma mitocondrial e o efeito da seleção negativa sobre mutações não sinônimas. Desta forma, obtiveram uma taxa de substituição média na ordem de  $10^{-8}$  tanto para a molécula completa como para as regiões controle e codificante em separado (SOARES et al., 2009). Além do mais, ao desconsiderarem os

principais *hotspots* mutacionais (por exemplo, polimorfismos nas posições 16.519, 152, 16.311, 146, 195) verificaram que as taxas de mutação das regiões controle e codificante eram muito semelhantes, corroborando a necessidade de uma taxa de mutação mitocondrial geral na determinação da antiguidade dos haplótipos (SOARES et al., 2009).

### Elevado número de cópias

O número de cópias de mtDNA por célula varia. Cada célula pode conter centenas de mitocôndrias no seu interior e cada mitocôndria tem em média 4 a 5 cópias de mtDNA. Neste sentido, uma única célula pode possuir entre centenas a milhares de cópias de moléculas de mtDNA. Esta particularidade torna o material genético da mitocôndria mais fácil de obter, sendo, por isso, muito utilizado no estudo de DNA antigo e de amostras degradadas (BUTLER, 2005; IBORRA; KIMURA; COOK, 2004).

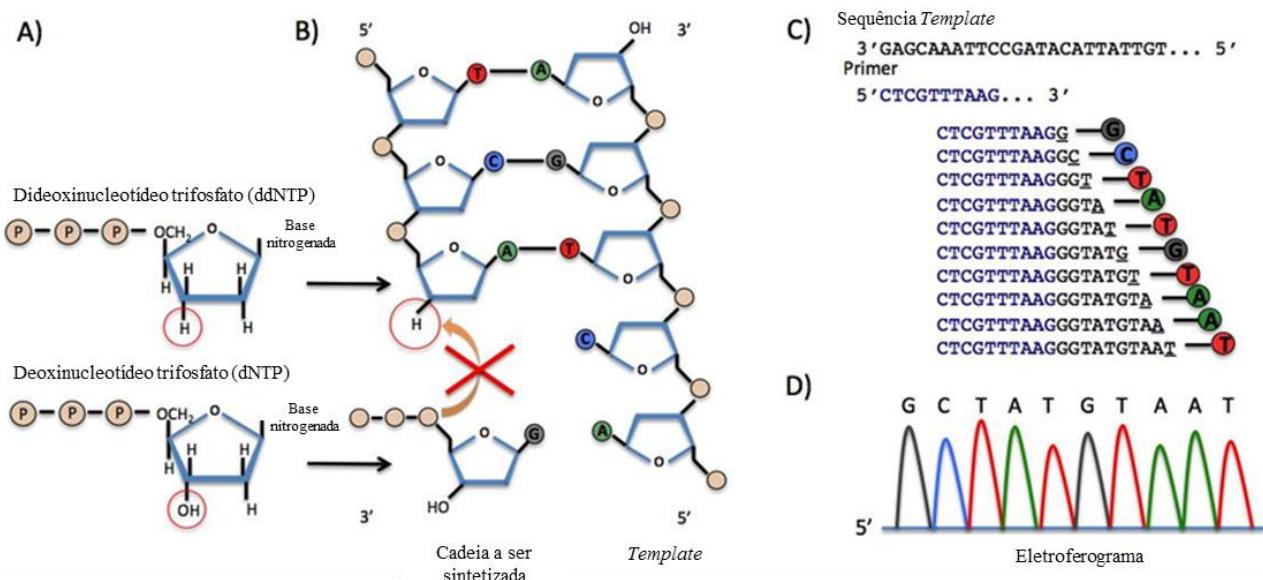
### **Metodologias de análise**

Os primeiros estudos sobre a variabilidade do genoma mitocondrial envolviam metodologias de restrição enzimática, com a análise de polimorfismos de comprimento de fragmento de restrição (RFLPs, do inglês *Restriction Fragment Length Polymorphism*) (BROWN, 1980). Mais tarde, com os avanços tecnológicos e obtenção de produtos de PCR e sequências de elevada qualidade, começou-se por sequenciar o segmento HVSI da região controle do mtDNA, localizado entre as posições 16.024 e 16.365, por ser a região mais variável. Com a crescente necessidade de aumentar o poder de discriminação nas comparações entre sequências, tornou-se comum sequenciar a região controle total do mtDNA, entre as posições 16.024 e 576, através do método de Sanger (SANGER; NICKLEN, 1977). Na química de sequenciamento de Sanger são incorporados ddNTPs na reação, bases que, por não apresentarem o grupo hidroxila na extremidade 3', impossibilitam a ligação de um novo nucleotídeo e, por isso, funcionam como terminadores de síntese de DNA (Figura 4a). Desta forma, quando a polimerase

introduz um novo ddNTP na cadeia a ser sintetizada, o crescimento desta termina (Figura 4b). Cada um dos ddNTPs está marcado com um fluorocromo, o que permite a sua detecção e identificação durante a análise (Figura 4c e Figura 4d). Nesta metodologia, o sequenciamento é realizado com *primers* direto e reverso, de forma a obter-se informação complementar, garantido um controle de qualidade dos resultados.

Por vezes, a separação de duas linhagens maternas não é possível apenas com o estudo da região controle. Na tentativa de ultrapassar esta limitação, começou-se por estudar polimorfismos SNPs localizados na região codificante do mtDNA. Estes marcadores são vantajosos por apresentarem uma sequência de menor comprimento que os STRs do DNA autossomal ou que as regiões hipervariáveis do mtDNA, permitindo um maior sucesso na análise de amostras degradadas.

Figura 4 - Esquema representativo do método de sequenciamento de Sanger



Legenda: No sequenciamento de Sanger são incorporados Deoxinucleotídeos trifosfatados (dNTPs) e Dideoxynucleotídeos trifosfatados (ddNTPs) à reação (a); os ddNTPs, por não apresentarem o grupo hidroxila na extremidade 3' (a), funcionam como terminadores da síntese da cadeia de DNA (b). Durante o sequenciamento, são adicionados de forma aleatória à cadeia *template* de DNA, dNTPs ou ddNTPs (c), o que origina sequências de comprimento variado (c). A detecção de cada posição nucleotídica é efetuada por eletroforese capilar, pela migração dos fragmentos de acordo com o seu tamanho. O dNTP está marcado com um fluorocromo, o que permite a sua detecção e identificação durante a análise (d).

Fonte: Adaptado de [https://www.researchgate.net/figure/The-Sanger-sequencing-method-A-simplified-schematic-of-the-Sanger-sequencing-method\\_fig8\\_303565848](https://www.researchgate.net/figure/The-Sanger-sequencing-method-A-simplified-schematic-of-the-Sanger-sequencing-method_fig8_303565848)

Na região codificante, existem SNPs com baixa taxa de mutação que, por isso, apresentam uma baixa probabilidade de recorrência e, consequentemente, uma elevada especificidade geográfica. Apesar da densidade de SNPs ser superior na região controle, por ser 14 vezes maior, a região codificante permite a análise de um maior número de *loci* polimórficos. A metodologia comumente utilizada para o estudo de SNPs é o mini sequenciamento (ou sequenciamento de base única), através de SNaPshot. Este procedimento é semelhante ao sequenciamento de Sanger, com a diferença de que apenas são adicionados à reação ddNTPs. Os *primers* são desenhados para terminarem na posição nucleotídica anterior à posição de interesse e, uma vez que não são adicionados nucleotídeos convencionais (dNTPs), a extensão é realizada para uma única base (PARSONS, 2006).

As metodologias convencionais baseadas na detecção de bases nucleotídicas ou na determinação do tamanho de fragmentos por eletroforese apresentam algumas limitações, como a perda de informação em amostras degradadas. A procura de novos métodos, mais precisos e eficientes, surgiu da necessidade de se obterem resultados de alta precisão e reproduzibilidade num curto espaço de tempo e a custos acessíveis.

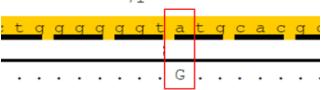
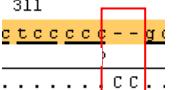
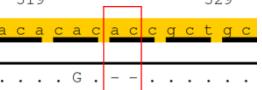
Durante as últimas décadas, têm surgido novas técnicas de sequenciamento em larga escala, denominadas de sequenciamento paralelo massivo (MPS do inglês *Massively Parallel Sequencing*), com aplicação em diversos campos da biologia. Estas técnicas, que não têm por base a química de Sanger, consistem na quebra de sequências de DNA em pequenos fragmentos (aproximadamente entre 400-600 pbs), sequenciamento destes em paralelo, seguido da montagem e ordenação num único fragmento contíguo. O produto sequenciado é diretamente detectado sem que haja a necessidade de realizar eletroforese capilar. Diversos estudos do genoma mitocondrial por MPS demonstraram a vantagem da utilização destas técnicas quando aplicadas à área forense, com obtenção de resultados de alta precisão e sensibilidade, alto rendimento, com um baixo custo e simples operação (HOLLAND; MCQUILLAN; O'HANLON, 2011; JUST; IRWIN; PARSON, 2015; KING et al., 2014; PARSON et al., 2015). Desta forma, é agora possível obter uma maior quantidade de informação, com sequências mais extensas (mais bases detectadas) e melhor cobertura por base. A detecção da variabilidade encontrada na molécula completa de mtDNA garante uma maior precisão na descrição dos haplogrupos. Efetivamente, novas linhagens têm sido descritas graças ao uso de metodologias MPS (por exemplo) (BRANDINI et al., 2018).

## Nomenclatura

Após sequenciamento do genoma mitocondrial, as sequências obtidas são descritas por comparação com a sequência de referência rCRS e apenas as posições que não coincidem entre ambas são anotadas. Desta forma, a sequência em estudo é representada pelo conjunto de posições diferentes em relação à rCRs, estando subentendido que as restantes posições são idênticas. O conjunto de polimorfismos identificados numa sequência denomina-se por haplótipo, o qual caracteriza uma determinada linhagem materna. Em forense, a nomenclatura dos haplótipos é feita com base em diretrizes detalhadamente descritas pela comunidade científica, por forma a evitar erros de interpretação (BÄR et al., 2000; PARSON et al., 2014b; TULLY et al., 2001). A existência de uma nomenclatura única, não só facilita o intercâmbio de resultados entre laboratórios, como também permite a partilha e comparação, a nível internacional, de perfis contidos em bases de dados. Esta é também essencial no desenvolvimento de testes de proficiência que permitam aferir a qualidade dos resultados obtidos pelos laboratórios.

Os polimorfismos são classificados pela posição nucleotídica em que ocorrem, seguida do nucleotídeo que resultou da alteração. Esta regra pode ser exemplificada pela posição 73 do mtDNA; na sequência de referência esta posição é caracterizada pela presença de uma adenina (A), a qual, numa grande porcentagem de linhagens, é substituída por uma guanina (G). Neste sentido, estando esta mutação localizada na posição 73 e sendo uma transição de A para G, o polimorfismo é identificado como “73G” ou “A73G” (Tabela 1). As substituições são o tipo de alteração mais comum no genoma mitocondrial. No entanto, os polimorfismos podem também apresentar-se sob a forma de inserções ou deleções. As inserções são identificadas pela posição nucleotídica em que ocorrem seguida de “.1” caso haja a inserção de uma base, “.2” caso sejam inseridas 2 bases, e assim sucessivamente; o nucleotídeo inserido deve ser indicado depois da indicação do número de inserções. No segmento poli-C de HVSII, entre as posições 310 e 316, a inserção de citosinas (C) é extremamente comum. A adição de uma C na posição 315 será identificada como “315.1C”; caso sejam duas Cs será “315.1C 315.2C” (Tabela 1). As deleções são classificadas pela posição nucleotídica seguida do prefixo “DEL” ou “del”, como por exemplo “523del” ou “523DEL” (Tabela 1) (PARSON et al., 2014b).

Tabela 1- Regras de anotação de polimorfismos na molécula de mtDNA

Tipo de polimorfismo	Imagen	Nomenclatura
Substituição	 73G	73G
Inserção	 315.1C 315.2C	315.1C 315.2C
Deleção	 523del 524del	523del 524del

Legenda: Os nucleotídeos a amarelo correspondem à rCRS; os pontos representam as posições da sequência em estudo que são idênticas à rCRS.

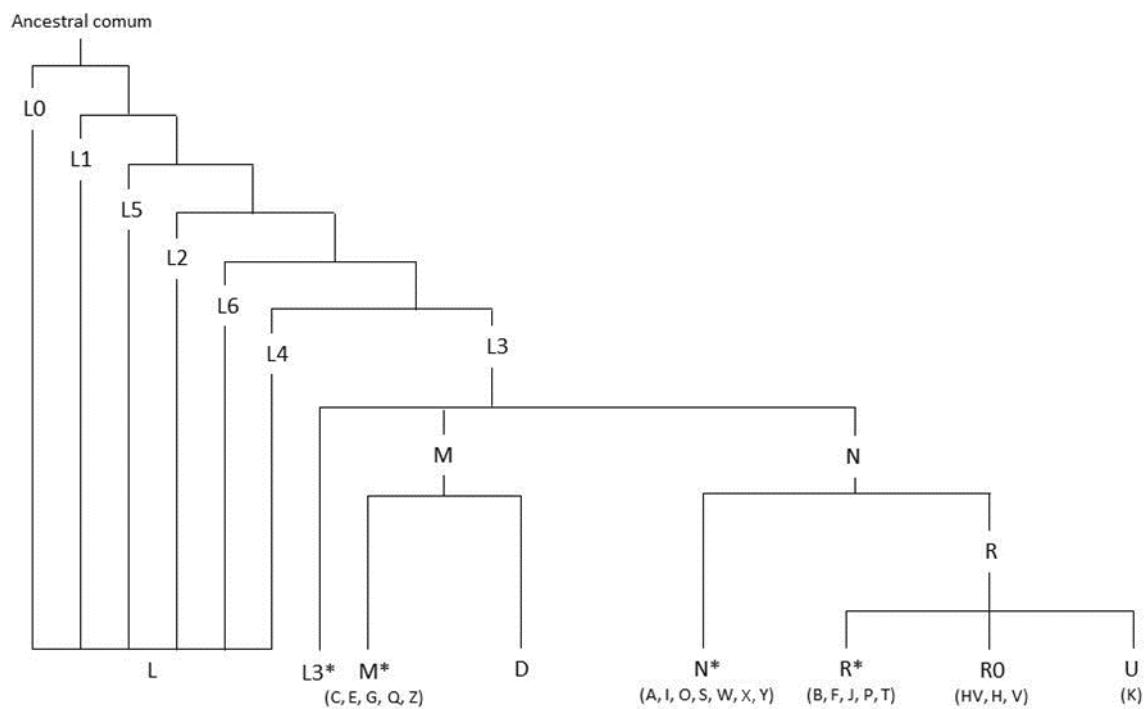
## Variabilidade do mtDNA

Os haplogrupos são nomeados por letras do alfabeto e formam-se a partir do agrupamento de haplótipos de mtDNA com polimorfismos específicos, que representam linhagens próximas umas das outras. Sequências de mtDNA que pertencem a um determinado haplogrupo, ou sub-haplogrupo dentro deste, apresentam alterações específicas, representativas de um ancestral comum. Em 2009, van Oven e Kayser desenvolveram uma árvore filogenética – *Phylotree* (VAN OVEN; KAYSER, 2009) - que englobava os polimorfismos identificados até ao momento no genoma mitocondrial, permitindo a diferenciação entre haplogrupos (Figura 5). A *Phylotree* foi sofrendo atualizações ao longo do tempo, à medida que novos polimorfismos e sequências foram descritos. A última atualização (quando da escrita desta tese) ocorreu em fevereiro de 2016 - <http://www.phylotree.org/>. Os ramos mais profundos e antigos da *Phylotree* incluem haplogrupos da região Subsaariana, nomeadamente os haplogrupos L0 a L6. O haplogrupo L3 ramifica-se nos haplogrupos M e N, que se dispersaram pela Eurásia e deram origem às linhagens presentes em populações desta região. As linhagens europeias derivam do haplogrupo N, onde se incluem os haplogrupos H, I, J, K, T, U, V, W e X que, no seu conjunto, estão presentes em 98% da população europeia. As linhagens asiáticas derivam tanto do haplogrupo M como do haplogrupo N. No primeiro incluem-se as ramificações C, D, E e Z e no segundo os ramos A, B, F e J. Nas

populações nativo-americanas verifica-se a presença dos haplogrupos A, B, C e D, característicos da região asiática, o que indica que a sua presença na região americana é anterior à colonização europeia (Figura 6)

Os ramos mais profundos e antigos da *Phylotree* incluem haplogrupos da região Subsaariana, nomeadamente os haplogrupos L0 a L6. O haplogrupo L3 ramifica-se nos haplogrupos M e N, que se dispersaram pela Eurásia e deram origem às linhagens presentes em populações desta região. As linhagens europeias derivam do haplogrupo N, onde se incluem os haplogrupos H, I, J, K, T, U, V, W e X que, no seu conjunto, estão presentes em 98% da população europeia. As linhagens asiáticas derivam tanto do haplogrupo M como do haplogrupo N. No primeiro incluem-se as ramificações C, D, E e Z e no segundo os ramos A, B, F e J. Nas populações nativo-americanas verifica-se a presença dos haplogrupos A, B, C e D, característicos da região asiática, o que indica que a sua presença na região americana é anterior à colonização europeia (Figura 6).

Figura 5 - Organização dos haplogrupos e suas ramificações



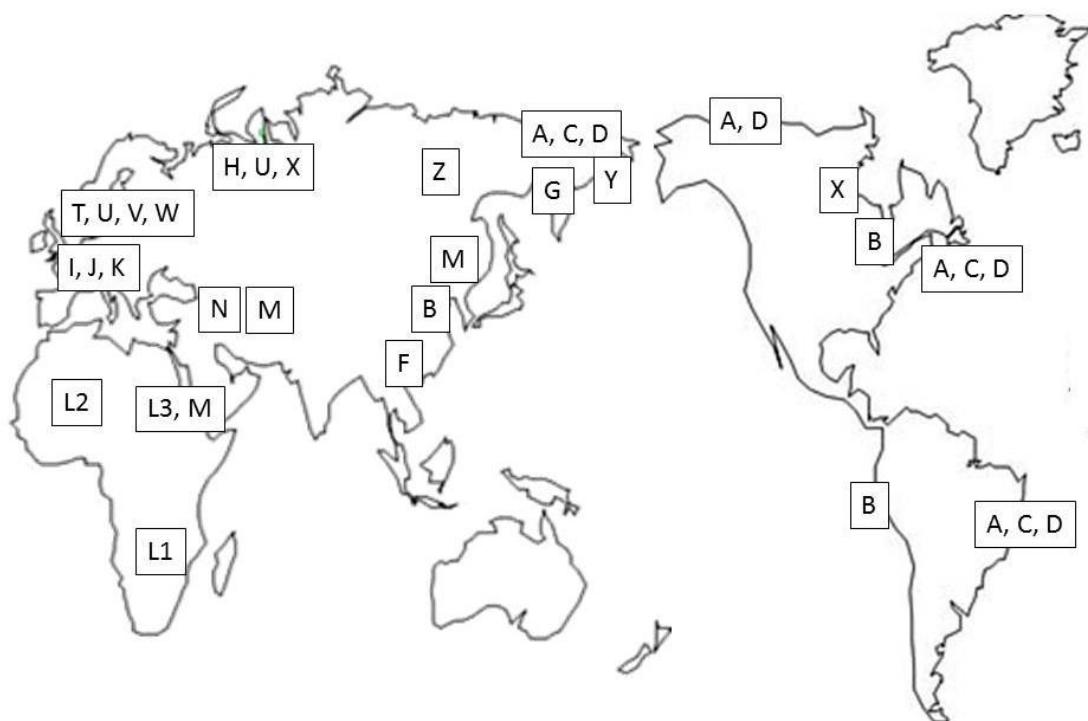
Legenda: Os haplogrupos africanos L0-L6 são os mais antigos da árvore. O L3 ramifica-se nos haplogrupos euroasiáticos M e N que por sua vez se ramificam nos restantes euroasiáticos e nativo-americanos.

Fonte: Adaptado de (VAN OVEN; KAYSER, 2009).

### Haplogrupos africanos

A expansão do grupo linguístico Bantu, com origem na região Centro-Oeste de África e posterior dispersão para o Sul do continente, é considerada um dos principais moduladores da composição genética da região subsaariana de África. Por outro lado, o influxo de linhagens oriundas da região do Próximo Oriente terá afetado, principalmente, a composição genética das populações do Norte e Este do continente (BANDELT; RICHARDS; MACAULAY, 2006).

Figura 6 - Distribuição dos haplogrupos no Mundo



Legenda: As linhagens L são características da região subsaariana de África. Os movimentos migratórios associados à saída de grupos populacionais de África e sua dispersão para as regiões Euroasiáticas, deram origem às linhagens M e N e suas ramificações. As linhagens nativo-americanas A, B C, D e X estão também presentes em grupos populacionais asiáticos.

Fonte: Adaptado de <https://www.familytreedna.com>

Atualmente, grande parte das linhagens detectadas no Norte de África pertence, ou têm raízes comuns, a haplogrupos Euroasiáticos [27]; existindo, no entanto, uma contribuição subsaariana, com cerca de 17 % de linhagens pertencentes ao haplogrupo L [28]. Por outro lado, na região subsaariana, cerca de 92 % das linhagens [28] pertence

ao haplogrupo L, apresentando uma distribuição variável ao longo do continente. O ramo L0 é mais comum na região Sul e Este de África (Figura 7). Por outro lado, o haplogrupo L1 está presente em maior frequência no Sudoeste africano (Figura 7). As linhagens L2 são mais comuns no Sudeste e Centro-Oeste (Figura 7). O L3 tem a sua frequência mais elevada no Sudoeste de África e Centro-Oeste (Figura 7). As linhagens L4 e L5 são mais frequentes no Este do que em outras regiões de África (Figura 7) [28].

Alguns ramos africanos foram previamente descritos como praticamente exclusivos de certos grupos étnicos, como é caso das linhagens L0d e L0k que estão presentes apenas em populações Khoisan [29–32]. Haplótipos pertencentes ao haplogrupos L são extremamente raros fora de África, excepto em regiões geográfica que, ao longo da história humana, receberam fluxo genético deste continente, como é o caso das populações miscigenadas da América.

### Haplogrupos euroasiáticos

Os haplogrupos M, N e R aparecem ter idades muito semelhantes, o que indica que, muito provavelmente, terão participado do mesmo processo migratório, relacionado com a saída de grupos populacionais da região leste da África (MACAULAY et al., 2005).

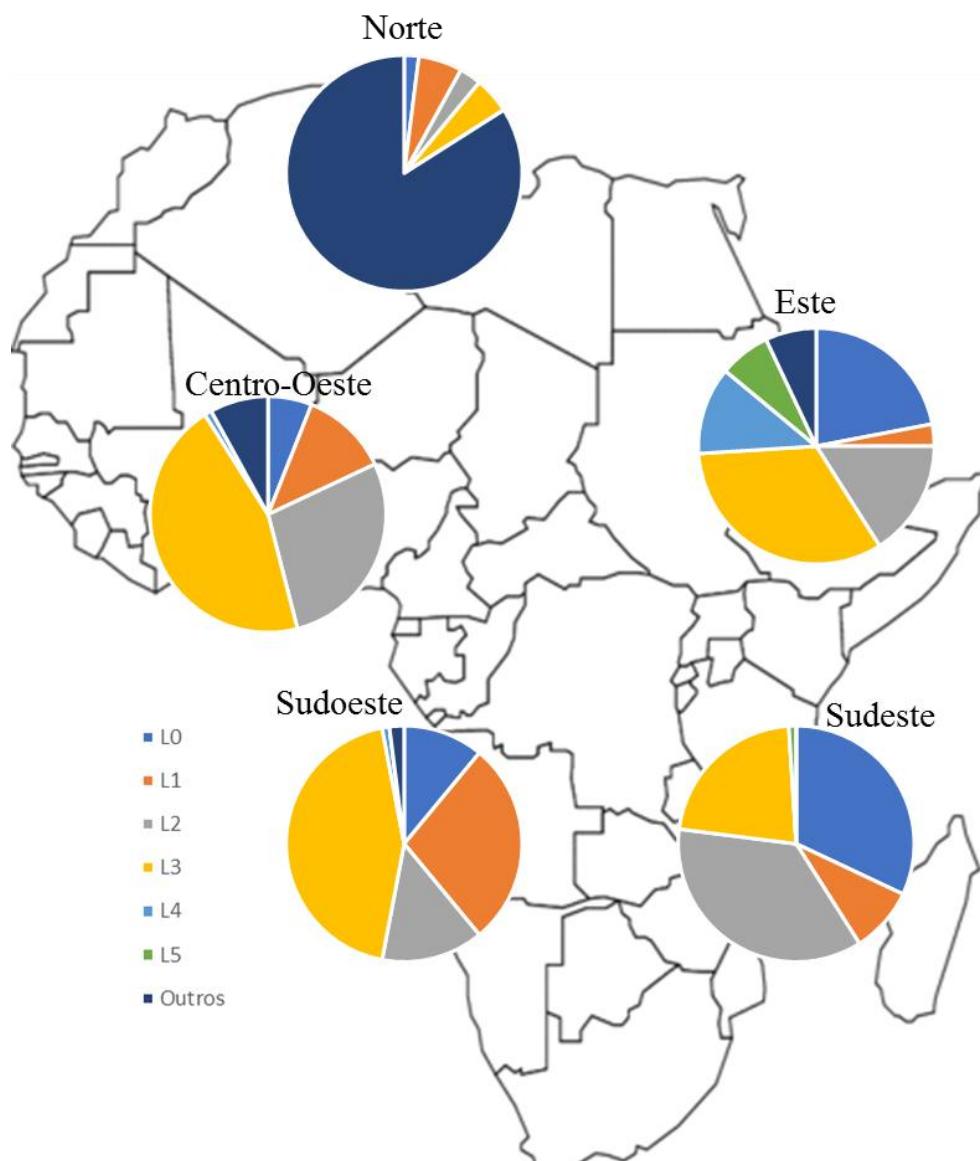
Atualmente, grande parte das linhagens maternas europeias pode ser incluída nos haplogrupos N1, N2, X, R0, JT e U. Os haplogrupos derivados do ramo R (R, JT e U) representam entre 80 % e 90 % da variação total do mitocondrial existente na maioria das populações europeias.

Os haplogrupos U, HV, JT, N1, N2 e X podem ser encontrados na Europa, Sudoeste asiático e Norte de África. Por outro lado, os haplogrupos R5-R8, M2-M6 e M4'67 estão presentes apenas no Sul da Ásia. Por último, os haplogrupos A-G, Z e M7-M9 encontram-se dispersos pelo Este asiático (KIVISILD, 2015).

### Haplogrupos Nativo-americanos

Com a migração para o continente americano, os ancestrais nativo-americanos foram-se separando das populações asiáticas e se tornando geneticamente distintos. No entanto, atualmente é possível detectar nos nativo-americanos linhagens maternas que estão presentes também em grupos asiáticos: haplogrupos A, B, C, D e X (BELLWOOD, 2014).

Figura 7 - Distribuição dos haplogrupos de mtDNA nas 5 regiões de África



Fonte: Adaptado de (CEREZO et al., 2016)

As populações sul-americanas apresentam elevada diversidade étnica e linguística. No entanto, devido aos fortes efeitos de deriva gênica a que estiveram sujeitas, tem sido difícil relacionar a sua estrutura cultural com a composição genética.

Alguns haplogrupos nativos podem ser detectados ao longo de todo o continente, como é o caso das linhagens A2, B2, C1b, C1c, C1d, D1 e D4h3a. Por outro lado, existem haplogrupos nativos que aparentam ter uma distribuição geográfica específica. Por exemplo, as linhagens D1j, D1g, B2i2 e C1b13 foram previamente descritas como restritas ao cone sul da América do Sul (BODNER et al., 2012; DE SAINT PIERRE et al., 2012a). Outras linhagens como D2a, D3 e X2a só foram até agora encontradas em populações da América do Norte.

### **Aplicação forense do mtDNA**

A análise do genoma mitocondrial não é aplicável em diversos cenários forenses, uma vez que não permite a identificação individual, mas sim de linhagens maternas. Contudo, em casos particulares, o mtDNA revela ser um marcador genético útil e a sua eficiência está associada, principalmente, a duas das suas características - herança materna e elevado número de cópias.

A herança materna é uma característica que se torna vantajosa na ausência de familiares diretos de vítimas, que possam fornecer uma amostra referência. Neste sentido, o estudo do genoma mitocondrial pode ser aplicado na análise de restos cadavéricos de pessoas desaparecidas, onde apenas é possível fazer uma comparação com amostras de referência de familiares maternos. Pela ausência de recombinação, indivíduos pertencentes à mesma linhagem materna, mesmo que separados por várias gerações, apresentam o mesmo haplótipo de mtDNA e podem, por isso, servir como referência em certas investigações forenses.

Apesar do DNA nuclear conter muito mais informação genética que o mtDNA, apenas existem duas cópias em cada célula, enquanto o genoma mitocondrial se encontra repetido centenas de vezes numa única célula. Além do mais, a estrutura circular da molécula e o seu confinamento entre duas membranas garantem a preservação do conteúdo genômico mitocondrial de forma mais eficiente. Desta forma, a probabilidade de sobrevivência do mtDNA em condições extremas é superior à do

nuclear. Esta característica é vantajosa quando o material biológico é muito antigo e/ou se encontra degradado, e o mtDNA é a única fonte da qual é possível obter perfis genéticos. O elevado número de cópias é ainda útil quando a quantidade de material biológico disponível para análise é reduzida, como por exemplo, em material genético recolhido de cenas de crime.

A interpretação de um resultado de mtDNA é geralmente obtida pela comparação direta entre uma sequência de origem desconhecida (por exemplo, proveniente de um vestígio encontrado no local do crime) e uma de origem conhecida (por exemplo, amostra do suspeito). Cada laboratório deve delinear as diretrizes a seguir e determinar em que situações as sequências obtidas poderão ser provenientes, ou não, de um mesmo indivíduo.

Quando dois perfis são idênticos, não se pode excluir a hipótese de pertencerem a indivíduos da mesma linhagem materna, sendo necessário estimar a frequência do haplótipo na população. Esta é calculada a partir de bases de dados de referência, representativas das populações em causa. Só assim é possível atribuir um peso estatístico à prova apresentada em tribunal. Deste modo, os laboratórios forenses recorrem a bases de dados próprias, ou disponíveis na literatura, que incluem sequências relevantes no âmbito da investigação.

A mais conhecida base de dados de perfis de mtDNA aplicada à área forense é a EMPOP - EDNAP *forensic mtDNA population database* (W; A, 2007), criada e mantida pelo Instituto de Medicina Legal de Innsbruck. Todas as sequências lá incluídas sofrem um rigoroso controle de qualidade. A base de dados inclui atualmente tanto sequências da região controle, como SNPs da região codificante e alguns mitogenomas completos.

O método mais utilizado no cálculo da probabilidade de coincidência entre perfis de mtDNA tem por base a frequência do perfil numa determinada base de dados. O cálculo da frequência é feito por contagem direta, onde o número de coincidências é dividido pelo número total de perfis. Para tal, a composição da base de dados, isto é, o seu tamanho e a sua representatividade geográfica (origem e distribuição dos perfis incluídos), é um fator importante que influencia o resultado do cálculo da probabilidade de coincidência. O número e a sequência dos haplótipos podem variar entre diferentes populações pelo que a precisão e exatidão do valor de frequência obtido, a partir de uma busca numa base dados, vai depender da representatividade da mesma (PEREIRA; CUNHA; AMORIM, 2004).

## **Problemas na análise do mtDNA**

Existem vários problemas que podem dificultar a interpretação de resultados com base em sequências do genoma mitocondrial, sendo os mais frequentes a contaminação entre amostras, o sequenciamento de pseudogenes nucleares e a presença de heteroplasmia.

### Contaminações

A probabilidade de contaminação durante a manipulação do mtDNA é superior à do DNA nuclear, por o primeiro apresentar um número superior de cópias em cada célula. De forma geral não é possível obter resultados a partir de amostras que apresentem mistura, pela complexidade do sinal obtido nas sequencias analisadas. Mais ainda, a proporção entre a altura dos picos correspondentes a duas bases diferentes é aleatória, não garantindo a identificação de proporção de cada componente, uma vez que a taxa de incorporação de nucleotídeos pode variar.

A análise de misturas tem sido mais aprofundada nos últimos anos, com a implementação de tecnologias de sequenciamento paralelo massivo na análise de mtDNA, uma vez que são identificadas misturas que não eram antes detectadas, quando usadas metodologias convencionais de sequenciamento.

Para evitar a ocorrência de contaminação e minimizar a presença de erros nos resultados apresentados, devem ser tomadas medidas de prevenção. Tully *et al.* (TULLY et al., 2001) e Parson *et al.* (PARSON et al., 2014a) propuseram várias medidas preventivas, algumas das quais são habituais em qualquer análise de DNA em investigações forenses, como sejam: a separação entre as zonas de Pre- e Pos-PCR; o tratamento de amostras de referência e problema em separado; o uso de controles positivos e negativos; o uso de material descontaminado; e a análise em duplicatas.

### Pseudogenes nucleares

Os NUMTs (do inglês *NUclear Mitochondrial DNA segments*) são pseudogenes nucleares resultantes de porções do genoma mitocondrial que migraram e se integraram no genoma nuclear (BUTLER, 2005). Na presença de amostras de qualidade e/ou quantidade reduzidas é por vezes necessário alterar condições da PCR, por aumento do número de ciclos e/ou diminuição da temperatura de *anneling* dos *primers*. Esta diminuição da especificidade da PCR pode levar à amplificação de pseudogenes nucleares, os quais aparecem como contaminação no resultado de mtDNA.

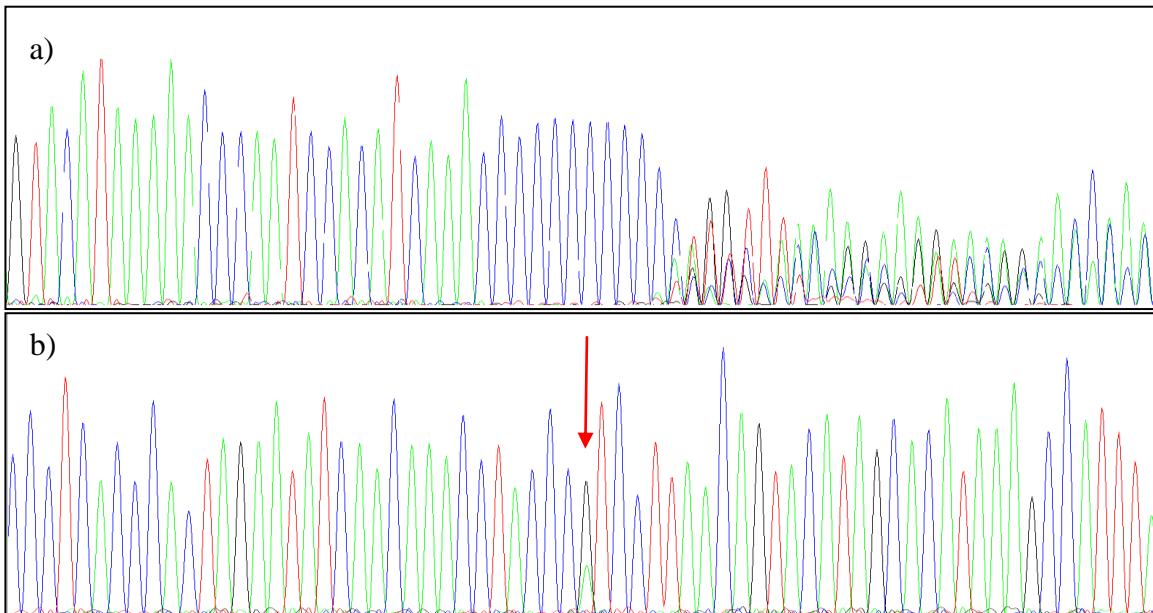
### Heteroplasmia

O fenômeno caracterizado pela presença de diferentes moléculas de mtDNA numa única mitocôndria, numa única célula, num tecido ou num mesmo indivíduo, denomina-se por heteroplasmia. A condição oposta, onde todas as moléculas apresentam sequências idênticas, é denominada homoplasmia. A análise de resultados de mtDNA pode ser dificultada pela presença sequências de mtDNA diferentes, pertencentes ao mesmo indivíduo. A heteroplasmia pode estar associada a inserções e/ou deleções, o que faz variar o comprimento das sequências de mtDNA presentes num mesmo indivíduo e é por isso denominada por heteroplasmia de comprimento (Figura 8a). Por outro lado, se a heteroplasmia resultar de substituições, denomina-se por heteroplasmia de posição (Figura 8b) (MELTON, 2004).

A presença de heteroplasmia é um dos fatores que mais dificulta a interpretação de resultados de mtDNA. Para evitar erros de interpretação, a análise de uma posição ou região heteroplásica deve ser confirmada pelo sequenciamento em duplicata, de preferência utilizando-se primers diferentes.

Para que haja uma padronização e consistência nos resultados publicados, a Sociedade Internacional de Genética Forense (ISFG, do inglês *International Society for Forensic Genetics*) publicou diretrizes detalhadas para classificação de sequências heteroplasmicas (PARSON et al., 2014b). Nas heteroplasmias de comprimento, foi recomendado reportar a variante heteroplasmica dominante, isto é, aquela que está presente em maior frequência. No caso das heteroplasmias de posição, a anotação deve ser realizada de acordo com os códigos de nomenclatura da IUPAC (do inglês International Union of Pure and Applied Chemistry) (Tabela 2).

Figura 8 - Exemplos de sequências de mtDNA com heteroplasmia



Legenda: a) heteroplasmia de comprimento e b) heteroplasmia de posição (apontada pela seta vermelha)

Tabela 2 - Códigos IUPAC

Código	Tradução
B	C, G ou T
D	A, G ou T
H	A, C ou T
V	A, C ou G
R	A ou G
Y	C ou T
K	G ou T
M	A ou C
S	G ou C
W	A ou T
N	Qualquer base

### Populações da América do Sul

Em 1920, o antropólogo Ales Hrdlicka publicou as suas teorias sobre a colonização do continente americano pelo Homem moderno, indicando a região da Sibéria como possível origem para as populações Nativo-americanas. Mais

recentemente, dados arqueológicos, linguísticos e genéticos vieram corroborar esta hipótese e esclarecer as possíveis rotas migratórias pelo continente (GOEBEL; WATERS; O'ROURKE, 2008). Assim, os primeiros colonizadores teriam chegado à América pela região da Beríngia, durante o período Pleistoceno. A concentração de precipitação a elevadas altitudes teria levado à formação de glaciares e, consequentemente, à diminuição do nível do mar, que teria exposto a região continental do Mar *Bering* e aberto caminho térreo entre as atuais regiões do Alasca e Sibéria (Figura 9). A posterior dispersão para o resto do continente teria sido rápida, a partir de um número reduzido de fundadores, motivo pelo qual se encontram padrões genéticos iguais ao longo de todo o território americano (BELLWOOD, 2014; BONATTO; SALZANO, 1997; SCHURR, 2004).

A entrada e dispersão dos Nativo-americanos no Sul do continente é explicada por duas principais teorias. Por um lado, teoriza-se que os grupos populacionais oriundos da região Norte teriam seguido pela costa até ao Sul e, mais tarde teriam atravessado a cordilheira dos Andes em diferentes latitudes. Por outro lado, pensa-se que teria havido uma separação aleatória dos grupos populacionais na região norte da América do Sul, com dois movimentos migratório distintos: um para a região da costa (oeste) e outro para o interior do continente (leste) (Figura 9). Esta última hipótese tem sido corroborada com diversos estudos genéticos que relatam uma diferença na diversidade encontrada entre as regiões leste e oeste do subcontinente (BELLWOOD, 2014; BODNER et al., 2012; ROTHHAMMER; DILLEHAY, 2009; YANG et al., 2010).

Figura 9 - Entrada e dispersão do Homem moderno no continente americano



Fonte: Adaptado de <https://infograph.venngage.com/p/207291/people-of-the-new-world>

No início do século XVI, a composição genética das populações americanas foi fortemente influenciada com a chegada dos colonizadores europeus e o tráfico transatlântico de escravos (HOMBURGER et al., 2015; MESA et al., 2000; SALAS et al., 2004, 2005). Nos dias de hoje, a América do Sul é caracterizada por uma elevada riqueza biológica e diversidade cultural e étnica.

A contribuição diferencial de nativo-americanos, colonizadores europeus e escravos africanos para o *background* genético das populações do subcontinente sul americano, fundamenta a necessidade de construção de bancos de dados de DNA que permitam capturar a alta diversidade e heterogeneidade genética presente nos diferentes grupos populacionais. Por outro lado, o detalhamento da composição genética coadjuva na compreensão dos movimentos migratórios que deram origem às populações sul-americanas atuais.

## Rio de Janeiro

O Brasil está localizado na costa Atlântica da América do Sul, com uma extensão litoral de cerca de 7 mil km. Faz fronteira com os restantes países da América do Sul, exceto Equador e Chile. O Rio de Janeiro é o estado do Brasil com a maior densidade demográfica e está localizado na região Sudeste do país (Figura 10).

Os navegadores portugueses chegaram ao Brasil pela primeira vez no ano de 1500 e, desde os primeiros anos de colonização, o território tornou-se um forte fornecedor de matérias-primas para o comércio europeu. No Século XVII, o Rio de Janeiro tornou-se um dos portos marítimos mais importantes de exportação do país, após a descoberta de ouro e diamantes em Minas Gerais, a capitania vizinha. Por esta altura, com o aumento do comércio e a constante necessidade de trabalho operário, tornou-se comum a prática de escravatura, primeiro direcionada para os nativo-americanos e mais tarde para os africanos (JAFFARY, 2016; LOCKARD, 2014).

Figura 10 - Mapa do Brasil, destacando as 5 regiões geográficas



Antes do período de colonização, cerca de 2,4 milhões de pessoas habitavam o território que hoje abrange o Brasil, distribuídos por dois principais grupos linguísticos: Tupi e Tapuias (este último refere-se a grupos não Tupi e inclui majoritariamente falantes do macro grupo Je). O grupo Tupi está incluído na família linguística Tupi-Guarani, que compreende mais de 40 grupos distribuídos pela América do Sul. Atualmente, os Tupi-Guaraní habitam também regiões da Bolívia, Peru, Argentina, Paraguai, Guiana Francesa, Venezuela e Colômbia. À chegada dos colonizadores europeus, os indivíduos Tupi que ocupavam o território brasileiro estavam distribuídos, majoritariamente, por dois principais grupos: os Mundurucú, que habitavam a região Centro-Oeste, no atual estado de Mato Grosso; e os Tupinambás, que habitavam as áreas da costa, desde o sul (atual região do Rio de Janeiro), passando pela Bahia, até ao Norte (estado do Maranhão) (MEADE, 2009). Após os 300 anos de colonização portuguesa, o número de nativos no Brasil reduziu drasticamente para cerca de 1 milhão, devido a fenômenos como doenças epidêmicas (JAFFARY, 2016; The Cambridge History of Latin America, 2008).

Entre 1820 (ano da independência do Brasil) e 1970, estima-se que cerca de 5,6 milhões de migrantes estrangeiros terão chegado ao Brasil (COHEN, 2010). A nacionalidade portuguesa sempre predominou no Rio de Janeiro, assim como na maioria das regiões do Brasil. Para além da primeira vaga de migrantes portugueses durante o século XVI, um grande número também viajou para o Rio de Janeiro, no início do século XX.

A ancestralidade africana da população brasileira atual resulta, na sua maioria, do período de escravatura, em que um elevado número de escravos africanos foi enviado para o Brasil (ARAUJO, 2015; BETHELL, 2011), transportados, principalmente, das regiões Centro-Oeste, Sudoeste e Sudeste de África. Entre o final do século XVII e meados do século XVIII, o porto do Rio de Janeiro foi um dos maiores importadores de escravos para o Brasil, com grande número de africanos vindos das regiões Centro-Oeste e Sudoeste de África (KLEIN; LUNA, 2009) (Figura 11).

Figura 11- Principais portos de tráfico de escravos em África



Fonte: adaptado de <http://www.slavevoyages.org/tat/assessment/intro-maps.faces>

Vários estudos têm sido desenvolvidos na tentativa de clarificar a história das populações brasileiras, com a descrição dos padrões de variação da ancestralidade genética ao longo do país.

Os estudos com base na análise de marcadores autossômicos revelaram uma prevalência de ancestralidade europeia ao longo de todo o território brasileiro, com valores máximos no Sul (LINS et al., 2011; PENA et al., 2011; SALOUM DE NEVES MANTA et al., 2013; SANTOS et al., 2016, 2010). O mesmo foi observado para marcadores localizados no cromossomo Y que descrevem a ancestralidade paterna (DE F. FIGUEIREDO et al., 2016; OLIVEIRA et al., 2014; RESQUE et al., 2016). A ancestralidade materna europeia foi também predominante no Sul. No entanto, nas outras regiões geográficas do Brasil, as proporções de ancestralidade materna africana e nativa foram superiores (ALVES-SILVA et al., 2000; BARBOSA et al., 2008; BERNARDO et al., 2014; FRIDMAN et al., 2014; PALENCIA et al., 2010; PRIETO et al., 2011; SANCHES et al., 2014).

No que diz respeito ao caso particular do Rio de Janeiro, alguns trabalhos focaram-se no estudo de marcadores do cromossomo Y e autossômicos. Em ambos os

casos se observou uma elevada proporção de ancestralidade europeia (ANA et al., 2017; JANNUZZI et al., 2018; OLIVEIRA et al., 2014).

Outros estudos focaram-se no estudo da ancestralidade materna do Rio de Janeiro. Em 2007, Bernardo e colaboradores (BERNARDO et al., 2014) analisaram as regiões HVSI e HVSII do mtDNA, numa amostra de 190 indivíduos do Rio de Janeiro. No seu estudo, obtiveram uma elevada diversidade haplotípica ( $0,9994 \pm 0,0006$ ), onde a maior porcentagem de haplótipos pertence a haplogrupos africanos (57,9 %). Fridman *et al.* (2014) (FRIDMAN et al., 2014) analisaram a região controle do mtDNA de um grupo de amostras do Rio de Janeiro, em conjunto com amostras de outros estados do Sudeste brasileiro. A porcentagem de linhagens maternas nativo-americanas e africanas, encontradas na região Sudeste, foi muito semelhante (36,9 % e 35,2 %, respectivamente).

### Paraguai

Os primeiros ocupantes do território do Paraguai terão sido indígenas caçadores-recoletores. Posteriormente, indivíduos pertencentes à família linguística Tupi-Guarani chegaram à região e, apresentando um estilo de vida menos nômade, estabeleceram-se no território. Antes da chegada dos colonizadores espanhóis no século XVI, a região Este do Paraguai era ocupada por grupos Guarani, enquanto a região Oeste compreendia, pelo menos, 5 grupos linguísticos diferentes (JERMYN; LIN, 2010).

Com a chegada dos colonizadores europeus ao território, foi criada a primeira cidade espanhola, Assunção (ano de 1537) (HEBBLETHWAITE, 2014; LAMBERT, 2012). No entanto, a inexistência de riqueza mineral rapidamente diminuiu o interesse dos espanhóis no território e o Paraguai tornou-se numa região isolada, geograficamente, culturalmente e politicamente separada das outras forças econômicas do período colonial na América do Sul. Os poucos colonizadores que permaneceram no território implementaram o uso de trabalho forçado em nativos e africanos (HEBBLETHWAITE, 2014).

Durante a “Guerra da Tríplice Aliança”, lutada pelo Paraguai contra Argentina, Brasil e Uruguai, entre os anos de 1865 e 1870, o país sofreu uma redução no efetivo populacional para menos de metade (WARREN, 1985). Com o final da guerra, a imigração para o território foi encorajada, resultando na entrada de um grande número

de indivíduos Euroasiáticos (maioritariamente Italianos, Alemães, Franceses, Espanhóis, Coreanos e Japoneses). Dados estatísticos indicam que, atualmente, mais de 90 % dos imigrantes no Paraguai chegam dos países vizinhos Brasil e Argentina (GLOBAL; II, 2004).

O Paraguai é separado pelo rio Paraná em duas regiões geográficas: região Este ou oriental (também conhecida como *Paranéña*) e a região Oeste ou ocidental (também conhecida como *Chaco*) (HEBBLETHWAITE, 2014). O país compreende 17 departamentos, 14 dos quais estão localizados na região Este, onde 98 % da população vive. Pelos eventos históricos descritos anteriormente, a população do Paraguai apresenta uma composição étnica peculiar, diferente de outros países da América do Sul.

São poucos os estudos publicados que descrevem a composição genética de populações miscigenadas ou nativas do Paraguai (CATANESI et al., 2007; E et al., 2003; RIBEIRO et al., 2018; VULLO et al., 2016).

Relativamente à ancestralidade materna do Paraguai, dois estudos foram publicados incluindo apenas a análise da região HVSI do mtDNA em grupos nativos (DORNELLES et al., 2004; SCHMITT et al., 2004). Dorneles *et al.* (DORNELLES et al., 2004) sequenciaram a região HVSI (entre as posições 16024 e 16384), em 34 nativos Ayoreo da região do Chaco, nos quais aproximadamente 80 % das amostras pertencem ao haplogrupo C e as restantes ao haplogrupo D. Schmitt *et al.* (SCHMITT et al., 2004) estudaram o segmento HVSI em 64 indivíduos pertencentes ao grupo étnico Aché (família linguística Tupi-Guarani) do Este do Paraguai. Em contraste com o que foi observado para a região Oeste, todos os haplótipos pertencem aos haplogrupos B (90 %) e A (10 %).

## Peru

O Peru estende-se ao longo da costa Pacífica Sul do continente americano e está limitado a Norte pelo Equador e pela Colômbia, a Este pelo Brasil e Bolívia e a Sul pelo Chile. A região costeira apresenta a maior densidade demográfica (HUNEFELDT, 2010). Assim, grande parte da população indígena ocupa regiões de planalto ao longo da costa. No entanto, uma porção menor está fixada na floresta tropical, distribuída por

quatro principais grupos linguísticos: Arawak, Carib, Tupi e Je. Pensa-se que os indivíduos da família linguística Arawak terão sido responsáveis por uma das principais diásporas na América do Sul no período pré-colonial e, à chegada dos colonizadores, encontravam-se dispersos por diversas regiões do subcontinente (Figura 12). Existem diferentes hipóteses para a origem e posterior dispersão deste grupo linguístico na América do Sul. Dados arqueológicos, linguísticos e históricos indicam que esta família linguística terá tido origem no norte da Amazônia, entre o Rio Negro e o Orinoco, com posterior dispersão ao longo destes (HILL; SANTOS-GRANERO, 2006). No entanto, mais recentemente tem sido especulado que a sua origem poderá estar na região Ocidental da Amazônia com posterior dispersão para o Caribe (WALKER; RIBEIRO, 2011). Os Ashaninka, pertencentes à família linguística Arawak, são o maior grupo indígena amazônico do Peru e estão distribuídos por diversos departamentos: Ayacucho, Cusco, Huanuco, Junin, Pasco e Lima. Segundo o censo peruano de 2009, cerca de 88.000 Ashaninkas viviam no país à data (HUNEFELDT; HARRIS, 2004; MCEWAN, 2006).

Logo após a conquista do território peruano por colonizadores espanhóis, em 1532, deu-se início a um processo de mistura étnica, ocorrendo principalmente entre homens europeus e mulheres indígenas. Como aconteceu em outros territórios, na época colonial, um grande número de escravos africanos foi levado para o Peru (PAERREGAARD, 2008). Após a independência em 1821, o governo peruano estimulou a migração para o território, com a implementação de regulamentações para ocupação de áreas rurais, resultando na chegada de um número considerável de imigrantes vindos de Itália, Portugal, Alemanha, Bélgica e Croácia. Com o final da escravatura, em 1854, e o aumento da necessidade de mão de obra, centenas de milhares de chineses e japoneses chegaram também ao Peru, principalmente entre a segunda metade do século XIX e início do século XX (PAERREGAARD, 2008).

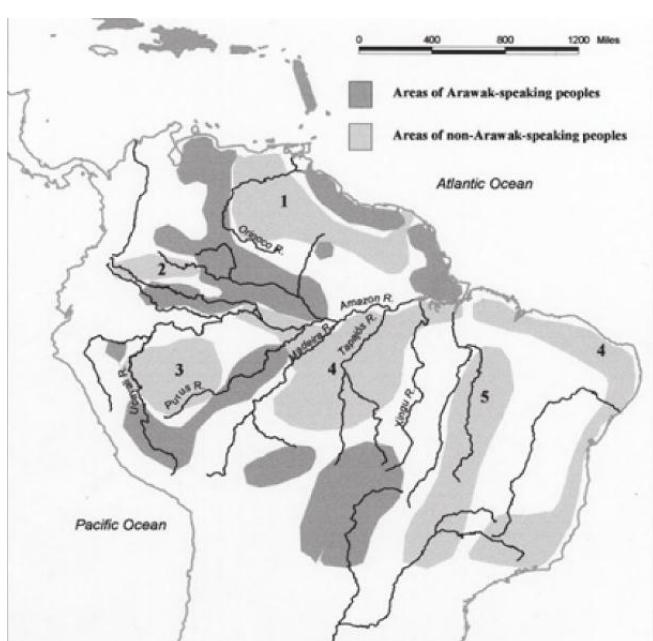
Nos dias de hoje, o Peru é uma sociedade multiétnica com mistura de várias culturas, em especial espanhola, africana e indígena.

Vários estudos têm sido publicados com o objetivo de revelar a composição genética de populações miscigenadas e nativas da região andina do Peru, com recurso a marcadores autossômicos (por exemplo (HOMBURGER et al., 2015; MESSINA et al., 2018)), marcadores da região não recombinante do cromossomo Y (por exemplo (JOTA et al., 2016; SANDOVAL et al., 2016)) e da molécula do mtDNA (por exemplo (BRANDINI et al., 2018; MESSINA et al., 2018; SANDOVAL et al., 2016)). No

entanto, pela sua localização e isolamento geográfico, os grupos indígenas amazônicos encontram-se pouco representados em análises populacionais.

Os poucos estudos com foco nas linhagens maternas de populações amazônicas incluem apenas a análise de alguns segmentos da região controle do mtDNA e/ou de um número reduzido de indivíduos (BARBIERI et al., 2014; DI CORCIA et al., 2017).

Figura 12 - Distribuição dos principais grupos Arawak (cinza escuro), à chegada dos colonizadores europeus



Legenda: Áreas ocupadas por indivíduos da família linguística Arawak na América do Sul (cinza escuro) aquando da chegada Europeia. Áreas ocupadas por indivíduos de outras famílias linguísticas na América do Sul (cinza claro) aquando da chegada Europeia.

Fonte: (HILL; SANTOS-GRANERO, 2006)

## 1 OBJETIVOS

### 1.1 Objetivos gerais

O presente trabalho de doutorado teve como finalidade estudar a ancestralidade materna de populações da América do Sul, através da análise de polimorfismos localizados na molécula de mtDNA. Por um lado, pretendeu-se contribuir para o desenvolvimento de bases de dados de sequências de mtDNA de elevada qualidade, representativas das populações do Rio de Janeiro, Paraguai e Ashaninka do Peru, que possam ser utilizadas para fins forenses.

Por outro lado, com a descrição detalhada das linhagens maternas, pretendeu-se perceber de que forma os diferentes padrões migratórios terão influenciado a composição genética das populações atuais da América do Sul, através do estudo de três populações com histórias muito diversas.

### 1.2 Objetivos específicos

Por forma a cumprir o objetivo geral proposto, foram determinados os seguintes objetivos específicos:

- a) Caracterizar o patrimônio genético de um grupo populacional de indivíduos não relacionados entre si, residentes na cidade do Rio de Janeiro, quanto à sua ancestralidade materna, através do sequenciamento da região controle do mtDNA, pela metodologia de sequenciamento de Sanger;
- b) Elaborar uma base de dados representativa da região e interpretar a origem das linhagens maternas obtidas. Com base no patrimônio genético materno comprovadamente africano, europeu e nativo-americano comparar as amostras brasileiras com aquelas já descritas na literatura, através da análise das distâncias genéticas entre pares de populações;

- c) Caracterizar o patrimônio genético de indivíduos residentes no Paraguai quanto à sua ancestralidade materna, através do sequenciamento da região controle do mtDNA, pela metodologia de sequenciamento de Sanger;
- d) Com base no patrimônio genético materno comprovadamente africano, europeu e nativo-americano, comparar as amostras paraguaias com aquelas já descritas na literatura, através da análise das distâncias genéticas entre pares de populações;
- e) Caracterizar o patrimônio genético de origem materna de indivíduos pertencentes ao grupo linguístico Ashaninka do Peru, através do sequenciamento da região controle do mtDNA, pela metodologia de sequenciamento de Sanger;
- f) Com base no patrimônio genético materno, comparar as amostras peruanas com aquelas já descritas na literatura, através da análise das distâncias genéticas entre pares de populações;
- g) Caracterizar o patrimônio genético de indivíduos residentes no Alto Paraná, Paraguai, quanto à sua ancestralidade materna, através do sequenciamento de mitogenomas, pela metodologia de sequenciamento paralelo massivo. Com base no patrimônio genético materno comprovadamente africano, europeu e nativo-americano, comparar as amostras paraguaias com aquelas já descritas na literatura, através da análise das distâncias genéticas entre pares de populações;
- h) Caracterizar haplótipos pertencentes ao haplogrupo B2+16051G+16129A+152C da população Ashaninka, através de metodologias de sequenciamento paralelo massivo. Comparar os polimorfismos característicos das linhagens B2 com os descritos para outras populações Arawak e investigar a sua possível relação;
- i) Comparar os resultados obtidos nas 3 populações analisadas e fazer a sua interpretação à luz dos conhecimentos sobre a história destas populações.

## 2 METODOLOGIA

### 2.1 Seleção das amostras

Em todas as coletas realizadas, os participantes assinaram consentimento individual informado, concordando com a sua participação nos estudos, sob condições estritamente confidenciais.

#### 2.1.1 Rio de Janeiro

Foram coletadas manchas de sangue de 205 indivíduos não relacionados entre si, residentes na região metropolitana do Rio de Janeiro (Figura 13). A coleta foi realizada no Laboratório de diagnóstico por DNA da Universidade do Estado do Rio de Janeiro, a voluntários que se apresentaram no laboratório, após divulgação do projeto.

O estudo destas amostras foi aprovado pelo Comité de ética da Universidade do Estado do Rio de Janeiro (CAAE: 34301714.7.0000.5259).

O DNA foi extraído a partir de manchas de sangue depositadas em papel FTA®, utilizando o resina quelante Chelex® 100 (BioRad) a 5%, de acordo com a técnica desenvolvida por Walsh e colaboradores, em 1991 (WALSH; METZGER; HIGUCHI, 1991). O produto extraído manteve-se armazenado a 4 °C.

Estas amostras foram previamente estudadas para um conjunto de 46 marcadores autossômicos informativos de ancestralidade (AIMs, do inglês *Ancestry Informative Markers*) (PEREIRA et al., 2012).

#### 2.1.2 Paraguai

Um total de 299 manchas de sangue foi coletado de indivíduos do sexo masculino residentes em 13 departamentos do Paraguai: Alto Paraná (n = 36), Caaguazú

(n = 15), Caazapá (n = 6), Canindeyú (n = 1), Central (n = 74), Concepción (n = 3), Cordillera (n = 13), Guairá (n = 95), Itapúa (n = 14), Misiones (n = 5), Ñeembucú (n = 1), Paraguarí (n = 34), San Pedro (n = 1), sem informação (n = 1) (Figura 13 e Figura 14).

O estudo destas amostras foi aprovado pelo comité de ética em investigação clínica do Instituto de previão social, Assunção, Paraguai. A coleta foi realizada nos respectivos departamentos, em colaboração com o *Instituto de Previsión Social, Asunción, Paraguai*, a *Legislación Farmacéutica, Universidad Católica, Hernandarias, Paraguai* e o Hospital Regional, Encarnación, Paraguai.

O DNA foi extraído a partir de manchas de sangue depositadas em papel FTA®, utilizando o reagente Chelex® 100 (BioRad) a 5%, pela técnica descrita por Walsh e colaboradores, em 1991 (Walsh *et al.* 1991). O produto extraído manteve-se armazenado a 4 °C.

Este grupo de amostras faz parte de uma colaboração do Laboratório de Diagnóstico por DNA da UERJ com o EAAF (*Equipo Argentino de Antropología Forense*) em Córdoba, Argentina.

### **2.1.3 Peru**

Um total de 171 amostras de sangue foi coletado de indivíduos não relacionados entre si, pertencentes a 41 comunidades localizadas nas margens dos Rios Pichis e Palcazú, no distrito de Puerto Bermúdez, região de Pasco, Peru (Figura 13).

O estudo destas amostras foi aprovado pela *Asociación de Comunidades Ashaninka del Valle Pichis*, Distrito de Puerto Bermúdez, Provincia de Oxampampa, Departamento de Pasco – Republica do Perú (Registro Público nº 11006635)

O DNA foi extraído a partir de manchas de sangue depositadas em papel de filtro, através de um protocolo padrão de fenol-clorofórmio (Anexo A).

As amostras do grupo linguístico Ashaninka foram obtidas através de uma colaboração com Dean Herman Tineo, aluno de mestrado da *Universidad Nacional Mayor de San Marcos*, Lima, Peru, o qual procedeu à coleta das mesmas nas regiões referidas anteriormente.

Figura 13 - Locais de coleta das amostras

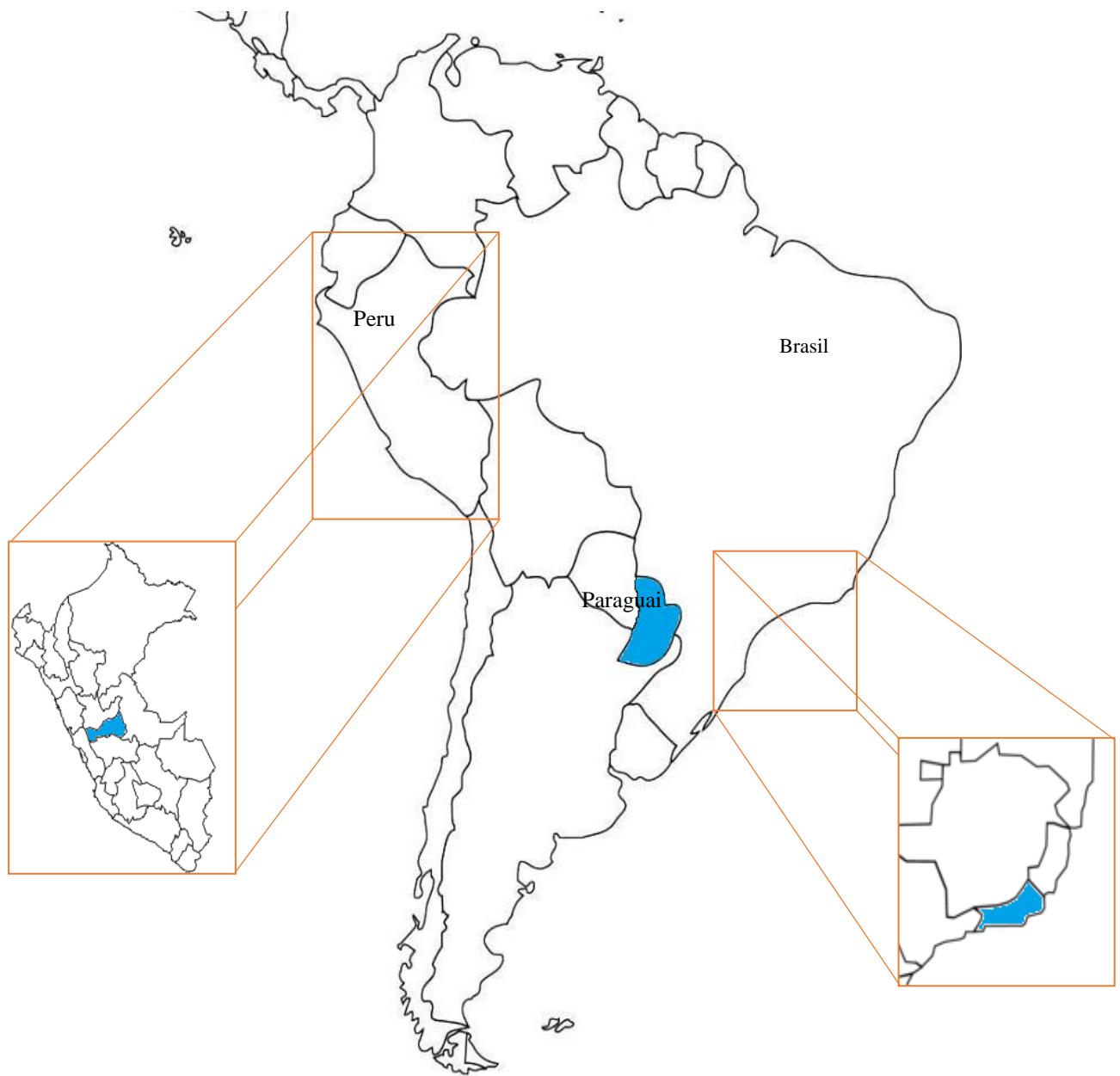
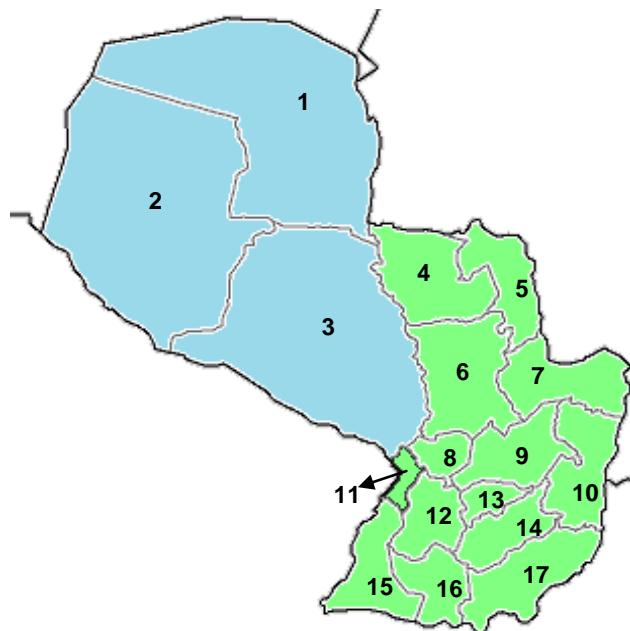


Figura 14 - Mapa dos departamentos do Paraguai



Legenda: A azul (números 1 a 3) estão representadas as províncias Oeste do Paraguai, pertencentes à região Gran Chaco; a verde (números 4 a 17) as províncias a Este do rio Paraguai, das quais foram coletadas amostras para o presente estudo. 1 - Alto Paraguay, 2 - Boquerón, 3 - Presidente Hayes, 4 - Concepción, 5 - Amambay, 6 - San Pedro, 7 - Canindeyú, 8 - Cordillera, 9 - Caaguazú, 10 - Alto Paraná, 11 - Central, 12 - Paraguari, 13 - Guairá, 14 - Caazapá, 15 - Neembucú, 16 - Misiones, 17 - Itapúa.

## 2.2 Sequenciamento do mtDNA

### 2.2.1 Sequenciamento de Sanger

#### 2.2.1.1 Rio de Janeiro

A região controle do mtDNA (entre as posições nucleotídicas 16024 e 576) foi amplificada utilizando o reagente *Qiagen multiplex PCR mix* (Qiagen). Para um volume de reação de 5 µl foram utilizados 2,5 µl de *kit* Qiagen, 0,5 µl da mistura de *primers* direto e reverso, 0,5 µl do produto extraído e 1,5 µl de H<sub>2</sub>O. Os *primers* utilizados foram o L15997 (5'- CAC CAT TAG CAC CCA AAG CT -3') e H639 (5'- GGG TGA TGT GAG CCC GTC TA -3'), a uma concentração de 0,2 µM cada. A reação de PCR

iniciou-se com um passo de desnaturação a 95 °C durante 10 min, seguida de 35 ciclos de 94 °C durante 30 s, 60 °C por 90 s, e 72 °C durante 60 s, e finalizou-se com um passo de extensão de 72 °C durante 10 min. A purificação do produto amplificado foi executada com as enzimas exonuclease I e fosfatase alcalina de camarão, combinadas no reagente ExoSap-IT (GE Healthcare). Por cada 5 µl de produto amplificado foram utilizados 2 µl de ExoSap-IT. A ativação das enzimas foi efetuada a 37 °C por 15 min, seguida da inativação a 80 °C por 15 min. No sequenciamento foi utilizado o kit *Big Dye Terminator* v3.1 (Applied Biosystems), juntamente com os *primers* diretos L15997 e L16555 (5'- CCC ACA CGT TCC CCT TAA AT -3') e os *primers* reversos H016 (5'- CCC GTG AGT GGT TAA TAG GGT - 3') e H639. Para um volume final de 5 µl foram utilizados 1 µl de BigDye, 1 µl de Buffer, 0,5 µl de *primer* a 2,5 µM, 2 µl de H<sub>2</sub>O e 0,5 µl do produto purificado. As condições de sequenciamento foram as seguintes: desnaturação inicial a 92 °C por 2 min, um total de 35 ciclos com temperaturas de 96 °C por 15 s, 50 °C por 9 s, e 60 °C por 2 min, terminando com uma extensão final a 60 °C durante 10 min. A purificação final, para remoção dos ddNTPs livres, foi realizada através da passagem do produto de sequenciamento por colunas de Sephadex (Illustra Sephadex DNA Grade; GE Healthcare). A separação e detecção das sequências foram realizadas por eletroforese capilar no aparelho ABI 3500 (Applied Biosystems).

Seguindo as recomendações de Parson e Bandelt (PARSON; BANDELT, 2007), em casos onde a dupla cobertura não foi atingida no sequenciamento com os primers referidos anteriormente, foi realizada uma segunda amplificação com os primers L15900 (5'-TAAACTAATACACCAGTCTTGTAACCC-3') e H599 (5'-TTGAGGAGGTAAGCTACATA-3'). O sequenciamento foi posteriormente efetuado com os primers L15900, L16268 (5'-CACTAGGATACCAACAAACC-3'), H016 e H599, seguindo as condições previamente descritas. Nas amostras em que foi observada heteroplasmia de comprimento associada a inserções AC entre as posições 513 e 525 e/ou pela transição na posição 460, o sequenciamento foi também obtido com os primers L314 (5'-CCGCTTCTGGCCACAGCACT-3'), H338 (5'-GTTTAAGTGTGTGCCAGAAG-3') e H484 (5'-TGAGATTAGTAGTATGGGAG-3').

As sequências de mtDNA foram comparadas com a rCRS, através do software SeqScape v2.7. Na classificação dos haplótipos seguiram-se as diretrizes propostas pela ISFG (PARSON et al., 2014a).

### 2.2.1.2 Peru e Paraguai

A região controle do mtDNA (entre as posições nucleotídicas 16.024 e 576) foi amplificada numa reação de 25 µl, com os seguintes componentes: 15,8 µl de H<sub>2</sub>O, 2,5 µl de tampão 10x, 2,5 µl de BSA a 2,5 mg/ml, 2 µl de dNTPs a 10 mM, 0,3 µl do *primer* direto L15851 (5' ATC TCC CTA ATT GAA AAC AAA ATA CTC AAA 3') a 10 µM, 0,3 µl do *primer* reverso H639 (5'- GGG TGA TGT GAG CCC GTC TA -3') a 10 µM, 0,2 µl da polimerase Advantage (Clotech) a 5U/µl e 1 µl de produto extraído. A reação de PCR iniciou-se com um passo de desnaturação a 95 °C durante 2 min, seguida de 35 ciclos de 95 °C durante 15 s, 56 °C por 30 s, e 72 °C durante 90 s, e finalizou-se com um passo de extensão de 72 °C durante 10 min. A purificação do produto amplificado foi executada com as enzimas exonuclease I e fosfatase alcalina de camarão, combinadas no reagente ExoSap-IT (GE Healthcare). Por cada 5 µl de produto amplificado foram utilizados 2 µl de ExoSap-IT. A ativação das enzimas foi efetuada a 37 °C por 15 min, seguida da inativação a 80 °C por 15 min. O sequenciamento foi efetuado com o kit de *Big Dye Terminator* v1.1 (AppliedBiosystems), juntamente com os *primers* L15971, L16268 (5' CAC TAG GAT ACC AAC AAA CC 3'), L15 (5' CAC CCT ATT AAC CAC TCA CG 3'), H16 (5' TGA TAG ACC TGT GAT CCA TCG TGA 3'), L314 (5' CCG CTT CTG GCC ACA GCA CT 3') e H639. Para um volume final de 10 µl foram utilizados 2 µl de BigDye, 2 µl de Buffer, 0,3 µl de *primer* a 10 µM, 3,7 µl de H<sub>2</sub>O e 2 µl do produto purificado. As condições de sequenciamento foram: desnaturação inicial a 96 °C por 1 min, um total de 25 ciclos com temperaturas de 95 °C por 15 s, 50 °C por 5 s, e 60 °C por 4 min. A purificação final, para remoção dos ddNTPs livres, foi realizada através da passagem do produto de sequenciamento por colunas de Sephadex (IllustraSephadex DNA Grade; GE Healthcare). A separação e detecção das sequências foram realizadas por eletroforese capilar no aparelho ABI 3500 (AppliedBiosystems).

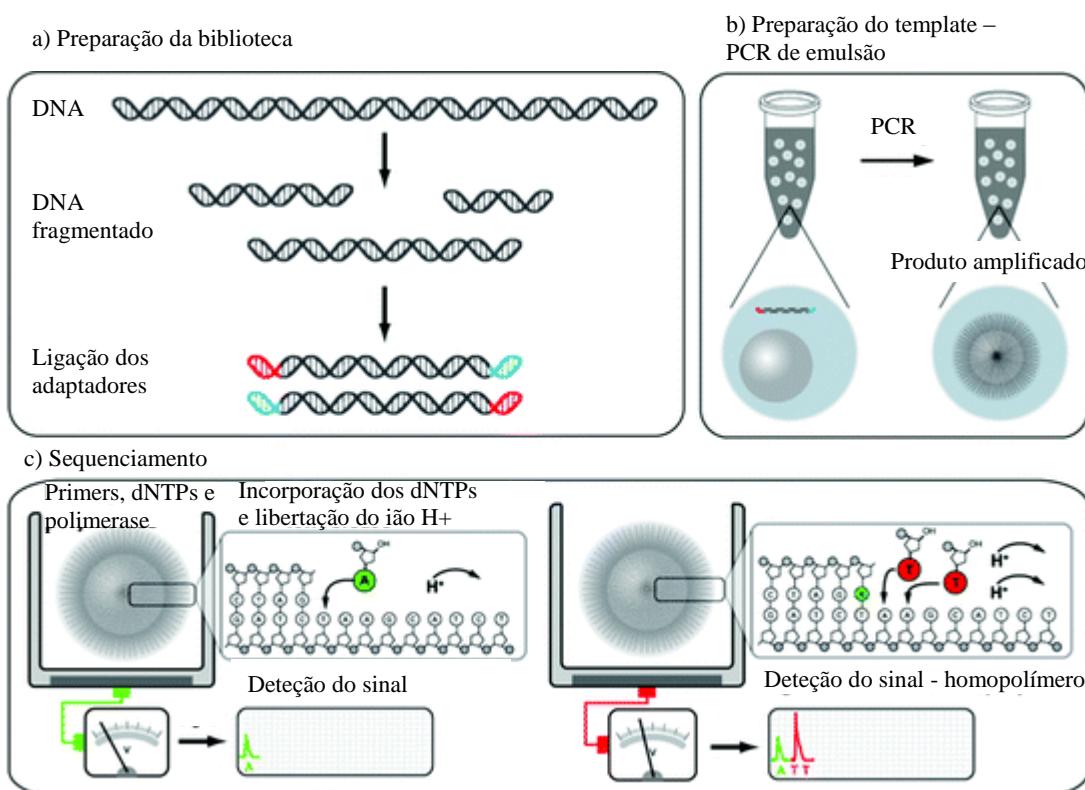
As sequências de mtDNA obtidas foram comparadas com a rCRS, através do software Sequencher v.4.9 (Gene Codes). Para a classificação dos haplótipos, seguiram-se as diretrizes propostas pela ISFG (PARSON et al., 2014a).

## 2.2.2 Sequenciamento massivo paralelo

A preparação da biblioteca para os mitogenomas foi realizada com os *kits* *Precision ID mtDNA Whole Genome Panel* e *Ampliseq Library*. Na amplificação, foram utilizados dois *pools* com 81 pares de *primers* cada, que geraram um total de 162 amplicões, com um tamanho médio de 163 pbs e uma sobreposição entre amplicões de 11 pbs (Figura 15a). Usou-se o método conservativo de amplificação, no qual os dois *pools* de *primers* foram amplificados em reações separadas de 10 µl, que foram posteriormente misturadas (volume final de 20 µl). As regiões de interesse foram amplificadas de acordo com as seguintes condições: ativação enzimática a 99 °C por 2 min, seguida de 18 ciclos de desnaturação a 99 °C por 15 s, *annealing* e extensão a 60 °C por 4 min (Figura 15b). Após digestão de *primers* com o reagente FuPA, os adaptadores IonCode foram ligados a cada amostra (Figura 15a). Para limpeza da reação e remoção de dímeros dos adaptadores, foi utilizado o reagente AMPure XP. A quantificação foi realizada com o kit Ion Library Taq Man Quantitation kit (Thermo Fisher Scientific). Para balancear a concentração, as amostras foram diluídas a uma concentração de 30 pM. Os chips foram carregados automaticamente no instrumento Ion Chef (Thermo Fisher Scientific) e as amostras sequenciadas no instrumento Ion S5 (Thermo Fisher Scientific) (Figura 15c). Neste procedimento, a sequenciação é realizada por síntese, onde a cadeia complementar é construída com base numa sequência *template*. O chip contendo os fragmentos de mtDNA a sequenciar é “inundado” com uma única espécie de dNTPs de cada vez (Figura 15c). Caso o dNTP seja complementar, a incorporação do nucleotídeo irá libertar iões de hidrogênio, detectados por um sensor (Figura 15c). Na presença de homopolímeros, a incorporação de mais de um dNTP leva a um maior número de iões de hidrogênio libertados e aumento proporcional da intensidade de sinal (Figura 15d).

As sequências foram comparadas com a sequência de referência de Cambridge revista (rCRS) (ANDREWS et al., 1999), no software IGV v 2.4.10 (ROBINSON et al., 2011; THORVALDSDÓTTIR; ROBINSON; MESIROV, 2013), de acordo com as directrizes da ISFG para a classificação de haplótipos de mtDNA (PARSON et al., 2014a).

Figura 15 - Etapas do sequenciamento massivo paralelo no aparelho Ion S5



## 2.3 Classificação de haplogrupos e análise estatística

As ferramentas “Network” e “EMPcheck” da EMPOP foram usadas para detectar potenciais erros nos haplótipos obtidos (PARSON; DÜR, 2007; ZIMMERMANN et al., 2014).

Os haplogrupos foram classificados com a ajuda da base de dados EMPOP database v3/R11 (PARSON; DÜR, 2007) e confirmados manualmente na Phylotree, build 17, atualizada em Fevereiro de 2016 (VAN OVEN; KAYSER, 2009).

O alinhamento das sequências foi realizado no Haplosearch (FREGEL; DELGADO, 2011), com posterior correção de alinhamento da EMPOP (PARSON; DÜR, 2007).

As frequências dos haplogrupos na população foram calculadas por contagem direta. Os parâmetros estatísticos de diversidade haplotípica, número médio de diferenças par-a-par e as distâncias genéticas ( $Fst$ ) foram calculados utilizando o software Arlequin v. 3.5.1.2 (EXCOFFIER; LAVAL; SCHNEIDER, 2005). As árvores filogenéticas foram desenhadas com os métodos *reduced median* e *median joining*,

implementados no software Network v5.0 (<http://www.fluxusengineering.com>). As distâncias genéticas par-a-par foram visualizada em três dimensões utilizando a análise de escalonamento multidimensional (MDS) incluída no software StatSoft, Inc. (2007) STATISTICA, ver.8.0 ([www.statsoft.com](http://www.statsoft.com)).

As árvores de máxima parcimônia filogenética de haplogrupos foram obtidas com o software mtPhyl (<https://sites.google.com/site/mtphyl/home>) e corrigidas manualmente com base na *phylotree build* 17 (VAN OVEN; KAYSER, 2009).

### 3 RESULTADOS E DISCUSSÃO

Os dados obtidos nesta tese foram divulgados ao longo do doutorado em congressos internacionais (apresentações orais e pôster) e publicados na forma de artigos científicos em revistas internacionais. Os resultados alcançados com o estudo das linhagens maternas das populações miscigenadas do Rio de Janeiro (Brasil) e Paraguai e população nativa Ashaninka (Peru), são apresentados em três partes, de acordo com os objetivos inicialmente propostos.

#### 3.1 Caracterização do patrimônio genético de origem materna de indivíduos residentes na cidade do Rio de Janeiro, Brasil, através da análise do mtDNA

Os resultados preliminares obtidos da análise de linhagens maternas de indivíduos residentes na cidade do Rio de Janeiro foram divulgados em apresentações orais no congresso internacional *Haplod markers 2016 – Update on DNA variation* (Berlim, Alemanha, Maio 2016) e no IV Simpósio Internacional de Identificação Humana por DNA (Rio de Janeiro, Brasil, Abril 2017).

Na análise de associação entre haplogrupos de mtDNA e marcadores autossômicos informativos de ancestralidade, foi utilizada uma base de dados para a população do Rio de Janeiro, atualizada em 2017 e publicada no *Forensic Science International: Genetics Supplement Series* com o título “*Contrasting admixture estimates in Rio de Janeiro obtained by different sampling strategies*” (Apêndice A). Das 1.042 amostras genotipadas para 46 AIM-Indels, 204 coincidem com as amostras sequenciadas para o mtDNA.

Os resultados finais da caracterização do patrimônio genético de origem materna de indivíduos residentes na cidade do Rio de Janeiro foram publicados na revista *Forensic Science International Genetics*, no artigo que aqui se apresenta com o título “*Defining mtDNA origins and population stratification in Rio de Janeiro*” (Artigo 1). O material suplementar do artigo está disposto nesta tese no Apêndice B.

## Artigo 1

Título traduzido: **Definindo as origens do mtDNA e estratificação populacional no Rio de Janeiro**

Resumo traduzido:

A composição genética da população brasileira foi formada por misturas interétnicas entre nativo-americanos, colonizadores europeus e escravos africanos. Esta estrutura, característica de grande parte das populações americanas, aponta a necessidade da implementação de bancos de dados forenses que permitam capturar a alta diversidade, geralmente associada a populações miscigenadas.

As 205 sequências da região controle do mtDNA apontaram para uma elevada diversidade haplotípica ( $0,9994 \pm 0,0006$ ) na população do Rio de Janeiro. Comparações par-a-par entre as sequências, revelaram uma elevada proporção de pares de haplótipos que apresentavam mais que uma diferença entre si. Quando se ignoraram os tratos homopoliméricos (inserções nas posições 309, 315, 573 e 16189 do genoma mitocondrial), as comparações par-a-par revelaram sequências iguais em 0,18 % das vezes, e diferenças numa única posição em 0,32 % das comparações. Obteve-se uma elevada porcentagem de linhagens maternas africanas (42%), cuja origem maioritária incide na região Sudoeste de África. Para os haplogrupos euroasiáticos e nativo-americanos (representando 32% e 26% das amostras, respectivamente), não foi possível avaliar uma origem geográfica clara ou uma afiliação linguística. Ao se agrupar os haplogrupos de mtDNA de acordo com a sua origem continental (nativo-americana, europeia e africana), foram observadas diferenças para as proporções de ancestralidade estimadas com os AIM-Indels, o que sugere algum nível de subestrutura genética no Rio de Janeiro. Os resultados deste estudo estão de acordo com os relatos históricos relativos ao processo de miscigenação, relatando uma elevada contribuição materna africana e uma contribuição materna nativo-americana significativa. Para além disso, a presença de algum nível de associação entre o mtDNA e informação autossômica deve ser tida em consideração quando se combina estes dois tipos de marcadores em análises forenses.



## Research paper

## Defining mtDNA origins and population stratification in Rio de Janeiro



Filipa Simão<sup>a</sup>, Ana Paula Ferreira<sup>a</sup>, Elizeu Fagundes de Carvalho<sup>a</sup>, Walther Parson<sup>b,c,\*</sup>, Leonor Gusmão<sup>a</sup>

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

<sup>b</sup> Institute of Legal Medicine, Medical University of Innsbruck, Innsbruck, Austria

<sup>c</sup> Forensic Science Program, The Pennsylvania State University, University Park, PA, USA

## ARTICLE INFO

**Keywords:**  
Mitochondrial DNA  
Population genetics  
Haplogroups  
Admixture  
Population  
Forensic genetics

## ABSTRACT

The genetic composition of the Brazilian population was shaped by interethnic admixture between autochthonous Native Americans, European settlers and African slaves. This structure, characteristic of most American populations, implies the need for large population forensic databases to capture the high diversity that is usually associated with admixed populations. In the present work, we sequenced the control region of mitochondrial DNA from 205 non-related individuals living in the Rio de Janeiro metropolitan region. Overall high haplotype diversity ( $0.9994 \pm 0.0006$ ) was observed, and pairwise comparisons showed a high proportion of haplotype pairs with more than one-point differences. When ignoring homopolymeric tracts, pairwise comparisons showed no differences 0.18% of the time, and differences in a single position were found with a frequency of 0.32%. A high percentage of African mtDNA was found (42%), with lineages showing a major South West origin. For the West Eurasian and Native American haplogroups (representing 32% and 26%, respectively) it was not possible to evaluate a clear geographic or linguistic affiliation. When grouping the mtDNA lineages according to their continental origin (Native American, European and African), differences were observed for the ancestry proportions estimated with autosomal ancestry-informative markers, suggesting some level of genetic substructure. The results from this study are in accordance with historical data where admixture processes are confirmed with a strong maternal contribution of African maternal ancestry and a relevant contribution of Native American maternal ancestry. Moreover, the evidence for some degree of association between mtDNA and autosomal information should be considered when combining these types of markers in forensic analysis.

## 1. Introduction

Genetic profiling is commonly used to assist criminal investigations. The study of genetic markers in forensic casework requires the implementation of DNA databases representative of the composition of reference populations. The accuracy of these databases is crucial to correctly estimate the frequency of a genetic profile, which is the only way to attribute statistical weight to the evidence. Additionally, to avoid erroneous estimates of the probability of a certain observation, it is also important to evaluate if population substructure exists, a condition that is frequently observed in American countries as a result of non-random mating among subgroups within populations.

The colonization of America by European settlers led to one of the largest human migrations, characterized by non-uniform movements in terms of number, origin and gender of the migrants. The interethnic crosses resulted in heterogeneous populations that were, in most cases, shaped by three well differentiated continental groups: autochthonous

Native Americans, European settlers and African slaves.

The Portuguese navigators arrived in Brazil in 1500, and since the early days of colonization, the territory became a major supplier of food and minerals to the European commerce. By this time, with the growth of commercial trading and the constant need of handwork, it became common among the colonizers to use slavery work, first targeted at indigenous people and later Africans [1,2]. In the 17th century, Rio de Janeiro became one of the most important seaports for exportation from the country, after the discovery of gold and diamonds in Minas Gerais, the neighbouring captaincy.

Mitochondrial DNA (mtDNA) has been frequently applied in population genetic studies in an attempt to clarify the history of human migrations and disclose stratification in admixed populations. Previous studies evaluated the mtDNA genetic composition of admixed Brazilian populations to unveil their maternal ancestry. In 2000, Alves-Silva and collaborators [3] described a variation throughout the country in the frequency of mtDNA continental fractions, corroborated by further

\* Corresponding author at: Medical University of Innsbruck, Mackenstr. 44, Innsbruck 6020, Austria.  
E-mail address: [walther.parson@med.ac.at](mailto:walther.parson@med.ac.at) (W. Parson).

investigations [4–9]. The highest proportion of Native American ancestry was found in the North region of Brazil [3]. According to the studies of Alves Silva and Barbosa [3,4], in the two Northeast states of Pernambuco and Alagoas, a higher proportion of African lineages is present, even though the second most represented ancestry was different in the two states (European in Pernambuco and Native American in Alagoas). On the other hand, the highest European proportion (over 60%) is present in the South [3,7]. The Southeast is the best-studied region of Brazil, with information available for mtDNA in all four states of São Paulo, Minas Gerais, Rio de Janeiro and Espírito Santo [3,5,6,8,9]. A high percentage of African maternal lineages was found in all states. However, similar proportions of European, African and Native American haplogroups were found in Minas Gerais [3]. No studies have yet been conducted regarding the admixed population from Central West Brazil.

To increase mtDNA discrimination power in Rio de Janeiro forensic research, we characterized the maternal lineages of the region by sequence analysis of the entire control region (CR) since published data only exist for HVS-I and HVS-II [5]. A sample from the Southeast (including Rio de Janeiro) was previously studied for HVS-I, HVS-II and HVS-III [6], but differential analysis was not performed for each state.

In autosomal loci, the presence of inside population substructure can be detected by a statistically significant higher observed heterozygosity than expected. Also, in an admixed population that is structured, associations between alleles at different markers will tend to persist not only for linked loci but even for markers in different chromosomes. However, a non-random population distribution of maternal lineages of very different origins can only be evaluated by combining information from mitochondria with other genomes. Therefore, in this work, data from ancestry informative markers in the same samples (data not published) was used to see if an association exists between mtDNA and autosomal variation, in the current Rio de Janeiro population.

In summary, we aimed to elaborate a genetic database for the mtDNA control region representative of Rio de Janeiro and to evaluate the presence of substructure in the population by ascertaining markers association, in view of a more accurate interpretation of maternal profiles in forensic casework. Additionally, we further analysed the obtained data in the context of the available mtDNA information on historically related populations, with the aim of contributing to a deeper understanding of the history of the South American maternal gene pool.

## 2. Material and methods

### 2.1. Population sample

Blood stains were collected from 205 non-related individuals living in the Rio de Janeiro (Brazil) metropolitan region, and all signed informed consent to participate in the work. This study was approved by the Ethics Committee of the State University of Rio de Janeiro (CAAE: 34301714.7.0000.5259).

### 2.2. mtDNA typing

The control region of mitochondrial DNA (between positions 16024 and 576) was amplified in a final volume of 5 µl, using 2.5 µl of Qiagen® multiplex PCR kit and the primers L15997 (5'-CACCATTAGCACCCAA AGCT-3') and H639 (5'-GGGTGATGTGAGOCCTGCTA-3') at a concentration of 0.2 µM each, with 1–5 ng of DNA. The PCR conditions were an initial denaturation at 95 °C for 10 min, followed by 35 cycles of 94 °C for 30 s, 60 °C for 90 s and 72 °C for 60 s, ending with a final extension at 72 °C for 10 min. The purification was performed with ExoProStar reagent (GE Healthcare). For sequencing, the BigDye Terminator v3.1 cycle Sequencing kit (Applied Biosystems) was used, according to the manufacturer protocol, together with the primers

L15997, L16555 (5'-CCACACCGTTCCCTTAAAT-3'), H016 (5'-COGTTAGTGGTTAACAGGT-3') and H639. The thermocycling conditions were as follows: initial denaturation at 96 °C for 2 min, followed by 35 cycles of 96 °C for 15 s, 50 °C for 9 s and 60 °C for 2 min, and a final extension at 60 °C for 10 min. For final cleaning, the sequences were purified using Sephadex (Illustra Sephadex DNA Grade; GE Healthcare). Separation and detection were achieved by capillary electrophoresis on an ABI 3500 (Applied Biosystems).

As recommended by Parson and Bandelt [10], in cases where double coverage was not possible with the primers indicated above, a second PCR was performed using the primers L15900 (5'-TAAACTAATACAC CAGTCITGTAAACC-3') and H599 (5'-TTGAGGAGGTAAGCTAC ATA-3'), and sequencing was done with primers L15900, L16268 (5'-CACTAGGATAACCAAACAAAC-3'), H016 and H599 under the previously described conditions. For samples with length heteroplasmy due to AC insertions between positions 513 and 525 and/or because of T > C transition in position 460, sequencing was also performed using the primers L314 (5'-CGCTTCTGGCCACAGCACT-3'), R338 (5'-GTTT AAGTGCTGTGCCAGAAG-3') and R484 (5'-TGAGATTAGTAGTATGG GAG-3').

### 2.3. Haplotype assignment and statistical analysis

The mtDNA sequences were compared to the revised Cambridge Reference Sequence (rCRS) [11] with the software SeqScape v2.7, and the haplotype classification was carried out following the guidelines proposed by International Society for Forensic Genetics (ISFG) [12]. Haplogroups were attributed via EMPOP according to PhyloTree build 17, February 2016 [13]. The EMPOP tools "Network" and "EMPcheck" were used to detect potential errors in the dataset [14,15]. Data were submitted to EMPOP for quality control [14] and will be available for forensic searches under the accession number EMP00697.

Haplotype frequencies were calculated by direct counting. The alignment was performed on HaploSearch [16] and adapted with the alignment correction of EMPOP [14]. The statistical parameters of haplotype diversity and genetic distance were calculated using Arlequin software [17]. Pairwise genetic distances were visualized in three-dimensional space using the multidimensional scaling (MDS) analysis included in the StatSoft, Inc. (2007) program STATISTICA (data analysis software system), ver.8.0 ([www.statsoft.com](http://www.statsoft.com)). Phylogenetic networks were designed with reduced median [18] and median-joining methods [19] on Network v5.0 (<http://www.fluxusengineering.com>).

## 3. Results and discussion

The haplotypes alongside corresponding haplogroups can be found in Table S1.

The lineages were classified according to their continental ancestry within African (42%), European (31%), Native American (26%) and Eurasian haplogroups (approximately 1%).

### 3.1. mtDNA diversity

Among the 205 samples analysed, 184 unique mtDNA haplotypes were identified, resulting in a haplotype diversity (H) of  $0.9994 \pm 0.0006$  (Table 1). The mismatch distribution of haplotypes in the Rio de Janeiro sample (Fig. S1A) shows a high number of differences in most pairwise comparisons (mean number of pairwise differences (MNPD) of 18.61) (Table 1). Since two or more differences are required by both the ISFG [20] and the SWGDAM [21] for exclusion purposes, according to these guidelines, the *a priori* chance of excluding non-maternally related individuals (mtCE) is 99.76% in the Rio de Janeiro population (Table 1).

Point heteroplasmy was observed in 8 samples (Table S2). Except in two cases involving positions 237 and 16147, all heteroplasmies occurred at positions considered major hotspots in the mtDNA genome

**Table 1**  
Haplotype diversity ( $H$ ), exclusion power (mtCE) and mean number of pairwise differences (MNPD) values calculated for a sample from Rio de Janeiro and for three subsamples of haplotypes with different continental origins.

Sample	n	Complete haplotypes			Haplotypes without length heteroplasmy polymorphisms *				
		H	mtCE	MNPD	H-mtCE	H	mtCE	MNPD	H-mtCE
Rio de Janeiro	205	0.9904	0.9976	18.613	0.0019	0.9982	0.9950	17.557	0.0032
Native American lineages	53	0.9964	0.9877	17.092	0.0087	0.9840	0.9768	15.725	0.0073
European lineages	66	0.9901	0.9953	10.846	0.0037	0.9977	0.9800	9.841	0.0177
African lineages	86	0.9986	0.9934	19.707	0.0052	0.9970	0.9918	19.072	0.0052

n, number of samples; H-mtCE, frequency of pairwise comparisons showing one difference.

\* Values were calculated after excluding insertions at positions 309, 315, 573 and 16193.

[22] and were detected more than once in a total of 319 observations by Irwin et al. [23]. No samples could be found in the EMPOP database with 237R (search made September 2017). Conversely, a total of 11 haplotypes presented point heteroplasmy at position 16147, all belonging to haplogroup H2a2a1, while our sample is from haplogroup B2b3a.

Length heteroplasmy was found between positions 16183 and 16194 and between positions 302 and 310. Additionally, the insertion of AC in the dimeric repeat between positions 513 and 525 (524.1A 524.2C) was associated with length heteroplasmy in some samples. Length heteroplasmy was also observed in one sample (PRJ168) associated with a T > C transition in position 460. Finally, the insertion of cytosines between positions 567 and 574 resulted in length heteroplasmy in 5 samples.

Since some of these homopolymeric stretches are usually ignored in the interpretation of mtDNA profiles in forensic casework [12], a new mismatch distribution analysis was performed disregarding insertions at positions 309, 315, 573 and 16193 (Fig. S1B). When removing these positions, the frequency of pairwise sequences with one or more differences (haplotype diversity) slightly decreased to 0.9982, and the proportion of pairwise comparisons that resulted in just one difference between haplotypes (H-mtCE) was approximately 30% higher (Table 1).

Differences observed between H and mtCE are expected to correlate with the mismatch distribution profiles of populations. To simulate the expected differences in Native American, European and African populations, mismatch distribution analysis was performed for haplotypes with the same continental origin. In Fig. S1C-H, we can see that MNPD values are lower for the European subset, followed by the Native American and the African subsets (Table 1). In accordance, the African subset showed the lowest difference between H and mtCE (H-mtCE), which is, however, higher than that observed in the admixed sample from Rio de Janeiro.

These results show that, when evaluating the *a priori* efficiency of mtDNA typing in forensic cases, haplotype diversity should be considered together with the haplotype variation profile in populations. Indeed, when excluding insertions and deletions in polycytosine regions, we can see that European/Eurasian have higher diversity than African haplogroups but a lower mtCE.

### 3.2. Differentiation analysis between Brazilian admixed populations

To evaluate if there are significant differences in haplotype frequencies, pairwise  $F_{ST}$  genetic distances were calculated using available data for Brazilian admixed populations with full CR information (Table 2). No statistically significant differences were observed in pairwise comparisons involving Rio de Janeiro, São Paulo [8] and Espírito Santo [9]. Significant differences were observed in pairwise comparisons between Santa Catarina (South region) [7] and all 3 samples from the Southeast (Table 2).

To increase the number of populations in pairwise comparisons, a new analysis was performed using only data from HVS-I, and the  $F_{ST}$  genetic distances are represented in the MDS plot in Fig. 1. The

**Table 2**  
Matrix of pairwise  $F_{ST}$  genetic distances among Rio de Janeiro and 3 admixed Brazilian populations (below diagonal) and the corresponding differentiation p-values (above diagonal) obtained for 10100 permutations for mtDNA control region.

	Rio de Janeiro	Santa Catarina	São Paulo	Espírito Santo
Rio de Janeiro	*	0.00020	0.22859	0.50926
Santa Catarina	0.01859	*	0.00020	0.00050
São Paulo	0.00113	0.01749	*	0.72092
Espírito Santo	-0.00051	0.01626	-0.00168	*

Note: significance level is 0.005 (after applying Bonferroni's correction); s.e. < 0.0043.

following population samples were added: Southeast [6], South (Rio Grande do Sul, Santa Catarina and Paraná states) [3], Minas Gerais [3], São Paulo Afro-Brazilians [24], North (Amapá, Pará, Rondônia and Acre states) [3], Pernambuco [3] and Alagoas [4]. The results obtained showed low  $F_{ST}$  values (below 0.4%) between Rio de Janeiro and all samples from Northeast and Southeast states, except the Afro-Brazilian sample from São Paulo, which revealed significant differences in all comparisons (Table S3).  $F_{ST}$  values above 1% were detected between Rio de Janeiro and the remaining populations from South and North regions. However, the probability of non-differentiation between Rio de Janeiro and the North was above the significance values, probably due to the smaller size of this sample compared with that from Santa Catarina (which showed similar genetic distance). In the MDS, the samples from the Northeast and Southeast form a cluster in the centre of the plot, clearly separated from the populations in the South and North regions (Fig. 1).

### 3.3. Native American haplogroups in Rio de Janeiro

The major percentage of Native American maternal lineages in Rio de Janeiro belongs to haplogroup B (38%), followed by haplogroups A (3.0%) and C (26%). At lower percentages, there were also samples belonging to haplogroup D, approximately 6%. This distribution is in accordance with what was found by Fridman et al. [6] for the Native American haplogroups in the south-eastern region of Brazil. The haplogroups' distribution and variation are represented in Fig. 2. The star-like distribution of the median-joining network indicates no recent expansions, as corroborated by the mismatch distribution profile (Fig. S2C-D). High haplotype diversity was observed with two haplotypes being shared between two individuals (one in haplogroup B and another in haplogroup C) and one haplotype being shared between three individuals from haplogroup B.

Before the colonization period in South America, it was estimated that approximately 2.4 million people belonging mainly to the Tupi and Gê linguistic groups lived in the territory that now comprises Brazil. During the 16th century the area surrounding Rio de Janeiro, as well as the remaining southeast coast of Brazil, was inhabited by Tupinambás (Tupi branch). However, after 300 years of Portuguese colonization, the number of Native Americans was significantly reduced [2,25].

Native American haplogroup frequencies in Rio de Janeiro were compared to those in South American native populations from four

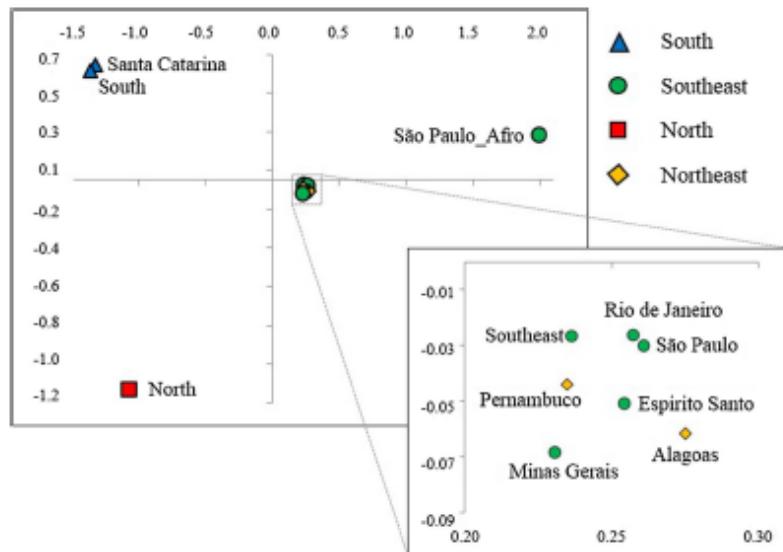


Fig. 1. MDS plot based on  $F_{ST}$  genetic distances between admixed Brazilian populations for HVSI mtDNA sequences.

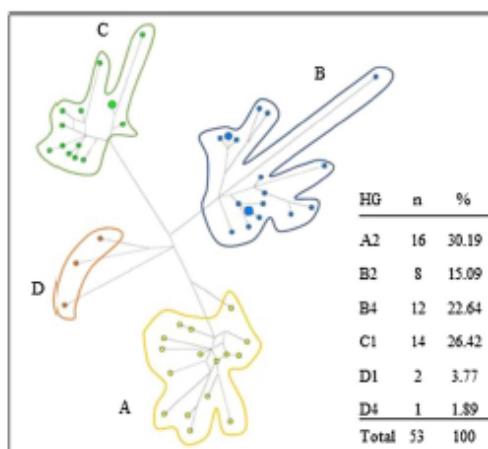


Fig. 2. Phylogenetic network representation of the Native American haplogroups in the Rio de Janeiro population. The size of the branches is proportional to the number of polymorphisms that separate each pair of haplotypes, and the size of the circles is proportional to the number of equal haplotypes.

macro linguistic groups [26]: Ge-Pano-Carib (which includes the Gé group) [27–29], Equatorial-Tucanoan (which includes the Tupi group) [27–29], Andean [29,30] and Chibcha-Paezan [29,31]. Due to different haplogroup resolution in each published population, haplotypes were combined into major haplogroups A, B, C and D. In this comparison (Fig. S2), it was not possible to infer the origin or affiliation of the Native American maternal lineages found in Rio de Janeiro since no strong geographic or linguistic association with mtDNA emerged. Indeed, previous studies using Y-chromosomal and autosomal markers in native South American populations revealed no clear association between their genetic composition and geographic and linguistic evidence

[32,33]. These findings were explained by a high genetic drift between small and isolated indigenous groups from a common initial package of lineages throughout the continent.

#### 3.4. Eurasian haplogroups in Rio de Janeiro

A total of 66 sequences found in Rio de Janeiro are distributed among haplogroups commonly spread through Europe, except for one sample belonging to haplogroup M7c1 (Table 3). A query in the EMPOP database for this sample sequence (PRJ023) revealed no match but several neighbouring haplotypes, all present in populations from Southeast Asia and North America. An inquiry about the sample donor revealed that both parents were born in East Timor.

Since the arrival of the first Europeans, Brazil has been subject to several immigration waves from Portugal, which have remained to this day. Even though immigration from other European countries was less important during the colonial period, this outlook changed in the last two centuries, starting with an important arrival of Germans, followed by Italians and later Spanish.

To investigate the impact of the different European migrations in

Table 3  
European haplogroups distribution in Rio de Janeiro population.

HG	n	HG	n	HG	n	HG	n
H	5	HV17	1	R	1	U5b2b1a	1
H10 + (16093)	1	I	1	R0	8	U6a1b1a	1
H13b1b	1	J1c	1	T2b	2	U6a3b	1
H15	1	J1c + 16261	1	T2b + 150	1	U6a6a	1
H1c	1	J1c?	1	T2b + 151	1	U6a7a1b	1
H24	1	J1d1a	1	T2b4a	1	V7	1
H2a2b	1	J2	1	T2c1d + 152	1	V9	1
H33c	1	J2alala	1	U3a1	1	W + 194	2
H3c2	1	J2b1c1	1	U3c123	1	W1g	1
H5	1	K1a	1	U4e2	1	X2m	1
H66a	1	K1c	1	U5a1	1		
HV	2	M1a8	1	U5b	2		
HV0	3	M7c1	1	U5b2b1a1	1	Total	66

Rio de Janeiro, a genetic distance analysis was performed between European haplotypes in Rio de Janeiro and population samples from Portugal [34], Spain [8], Italy [35], Germany [36], Netherlands [37], Austria [38], Finland [39], Denmark [40] and Greece [41]. Because data were not available for the CR in some of these populations, the analysis was restricted to the HVS-I and HVS-II segments.

Considering only the comparisons between Rio de Janeiro and the European samples, low  $F_{ST}$  genetic distances were observed (lower than 0.012), with differentiation p-values that were above the significance level (after applying Bonferroni's correction), except for the comparison with Greece (Table S4).

Previous studies showed that the European Y-chromosomal ancestry in Rio de Janeiro was mainly mediated by Portuguese contributors [42,43]. A main Portuguese maternal contribution would also be compatible with our results. However, the low differentiation between Portugal, Germany, Netherlands, Italy, Austria, Denmark, Greece and Spain precludes the use of the available mtDNA information to clearly determine the source of the current European maternal lineages in Rio de Janeiro.

### 3.5. African haplogroups in Rio de Janeiro

The 86 African haplotypes found in the Rio de Janeiro population are distributed among haplogroups L0, L1, L2, L3 and L4 (Fig. 3). The haplotype variation of these African haplogroups is represented in a median-joining network (Fig. 3), showing a clear separation of the haplotypes. High diversity can be observed within each clade, with only five different haplotypes being shared by two individuals (two within L0 and L3, and one within L1). This analysis shows no signs of important founder events in the African genetic background of Rio de Janeiro, which is also supported by the mismatch distribution profile of the African sub-lineages, as represented in Figs. S1G and H.

During the Atlantic slave trade, Brazil was the country that received the largest number of sub-Saharan Africans, arriving mostly from three regions: Central West Africa (including all coast between Guinea-Bissau and Equatorial Guinea), South West Africa (mainly from Congo and Angola), and Southeast Africa (principally from Mozambique) [25,44].

To investigate the most likely origin of the African maternal lineages, the observed haplogroup frequencies were compared with those available for the three regions mentioned above, as well as for North and East Africa. Since the available data has different resolution levels, haplotypes were pooled inside the branches that are present in our sample (see table included in Fig. 3). Table 4 shows the distribution of these haplogroups in East (Somali and Uganda [45]), North (Morocco [45]), South East (Mozambique and Zimbabwe [45]), South West

(Angola [45–47]), and Central West (Cameroon, Gambia, Ghana, Ivory Coast, Nigeria, Sierra Leone, Togo [45]) Africa and Rio de Janeiro (this study).

The most common subclades of African ancestry in our sample were L1c and L3e, representing approximately 40% of the L haplogroups, and they were also the most frequent in the South West African region (Table 4). These branches are very common in Central West, South West and Southeast Bantu populations (e.g., [45–48]).

Also frequent in Rio de Janeiro are L0a and L2a, which are the most common in the Southeast sample (Table 4). These subclades are both widely spread in sub-Saharan Africa, with high frequency in the East Coast [45,48,49].

Clusters L3d and L3f each have 9% frequency in our L-samples. Together with L1b, L2b and L3b, they are more frequent in the Central West than in other regions of Africa. The three remaining lineages (L0d, L2c, L4b) have a total frequency of 6% in our sample and are also rare in all compared populations. L0d (together with L0k) has been described as characteristic of Khoisan populations [50–53]. L4b is frequent in East African Nilotes [49], although represented by a different subclade (L4b2) than that found in Rio de Janeiro (L4b1). The overall frequency profile of the L-lineages in Rio de Janeiro is very similar to that from South West Africa, with differences in the frequency of the L3d and L3f haplogroups that can be explained by contributions from Central West populations.

For a more in-depth understanding of the African maternal gene pool origin,  $F_{ST}$  genetic distances were calculated between L-lineages in Rio de Janeiro and samples from African populations based on HVS-I and HVS-II sequences (data not shown). The MDS plot of pairwise  $F_{ST}$  matrix (Fig. 4) shows a cluster including Rio de Janeiro, Cabinda [45] and Angola [47], with no significant differences among them ( $F_{ST} \leq 0.0087$ ;  $p \geq 0.1134$ ). Larger distances were found when comparing Rio de Janeiro with four Bantu-speaking groups from south-western Angola [46] ( $F_{ST} \geq 0.0528$ ;  $p \leq 0.00036$ ). This group showed similarly large distances when compared with the two other samples from north Angola [45,47]. The comparison of Rio de Janeiro with samples from Central West (Guinea Bissau [54] and Ghana [55]), South East (Mozambique [45]) and East (Kenya [56] and Uganda [49]) Africa revealed similar genetic distances, with  $F_{ST}$ s varying between 0.0288 (Guinea Bissau) and 0.0399 (Kenya). As expected, highly significant differences ( $F_{ST} \geq 0.0801$ ;  $p \leq 0.000005$ ) were observed between Rio de Janeiro and populations from North Africa (Tunisia [57] and Morocco [57,58]) as well as a Khoisan group from Angola [50].

### 3.6. Comparison of mtDNA and autosomal gene pools

Ancestry proportions were calculated for this group of samples, except for one sample (PRJ\_09) (unpublished data), using the protocol described by Pereira et al. [59] for a multiplex of 46 ancestry informative markers (AIMs). Proportions of African, Native American and European contributions were calculated for each sample using the software STRUCTURE v2.3.3 [60,61].

When comparing the ancestries obtained for autosomal and mtDNA data (Fig. 5), an association was observed between these types of markers. In fact, the average African autosomal ancestry was higher in the subsample of individuals carrying African mtDNA haplogroups, while being lower in the samples with Native American and European mtDNA lineages. Conversely, European ancestry was higher in the group of individuals with European mtDNA lineages than in samples with Native American or African maternal lineages. Concerning the Native American ancestry observed for the autosomal markers, smaller differences were found between mtDNA subgroups, although a higher value was found in samples with Native American mtDNA haplogroups.

Based on the frequency distributions of the autosomal markers, the two subsamples of African and European mtDNA haplogroups were further compared by calculating the  $F_{ST}$  between them and the probability of non-differentiation (after 10,000 permutations). A significant

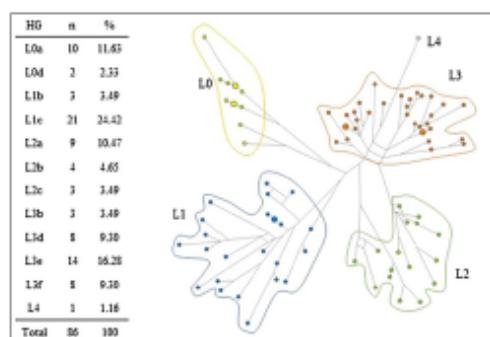


Fig. 3. Distribution of African haplogroups in the Rio de Janeiro population and phylogenetic network representation. The size of the branches is proportional to the number of polymorphisms that separate each pair of haplotypes, and the size of the circles is proportional to the number of equal haplotypes.

**Table 4**  
Frequency of African haplogroups in 5 different regions of Africa and in the Rio de Janeiro population.

Haplogroups																		
Region	Non-L	L0a	L0I	L0 – others	L1	L1b	L1c	L2a	L2b	L2c	L2 – others	L3b	L3d	L3e	L3f	L3 – others	L4-6	
Bast	6%	19%	0%	3%	0%	0%	3%	14%	1%	0%	1%	8%	3%	4%	3%	16%	20%	
North	83%	1%	0%	1%	0%	7%	0%	2%	0%	0%	1%	1%	1%	3%	1%	0%	0%	
South-East	0%	25%	5%	0%	1%	1%	9%	32%	1%	1%	2%	2%	9%	13%	1%	1%	1%	
South-West	0%	16%	4%	0%	1%	4%	20%	11%	3%	3%	1%	3%	5%	20%	7%	1%	0%	
Central West	2%	7%	0%	0%	0%	8%	4%	17%	4%	2%	2%	10%	8%	20%	13%	1%	1%	
Rio de Janeiro	0%	12%	2%	0%	0%	3%	24%	10%	9%	3%	0%	3%	9%	16%	9%	0%	1%	

The values for the most frequent haplogroups in each African region are indicated in bold.

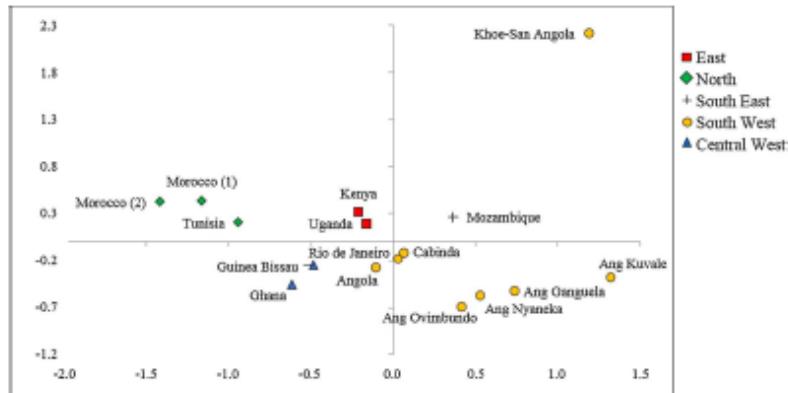
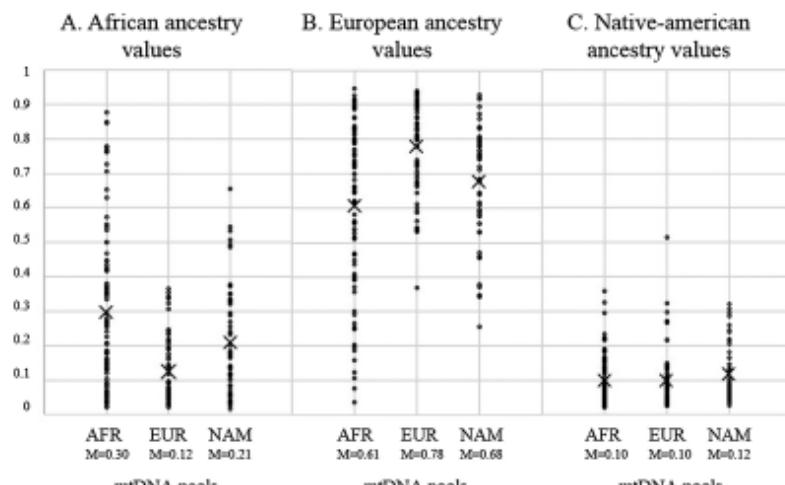


Fig. 4. MDS plot based on  $F_{ST}$  genetic distances between a sample of African lineages from Rio de Janeiro and populations from Africa, calculated using HVS-I and HVS-II data.



**Fig. 5.** Distribution of the African (A), European (B) and Native American (C) ancestry values for individuals carrying African (AF1), European (EUR) and Native American (NAM) mtDNA lineages. The mean value for each mtDNA root is indicated (M) and commented in the results by a cross.

differentiation was found ( $F_{ST} = 0.01500$ ;  $p < 0.000005$ ), supporting the stratification of Rio de Janeiro population.

#### 4. Conclusions

Our study provided a deeper and more detailed understanding of the genetic composition of Rio de Janeiro, one of the main entrances for people in Brazil during the colonial period.

The results show that control region sequences are very diverse in Rio de Janeiro, presenting high values of haplotype diversity and exclusion power. When lineages were grouped by continental origin, it was observed that although the European subset has the lowest probability of a match, the highest exclusion power (more than one-point differences) was found in the African subset. These results highlight the importance of considering the haplotype distribution profile in the population when evaluating the efficiency of mtDNA marker analysis in forensic case studies.

The high mtDNA haplotype diversity observed in the Rio de Janeiro population resulted from the admixture of well-differentiated source populations, with a higher percentage of African lineages, followed by European and Native American contributions. The African mtDNA subset in Rio de Janeiro showed a genetic profile that resembles those found in populations from South West Africa, with minor Central West and South East influences.

The lack of a clear pattern among reference populations prevented the ethnic origin of the Native American lineages currently found in Rio de Janeiro from being inferred. Additionally, because of the low genetic differentiation found among European source populations, inferences concerning the European lineages present in our sample would require mtDNA analysis of higher resolution.

The detection of an association between mtDNA haplogroups and AIMs ancestries confirmed the presence of substructure in the Rio de Janeiro population.

Overall, this study contributed to a better understanding of maternal ancestry in Rio de Janeiro and its genetic structure. Additionally, the haplotype information described here can be used to establish a high-quality mtDNA database for forensic purposes.

#### Acknowledgements

This work was supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES. PS was supported by Fundação de Amparo à Pesquisa do Rio de Janeiro – FAPERJ (process E-26/200.044/2016). LG was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq (ref. 305330/2016-0).

We would like to thank Nicole Huber and Bettina Zimmermann, Institute of Legal Medicine, Medical University of Innsbruck, for technical support.

#### Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.forsig.2018.02.003>.

#### References

- [1] C.A. Lockard, Societies, Networks, and Transition Volume II: Since 1450: A Global History, 3rd ed., Cengage Learning, 2014.
- [2] N.E. Jaffey, Gender, Race and Religion in the Colonization of the Americas, 1st ed., Ashgate Publishing Limited, 2016.
- [3] J. Alves-Silva, M. da Silva-Santos, P.E. Guimaraes, A.C.F. Ferreira, H.J. Bandelt, S.D. Pena, V.F. Prado, The ancestry of Brazilian mtDNA lineages, *Am. J. Hum. Genet.* 67 (2000) 444–461.
- [4] A.B.G. Barboza, L.A.P.D. Silva, D.A. Araújo, V.Q. Balbim, L. Mauricio-Da-Silva, Mitochondrial DNA control region polymorphisms in the population of Alagoas State, North-Eastern Brazil, *J. Forensic Sci.* 53 (2008) 142–146.
- [5] S. Bernardo, R. Hermida, M. Denísio, I.M. Silva, E.F. De Carvalho, mtDNA ancestry of Rio de Janeiro population, Brazil, *Mol. Biol. Rep.* 41 (2014) 1945–1950.
- [6] C. Fridman, R.S. Gonzalez, A.C. Gon, M.M.S.G. Cardoso, Haplotype diversity in mitochondrial DNA hypervariable region in a population of south-eastern Brazil, *Int. J. Legal Med.* 128 (2014) 589–593.
- [7] I. Palencia, L. Valverde, A. Álvarez, L.M. Cainf, S. Cardoso, M.A. Alfonso-Sánchez, M.F. Pinheiro, M.M. De Pancorbo, Mitochondrial DNA diversity in a population from Santa Catarina (Brazil): Predominance of the European input, *Int. J. Legal Med.* 124 (2010) 331–336.
- [8] I. Paiva, B. Zimmermann, A. Gómez, A. Rodríguez-Monge, G.G. Parra, C. Alves, A. Alonso, C. Fridman, S. Cardoso, G. Lima, M.J. Anjos, M.R. Whittle, M. Monteiro, R.M.B. Ciccarelli, A.M. Rech, C. Albarca, M.M. De Pancorbo, M.F. Pinheiro, M. Carvalho, D.R. Simão, W. Parson, The GHEP-EMPOP collaboration on mtDNA population data – A new resource for forensic casework, *Forensic Sci. Int. Genet.* 5 (2011) 146–151.
- [9] N.M. Sanches, G.G. Parra, R.F. Figueiredo, A.O. de Melo, R.M.B. Ciccarelli, Mitochondrial DNA control region diversity in a population from Espírito Santo state, Brazil, *Mol. Biol. Rep.* 41 (2014) 6645–6648.
- [10] W. Parson, H.-J. Bandelt, Extended guidelines for mtDNA typing of population data in forensic science, *Forensic Sci. Int. Genet.* 1 (2007) 13–19.
- [11] R.M. Andrew, I. Kubacka, P.F. Chinnery, R.N. Lightowler, D.M. Turnbull, N. Howell, Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA, *Nat. Genet.* 23 (1999) 147.
- [12] W. Parson, L. Gusmão, D.R. Härke, J. A. Irwin, W.R. Mayr, N. Morling, E. Pokorná, M. Pérez, A. Salas, P.M. Schneider, T.J. Parsons, DNA Commission of the International Society for Forensic Genetics: revised and extended guidelines for mitochondrial DNA typing, *Forensic Sci. Int. Genet.* 13C (2014) 134–142.
- [13] M. van Oven, M. Kayser, Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation, *Hum. Mutat.* 30 (2009) E366–E94.
- [14] W. Parson, A. Ditz, EMPOP-a forensic mtDNA database, *Forensic Sci. Int. Genet.* 1 (2007) 88–92.
- [15] B. Zimmermann, A.W. Rück, A. Ditz, W. Parson, Improved visibility of character conflicts in quasi-endian networks with the EMPOP NETWORK software, *Croat. Med. J.* 55 (2014) 115–120.
- [16] R. Fugel, S. Delgado, HaploSearch: a tool for haplotype-sequence two-way transformation, *Mitochondrion* 11 (2011) 366–367.
- [17] L. Excoffier, G. Laval, S. Schneider, Arlequin (version 3.0); an integrated software package for population genetics data analysis, *Evol. Bioinform. Online.* 1 (2005) 47–50.
- [18] H.J. Bandelt, P. Forster, R.C. Sykes, M.R. Richards, Mitochondrial portraits of human populations using median networks, *Genet. Soc. Am.* 74 (1995) 3–75 3.
- [19] H.J. Bandelt, P. Forster, A. Röhl, Median-joining networks for inferring intraspecific phylogenies, *Mol. Biol. Evol.* 16 (1999) 37–48.
- [20] A. Carrasco, W. Bar, P. Lincoln, W. Mayr, N. Morling, B. Okiesen, P. Schneider, B. Budowle, B. Brinkmann, P. Gill, M. Holland, G. Tally, M. Wilcox, DNA commission of the international society for forensic genetics: guidelines for mitochondrial DNA typing, *Forensic Sci. Int.* 110 (2000) 17–85.
- [21] Scientific Working Group on DNA Analysis Methods, Interpretation Guidelines for Mitochondrial DNA Analysis by Forensic DNA Testing Laboratories, (2013), pp. 1–23.
- [22] P. Soares, L. Ersmark, N. Thomsen, M. Moraes, T. Rita, A. Röhl, A. Salas, S. Oppenheimer, V. Macaulay, M.R. Richards, Correcting for purifying selection: an improved human mitochondrial molecular clock, *Am. J. Hum. Genet.* 84 (2009) 740–750.
- [23] J.A. Irwin, J.L. Saenger, H. Niedenbörker, K.M. Strauss, K.A. Stark, T.M. Diegoli, A. Brandstetter, W. Parson, T.J. Parsons, Investigation of heteroplasmy in the human mitochondrial DNA control region: a synthesis of observations from more than 5000 global population samples, *J. Mol. Evol.* 68 (2009) 516–527.
- [24] V.F. González, C.M.B. Carvalho, M.C. Bertolini, S.P. Bydlowski, S.D.J. Item, The phylogeography of African Brazilians, *Hum. Hered.* 65 (2008) 23–32.
- [25] L. Reichell, The Cambridge World History of Slavery vol. 3, (2011).
- [26] L. Campbell, American Indian Languages: the Historical Linguistics of Native America, Oxford University Press, New York, 1997.
- [27] V. Ramallo, R. Basso-Machado, C. Bravi, M.D. Coble, F.M. Sakano, T. Hünemeier, M.C. Bertolini, Demographic expansions in South America: enlightening a complex scenario with genetic and linguistic data, *Am. J. Phys. Anthropol.* 150 (2013) 423–463.
- [28] S. Matzuda, E. Guitard, E. Crubézy, J.M. Dugoujon, M.C. Bertolini, S.I. Bonatto, M.H. Hutz, E. Bois, P. Tisnáka, G. Lameyre, P.M. Salzano, Uniparental (mtDNA, Y-chromosome) polymorphisms in French Guiana and two related populations – implications for the region's colonization, *Ann. Hum. Genet.* 72 (2008) 145–156.
- [29] M.M. Torres, C.M. Bravi, M.C. Bertolini, C. Dugay, S. Callegeri-Jacques, D. Ortiz, G. Bedoya, H.G. De Rezende, A. Ruiz-Linares, A revert of the major founder Native American haplogroup C common in populations from Northern South America, *Am. J. Hum. Biol.* 18 (2006) 59–65.
- [30] M. De Saint Pierre, C.M. Bravi, J.M.B. Mori, N. Palka, M. Tanaka, E. Illop, S.L. Bonatto, M. Moraes, An alternative model for the early peopling of Southern South America revealed by analysis of three mitochondrial DNA haplogroups, *PLoS One* 7 (2012).
- [31] C. Xavier, J.J. Bailes, V. Gomez, J.M. Ospina, J. Aquino, W. Parson, A. Amorim, L. Gusmão, A. Goicoechea, Admixture and genetic diversity distribution patterns of non-recombining lineages of native american ancestry in colombian populations, *PLoS One* 10 (2015) 1–13.
- [32] S. Wang, C.M. Lewis, M. Jakobsson, S. Ramachandran, N. Ray, G. Bedoya, W. Rojas, M.V. Parra, J.A. Molina, C. Gallo, G. Mazzotti, G. Polent, K. Hill, A.M. Hurtado, D. Labuda, W. Klitz, R. Bernstein, M.C. Bertolini, F.M. Salzano, M.L. Peñal-Estevez, I.T. Tamayo, E. Illop, P. Rothhammer, L. Excoffier, M.W. Feldman, N.A. Rosenberg, A. Ruiz-Linares, Genetic variation and population structure in Native Americans, *PLoS Genet.* 3 (2007) 2049–2057.
- [33] L. Roewer, M. Nothnagel, L. Gusmão, V. Gomez, M. González, D. Corach, A. Salas, E. Alchami, T. Pálha, N. Santos, A. Ribeiro-Dos-Santos, M. Geppert, S. Willaume, M. Nagy, S. Zwyer, M. Betsa, C. Nájera, B. Martínez-Jiménez, F. González-Andrade, E. Figueredo de Carvalho, D.A. da Silva, J.J. Bailes, D. Turbin, A.M. López Parra, E. Arroyo-Pardo, U. Toscancini, I. Borja, C. Badelta, E. Ewart, S. Santos, M. Kwecaek, Continent-wide decoupling of Y-chromosomal genetic variation from language and geography in native South Americans, *PLoS Genet.* 9 (2013) e1003460.
- [34] S.L. Marques, A. Goicoechea, A.M. Rocha, M.J. Prata, A. Amorim, L. Gusmão, C. Alves,

- I. Alvarez, Portuguese mitochondrial DNA genetic diversity – An update and a phylogenetic revision, *Forensic Sci. Int. Genet.* 15 (2015) 27–32.
- [35] C. Turchi, L. Busconi, C. Pavlidis, P. Grigioni, A. Brandstätter, A. Achilli, W. Parson, A. Taglialatella, M. Alu, G. Bedoglio, C. Bini, I. Boschi, E. Camerani, S. Giac, L. Gianti, M. Lancia, V. Pascali, S. Pasquini, U. Ricci, C. Robino, Italian mitochondrial DNA database: results of a collaborative exercise and profile-level scoring, *Int. J. Legal Med.* 122 (2008) 199–204.
- [36] A. Brandstätter, R. Klein, N. Dufner, P. Wiegand, W. Parson, Application of a quasi-median network analysis for the visualization of character conflicts to a population sample of mitochondrial DNA control region sequences from southern Germany (Ulm), *Int. J. Legal Med.* 120 (2006) 310–314.
- [37] I. Chakraborty, M. Van Oyen, S. Bauer, B. Zimmersmann, G. Huber, C. Xavier, W. Parson, P. De Knijff, M. Kayser, High-quality mtDNA Control Region Sequences from 680 Individuals Sampled Across the Netherlands to Establish a National Forensic mtDNA Reference Database, Elsevier Ireland Ltd., 2016.
- [38] A. Brandstätter, H. Niedenbürger, M. Paulic, P. Grubwieser, W. Parson, Generating population data for the EMPOP database—An overview of the mtDNA sequencing and data evaluation processes considering 273 Austrian control region sequences as example, *Forensic Sci. Int.* 166 (2007) 164–175.
- [39] M. Hedman, A. Brandstätter, V. Rinnehoff, P. Skutinen, J.U. Palo, W. Parson, A. Sajantila, Finnish mitochondrial DNA HVs-I and HVs-II population data, *Forensic Sci. Int.* 172 (2007) 171–178.
- [40] M. Mikkelson, E. Stenseth, E.M. Rasmussen, N. Morling, Mitochondrial DNA HV1 and HV2 variation in danes, *Forensic Sci. Int. Genet.* 4 (2010) 87–88.
- [41] J. Irwin, J. Saunier, K. Strauss, C. Painter, T. Diegoli, K. Stark, L. Kovanci, A. Brandstätter, M.A. Carikou, W. Parson, T.J. Parsons, Mitochondrial control region sequences from northern Greece and Greek Cyprus, *Int. J. Legal Med.* 122 (2008) 87–95.
- [42] D.A. Silva, E. Carvalho, G. Costa, I. Tavares, A. Amorim, L. Gusmão, Y-chromosome genetic variation in Rio De Janeiro population, *Am. J. Hum. Biol.* 18 (2006) 829–837.
- [43] A.M. Oliveira, P.M. Domingos, V. Gomes, A. Amorim, J. Januzzi, E.F. De Carvalho, L. Gusmão, Male lineage strata of Brazilian population disclosed by the simultaneous analysis of STRs and SNPs, *Forensic Sci. Int. Genet.* 13 (2014) 264–268.
- [44] A.I. Araújo, African Heritage and Memories of Slavery in Brazil and the South Atlantic World, Cambria Press, 2015.
- [45] M. Cenzo, L. Gusmão, V. Čemý, N. Uddin, D. Syndercombe-Court, A. Gómez-Carballa, T. Göbel, P.M. Schneider, A. Salas, Comprehensive analysis of pan-African mitochondrial DNA variation provides new insights into continental variation and demography, *J. Genet. Genomics.* 43 (2016) 133–143.
- [46] M. Coelho, F. Sequeira, D. Luiselli, S. Beloza, J. Rocha, On the edge of Bantu expansions: mtDNA, Y chromosome and lactase persistence genetic variation in southwestern Angola, *BMC Evol. Biol.* 18 (2008) 1–18.
- [47] S. Haza, A. Salas, P. Calafell, F. Cortés-Real, J. Bertranpetti, Á. Carrasco, D. Comas, Insights into the western Bantu dispersal: mtDNA lineage analysis in Angola, *Hum. Genet.* (2004) 439–447.
- [48] A. Salas, M. Richards, T. De la Fe, M.-V. Llorente, B. Sobrino, P. Sánchez-Díaz, V. Macaulay, A. Carrasco, The making of the African mtDNA landscape, *Am. J. Hum. Genet.* 71 (2002) 1082–1111.
- [49] V. Gomez, M. Pala, A. Salas, V. Álvarez-Iglesias, A. Amorim, A. Gómez-Carballa, Carrasco Á, D.J. Clarke, C. Hill, M. Mormina, M.A. Shaw, D.W. Damme, R. Pereira, V. Pereira, M.J. Pata, P. Sánchez-Díaz, T. Rita, P. Sozen, L. Gusmão, M.R. Richards, Mosaic maternal ancestry in the Great Lakes region of East Africa, *Hum. Genet.* 134 (2015) 1013–1027.
- [50] I. Fendt, G. Huber, A.W. Röck, B. Zimmersmann, M. Bodner, R. Delport, K. Schmidt, W. Parson, Mitochondrial DNA control region data from indigenous Angolan Khosa-San lineages, *Forensic Sci. Int. Genet.* 6 (2012) 663–663.
- [51] Y.S. Chen, A. Okkera, T.G. Schurr, A.M. Kogelnik, K. Huoponen, D.C. Wallace, mtDNA variation in the South African Kung and Khwe and their genetic relationships to other African populations, *Am. J. Hum. Genet.* 66 (2000) 1382–1383.
- [52] S.A. Tishkoff, M.K. Gonder, B.M. Henn, H. Mortensen, A. Knight, C. Gignoux, N. Remandopoulou, G. Lema, T.B. Nyambo, U. Ramakrishnan, F. a. Reed, J.L. Mountain, History of click-speaking populations of Africa inferred from mtDNA and Y chromosome genetic variation, *Mol. Biol. Evol.* 24 (2007) 2180–2195.
- [53] A. Knight, P.A. Underhill, H.M. Mortensen, L.A. Zhivotovsky, A.A. Lin, R.M. Henn, D. Louis, M. Ruhlen, J.L. Mountain, African Y chromosome and mtDNA divergence provides insight into the history of click languages, *Curr. Biol.* 13 (2003) 464–473.
- [54] M. Carvalho, P. Braga, A.M. Bento, V. Gomes, H. Antunes, H.A. Costa, V. Inapa, F. a. Serra, I. Balá, M.J. Andrade, F.L. Anjos, Corre-Real, Gusmão, Paternal and maternal lineages in Guiné-Bissau population, *Forensic Sci. Int. Genet.* 5 (2011) 114–116.
- [55] I. Fendt, A. Röck, B. Zimmersmann, M. Bodner, T. Thye, F. Tschannen, E. Owusu-Dabo, T.M.K. Göbel, P.M. Schneider, W. Parson, mtDNA diversity of Ghana: a forensic and phylogeographic view, *Forensic Sci. Int. Genet.* 6 (2012) 244–249.
- [56] A. Brandstätter, C.T. Peterson, J. Irwin, S. Mykje, D.K. Koch, W. Parson, T.J. Parsons, Mitochondrial DNA control region sequences from Nairobi (Kenya): inferring phylogenetic parameters for the establishment of a forensic database, *Int. J. Legal Med.* 118 (2004) 294–306.
- [57] C. Turchi, L. Busconi, E. Giachino, V. Onofri, I. Fendt, W. Parson, A. Taglialatella, Polymorphisms of mtDNA control region in Tunisian and Moroccan populations: an enhancement of forensic mtDNA databases with Northern Africa data, *Forensic Sci. Int. Genet.* 3 (2009) 166–172.
- [58] R. Abualhaid, K. Stark-Andmaghi, M. Bouabdellah, D. Squalli, J.A. Irwin, S. Amzazi, Mitochondrial DNA control region variation from samples of the Moroccan population, *Int. J. Legal Med.* 127 (2013) 757–759.
- [59] R. Pereira, C. Phillips, N. Pinto, C. Santos, S.E.B. dos Santos, A. Amorim, Á. Carrasco, L. Gusmão, Straightforward inference of ancestry and admixture proportions through ancestry-informative insertion deletion multiplexing, *PLoS One* 7 (2012).
- [60] D. Falush, M. Stephens, J.K. Pritchard, Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies, *Genetics* 164 (2003) 1587–1587.
- [61] J.K. Pritchard, M. Stephens, P. Donnelly, Inference of population structure using multilocus genotype data, *Genetics* 155 (2000) 945–959.

### **3.2 Caracterização do patrimônio genético de origem materna de indivíduos residentes no Paraguai, através da análise do mtDNA**

Os resultados preliminares obtidos do sequenciamento da região controle do mtDNA de indivíduos residentes no Paraguai foram divulgados em pôster no congresso internacional *27th Congress of the International Society for Forensic Genetics* (Seoul, Coreia do Sul, Setembro de 2017) e posteriormente publicados em formato de artigo na revista *Forensic Science International: Genetics Supplement Serie*, com o título “*Paraguay: Unveiling migration patterns with ancestry genetic markers*” (Artigo 2). Os haplótipos estão apresentados na presente tese no Apêndice C.

Uma colaboração do Laboratório de Diagnóstico por DNA com Instituto de Medicina Legal de Innsbruck permitiu aprofundar o estudo das linhagens maternas do Paraguai, através do sequenciamento da molécula completa do mtDNA de 105 indivíduos residentes no departamento de Alto Paraná.

Resultados preliminares deste estudo foram divulgados em apresentação oral no congresso internacional *11th Haploid Markers Conference* (Bydgoszcz, Polónia, Maio de 2018). O artigo referente aos resultados finais obtidos com os 105 mitogenomas paraguaios foi publicado na revista internacional *Forensic Science International: Genetics*, com o título “*Defining mtDNA origins and population stratification in Rio de Janeiro*” (Artigo 3). O material suplementar está disposto no Apêndice D.

## Artigo 2

Título traduzido: **Paraguai: Revelando padrões de migração com marcadores genéticos de ancestralidade**

Resumo traduzido:

Antes da chegada dos colonizadores espanhóis, a região Este do Paraguai era habitada por indígenas Guarani. Depois da guerra do Paraguai em 1870, que terminou com a perda de uma grande porcentagem da população masculina, a imigração para o país foi encorajada. Dados de imigração indicam uma grande entrada de euro-asiáticos no território. Desde os anos 60, um grande número de brasileiros e argentinos chegaram também ao Paraguai. Um total de 299 amostras de indivíduos residentes na região Este do Paraguai foram sequenciadas para a região controle do mtDNA. Cerca de 88 % dos haplótipos pertencem a haplogrupos nativo-americanos. Um estudo preliminar utilizando AIMs no mesmo grupo de amostras indica uma elevada contribuição autossomal europeia e nativo-americana. A comparação dos dois tipos de marcadores revela que a ancestralidade europeia para autossomos é superior à esperada, considerando a média entre mtDNA e cromossomo Y. Estes resultados suportam um processo de mistura recente entre paraguaios e outras populações, provavelmente já miscigenadas, onde os homens terão contribuído para a elevada ancestralidade europeia.



Contents lists available at ScienceDirect

## Forensic Science International: Genetics Supplement Series

journal homepage: [www.elsevier.com/locate/fsigss](http://www.elsevier.com/locate/fsigss)

## Paraguay: Unveiling migration patterns with ancestry genetic markers

F. Simão<sup>a</sup>, A.P. Ferreira<sup>a</sup>, C. Vullo<sup>b</sup>, C. Xavier<sup>c</sup>, G. Huber<sup>c</sup>, A. Quiroz<sup>d</sup>, P. Machado<sup>e</sup>, V. Velázquez<sup>f</sup>, E.F. Carvalho<sup>a</sup>, L. Gusmão<sup>a,g,\*</sup>, W. Parson<sup>c,h</sup><sup>a</sup> Laboratório de Diagnóstico por DNA (LDD), Universidade do Estado do Rio de Janeiro, Rio de Janeiro, Brazil<sup>b</sup> DNA Forensic Laboratory, Argentinean Forensic Anthropology Team (RAAF), Córdoba, Argentina<sup>c</sup> Institute of Legal Medicine, Medical University of Innsbruck, Innsbruck, Austria<sup>d</sup> Instituto de Previdência Social, Asunción, Paraguay<sup>e</sup> Legislación Jurídica, Universidad Católica, Hernández, Paraguay<sup>f</sup> Hospital Regional, Encarnación, Paraguay<sup>g</sup> Institute of Pathology and Immunology of the University of Porto (IPATIMUP)/Instituto de Investigação e Inovação em Saúde, IIS, Universidade do Porto, Portugal<sup>h</sup> Forensic Science Program, The Pennsylvania State University, PA, USA

## ARTICLE INFO

## Keywords

mtDNA  
Population genetics  
Paraguay  
Admixed populations  
AIMs

## ABSTRACT

Before the arrival of Spanish settlers, the East region of Paraguay, was inhabited by Guaraní people. After the Paraguayan war in 1870, which ended in loss of a high percentage of the male population, the migration to the country was encouraged. Immigration data indicate a high input of Eurasians to the territory. Also, since 1960s, a large number of Brazilians and Argentines arrived in Paraguay. Samples from the eastern provinces of Paraguay were sequenced for the mtDNA control region and 88% presented native American haplogroups. A preliminary study on the same samples using AIMs indicates a high autosomal contribution from Europe and native America. The comparison of both type of markers showed that the European ancestry for autosomes is higher than expected when averaging mtDNA and the Y chromosome. This result supports recent admixture between Paraguayans and other populations probably already admixed, where the men contributed with high European ancestry.

## 1. Introduction

After the entrance of people in the American continent about 10–15 thousand years ago there was a rapid dispersal towards the South. The Paraguayan territory was first occupied by indigenous hunter gatherers. Later, the Guaraní arrived in the region and, presenting a less nomadic lifestyle, settled in the territory [1].

In the sixteenth century Paraguay was occupied by colonizers, similarly to what happened in other south American countries. The Europeans arrived in the territory in 1520, but only in 1537 the first Spanish city was settled (Asunción) [1,2].

The inexistence of mineral wealth quickly decreased the interest of Spanish colonizers in this territory, making Paraguay an isolated region, geographically, culturally and politically separated from the other economic forces from the colonial period in South America. Nevertheless, the few colonizers that remained in the territory initiated the use of indigenous and African slave work, although the latter was not as strong as in other South American counties [1].

Paraguay suffers a tremendous reduction in the number of male population (more than 50%) after the "War of the Triple Alliance",

fought against Argentina, Brazil and Uruguay. With the end of the war in 1870, the immigration to the country was encouraged. Subsequently, a high input of European immigrants occurred (Italians, Germans, French, Spanish, Koreans and Japanese) and, after the 1960s, many Brazilians and Argentines also entered Paraguay [2].

The country is currently divided into two geographical regions, separated by the Paraguay river. The most populated region lies east of the river that comprises 14 provinces, while the west, also known as Chaco, has three provinces.

As a result of the historical events described above, the people from Paraguay present a peculiar ethnic composition, different from other South American countries. Historical and linguistic data are sometimes incomplete, preventing a thorough characterization of the ancestry of today's populations. Therefore, the recruitment of information from other fields, as population genetics, can be enlightening. With this study, we aimed to investigate the maternal ancestry in Eastern Paraguay obtained by mtDNA control region sequencing. We also intended to elucidate the genetic structure and the migration patterns of the territory by combining mtDNA and autosomal ancestry informative indels (AIM-Indels).

\* Corresponding author at: Laboratório de Diagnóstico por DNA (LDD), Universidade do Estado do Rio de Janeiro, Rio de Janeiro, Brazil.  
E-mail address: leonorbgusmao@gmail.com (L. Gusmão).

## 2. Materials and methods

### 2.1. DNA samples

A total of 299 blood samples were collected from individuals living in different departments of eastern Paraguay, under informed consent signed by the participants. DNA was extracted with Chelex<sup>®</sup> 100 (BioRad) 5% as described by Walsh and collaborators in 1991 [3].

### 2.2. Genetic markers and genotyping

All samples were sequenced for the entire mtDNA control region (CR) following the EMPOP protocol. The amplification was performed using the primers F15971 (5' TTAACCTCACCATTAGCCC 3') and R639 (5' GGGTGATGTGAGCCCGTCA 3'). The PCR conditions were: denaturation at 95 °C for 2 min, 35 cycles of 95 °C for 15 s, 56 °C for 30 s, 90 °C for 90 s and final extension of 72 °C for 10 min. The purification of PCR products was performed with the enzymes ExoSap-IT (GE Healthcare). Total control region was sequenced with the primers F15971, F15 (5' CACCTATTAAACCCTCACG 3'), R16 (5' TGATAGACCTGTGATCCATCGTGA 3'), F16268 (5' CACTAGGATAACC AACAAACC 3') and R639, using the following conditions: 96 °C for 1 min and 25 cycles of 95 °C for 15 s, 50 °C for 5 s, 60 °C for 4 min. Sephadex (Illustra Sephadex DNA Grade; GE Healthcare) was used for final purification. An ABI 3500 (Applied Biosystems) was used for separation and detection of the sequences.

A subset of 39 randomly selected samples was also genotyped for 46 AIM-Indel markers, using the protocol previously described [4]. Fragments were separated and detected on an ABI 3500 (Applied Biosystems).

### 2.3. Haplotype assignment and statistical analysis

For haplotype classification, the sequences were compared to the revised Cambridge Reference Sequence (rCRS) with the software Sequencher v5.4.6, following the International Society for Forensic Genetics (ISPG) guidelines for mtDNA [5]. Haplotype grouping was performed via EMPOP according to PhyloTree build 17, February 2016 [6]. The EMPOP tools "Network" and "EMPcheck" were used to perform a posteriori quality control [7]. Haplotype frequencies were calculated by direct counting and haplotype diversity was determined using the Arlequin software v3.5 [8].

The Software GeneMapper v4.1 (Applied Biosystems) was used in Indels genotype assignments. Ancestry proportions were estimated with software STRUCTURE v2.3.3 [9], using reference populations from Africa, Europe and Native America [4]. The "Admixture model" was used, and allele frequencies were correlated and updated using only individuals with POPFLAG = 1.

## 3. Results and discussion

The sample of 299 sequences analyzed show a high haplotype diversity ( $0.9937 \pm 0.0011$ ). Most haplotypes are from Native American origin (88%), distributed among haplogroups A, B, C and D. The most common lineage was C1, representing 31% of the native American haplotypes. European and African haplogroups have lower frequencies, being present in 8% and 4% of the samples, respectively. The highest proportion of European haplotypes fall into haplogroup H, and the Africans into haplogroup L3.

A group of 39 samples was also typed for AIM-Indels. In this case, the highest proportion of ancestry was European (60%), followed by native American (31%) and African (9%) ancestries. Considering a possible immigration from South-East Asia, a second analysis was performed with four ancestral populations (Native America, Africa, Europe and South-East Asia). In this case a considerable mean value of South-East Asian ancestry was found (12%), with a 7% reduction in the Native

**Table 1**  
Ancestry estimated for mtDNA and AIM-Indels in 37 samples from Paraguay, and extrapolated value for male contribution.

	African	European	Native American
mtDNA	0.075	0.075	0.850
AIM-Indels	0.092	0.599	0.309
Y Chr*	0.109	1.123	-0.232

\* Extrapolated value based on mtDNA and AIM-Indels ancestry information.

American contribution, which can be explained by the recent immigration from Korea, Japan and China. However, we cannot exclude a possible biased result due to the lack of appropriate reference populations from Native American groups.

A comparison between the ancestries obtained by mtDNA and autosomal markers was performed to infer sex biased mating. The expected paternal contribution was calculated considering similar proportions for autosomal markers and the average obtained for lineage markers (Table 1). However, when assuming this, the expected male European contribution was higher than 100%, while the Native American contribution was negative (Table 1). Our results support recent admixture between Paraguayans and other populations, probably already admixed, where the men contributed with high European ancestry. Indeed, the same type of deviation was previously observed in a sample from Antioquia, as the result of European male gene flow after an initial admixture [10].

## 4. Conclusion

The present study represents the first attempt to evaluate the genetic composition of admixed Paraguayan population. A high mtDNA diversity was observed not only in the whole sample but also inside Native American lineages, which represent the main maternal background. Moreover, the comparison of mtDNA and autosomal information highlights the demographic importance of the admixture events that succeeded the first interaction between European males and native American females during colonial times, and the role of the following gene flow in modulating the genetic composition of the current Paraguayan population.

## Financial support

Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq, Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro – FAPERJ (ref. 200.044/2016).

## Conflict of interest

None.

## Acknowledgements

We would like to thank Bettina Zimmermann, Institute of Legal Medicine, Medical University of Innsbruck, for her technical support.

## References

- [1] M. Hebbelawain, Paraguay, Bradt Travel Guides, Chalfont St Peter, 2010.
- [2] P. Lambert, A. Nicklin (Eds.), *The Paraguay Reader: History, Culture, Politics*, Duke University Press Books, Durham and London, 2012.
- [3] P.S. Walsh, D.A. Metzger, R. Higuchi, Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material, *BioTechniques* 10 (1991) 505–513.
- [4] 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.

- [5] W. Parsons, L. Guanella, D.R. Haros, et al., DNA commission of the international society for forensic genetics: revised and extended guidelines for mitochondrial DNA typing, *forensic sci. int. Genet.* 13 (2014) 134–142.
- [6] M. van Oven, M. Kayser, Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation, *Hum. Mutat.* 30 (2009) E386–94.
- [7] W. Parsons, A. Dirr, EMPOP-a forensic mtDNA database, *Forensic Sci. Int. Genet.* 1 (2007) 88–92.
- [8] L. Excoffier, G. Laval, S. Schneider, Adequin (version 3.0), an integrated software package for population genetics data analysis, *Evol. Bioinform. Online* 1 (2005) 47–50.
- [9] 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.
- [10] W. Rojas, M.V. Parra, O. Campo, et al., Genetic make up and structure of Colombian populations by means of uniparental and biparental DNA markers, *Am. J. Phys. Anthropol.* 143 (2010) 13–20.

### Artigo 3

Título traduzido: **A herança materna do Alto Paraná revelada por sequências completas de mitogenomas**

Resumo traduzido:

Grande parte dos estudos existentes que descrevam linhagens maternas em populações da América do Sul é restrita à região controle (CR) ou, para algumas regiões geográficas, o número de amostras estudadas não representa adequadamente a diversidade existente. Este é o caso dos estudos de DNA mitocondrial (mtDNA) no Paraguai, que são limitados aos segmentos HVSI de dois grupos nativos. Para ultrapassar esta deficiência, analisámos mitogenomas de 105 indivíduos residentes no Alto Paraná, o segundo departamento mais populoso do país. A molécula foi sequenciada no aparelho Ion S5 com o kit *Precision ID mtDNA Whole Genome Panel*. A diversidade haplotípica e o poder de exclusão são maiores quando comparados com os valores da CR, destacando a importância do estudo de mitogenomas no ramo da genética forense. Grande parte dos haplótipos pertence às linhagens nativo-americanas A, B, C e D. A análise de máxima parcimônia utilizando dados de mitogenoma publicados no Projeto 1000 genomas revelou um elevado número de novos ramos dos haplogrupos nativo-americanos no Paraguai. Nenhum dos haplótipos encontrados no Alto Paraná coincidem com os das restantes amostras da América do Sul utilizadas para comparação: populações miscigenadas da Colômbia, Peru e Equador e nativas da Colômbia e Equador. A análise de distâncias genéticas  $F_{ST}$  revelou que a composição genética nativa do Alto Paraná está numa posição intermédia entre os grupos da Amazônia e as populações miscigenadas do Peru e Equador. Estes resultados suportam a teoria sobre a origem Amazônica dos Tupi-Guarani e, ao mesmo tempo, revelam alguma influência de outros grupos linguísticos.



Contents lists available at ScienceDirect

## Forensic Science International: Genetics

journal homepage: [www.elsevier.com/locate/fsigen](http://www.elsevier.com/locate/fsigen)

## Research paper

## The maternal inheritance of Alto Paraná revealed by full mitogenome sequences



Filipa Simão<sup>a</sup>, Christina Strobl<sup>b</sup>, Carlos Vullo<sup>c</sup>, Laura Catelli<sup>c</sup>, Patrícia Machado<sup>d</sup>, Nicole Huber<sup>b</sup>, Lisa Schnaller<sup>b</sup>, Gabriela Huber<sup>b</sup>, Catarina Xavier<sup>b</sup>, Elizeu F. Carvalho<sup>a</sup>, Leonor Gusmão<sup>a</sup>, Walther Parson<sup>b,e,\*</sup>

<sup>a</sup>Laboratório de Diagnóstico por DNA (LDD), Universidade do Estado do Rio de Janeiro, Rio de Janeiro, Brazil<sup>b</sup>Institute of Legal Medicine, Medical University of Innsbruck, Innsbruck, Austria<sup>c</sup>DNA Forensic Laboratory, Argentinean Forensic Anthropology team (EAAT), Córdoba, Argentina<sup>d</sup>Legislación Armada, Universidad Católica, Montevideo, Uruguay<sup>e</sup>Forensic Science Program, The Pennsylvania State University, PA, USA

## ARTICLE INFO

**Keywords:**  
Mitochondrial DNA  
Mitogenomes  
Haplogroups  
EMPOP  
Forensic science

## ABSTRACT

Most studies on maternal lineages of South American populations are restricted to control region (CR) markers and, for some geographical regions, the number of studied samples does not adequately represent the existing diversity. This is the case of mitochondrial DNA (mtDNA) studies on Paraguay that are limited to two Native ethnic groups. To overcome this deficiency, we analysed the mitogenomes from 105 individuals living in Alto Paraná, the second most populated department of the country. Using the Precision ID mtDNA Whole Genome Panel, the molecule was sequenced on Ion S5. The majority of the haplotypes belong to the Native American lineages A, B, C and D. Analyses of maximum parsimony using mitogenome data retrieved from publications and in The 1000 Genomes Project showed a high number of new native American subclades in Paraguay. Also, none of the haplotypes found in Alto Paraná match the remaining South American samples, which include admixed populations from Colombia, Peru and Ecuador, and natives from Colombia and Ecuador. *FST* genetic distance analysis showed that the native genetic background of Alto Paraná has an intermediate position between the Amazonian groups and the admixed populations from Peru and Ecuador, supporting the theory about the Amazonian origin of the Tupí-Guarani and, at the same time, showing the influence of other linguistic groups.

## 1. Introduction

Paraguay is a South American country separated into two demographic regions by the Paraná river: the Eastern or Oriental region (also known as Panamá) and the Western or Occidental region (also known as Chaco). Before the arrival of Spanish settlers in the beginning of the 16<sup>th</sup> century the Eastern region of Paraguay was occupied mainly by the Guaraní people, while the Western region comprised at least five different linguistic groups [1]. Except during the Jesuit missions period (from the beginning of 17<sup>th</sup> to the mid of 18<sup>th</sup> centuries), Paraguay remained geographically, culturally and politically separated from other economic forces of the period, due to lack of mineral wealth [2]. During the "War of the Triple Alliance" (fought against Argentina, Brazil and Uruguay, between 1865 and 1870), the country suffered a reduction on population size to less than half [3]. After that, the immigration to the territory was encouraged and, shortly, a high number

of Europeans and Asians (from Italy, Germany, France, Spain, Korea and Japan) arrived in Paraguay. After 1960, immigrants also came from the neighbouring countries, namely Brazil and Argentina [4]. Currently, Paraguay comprises 17 departments, 14 of which are located in the Eastern region where 98% of the population lives (Fig. 1a). Alto Paraná, located in east of the Eastern region, is the second most populated department of the country.

In the past decades, it became common to resort to different types of genetic markers to investigate complex forensic case studies. Because of its high copy number and low susceptibility to degradation it is possible to obtain genetic information from mitochondrial DNA (mtDNA) in cases where the typing of autosomal markers cannot be accomplished. Currently, most laboratories use Sanger sequencing techniques to study the mtDNA variability of the control region (CR). However, recently, massive parallel sequencing (MPS) platforms have been applied in different scientific fields, allowing a more in-depth and high-volume

\* Corresponding author at: Institute of Legal Medicine, Medical University of Innsbruck, Innsbruck, Austria.  
E-mail address: [walther.parson@i-med.ac.at](mailto:walther.parson@i-med.ac.at) (W. Parson).

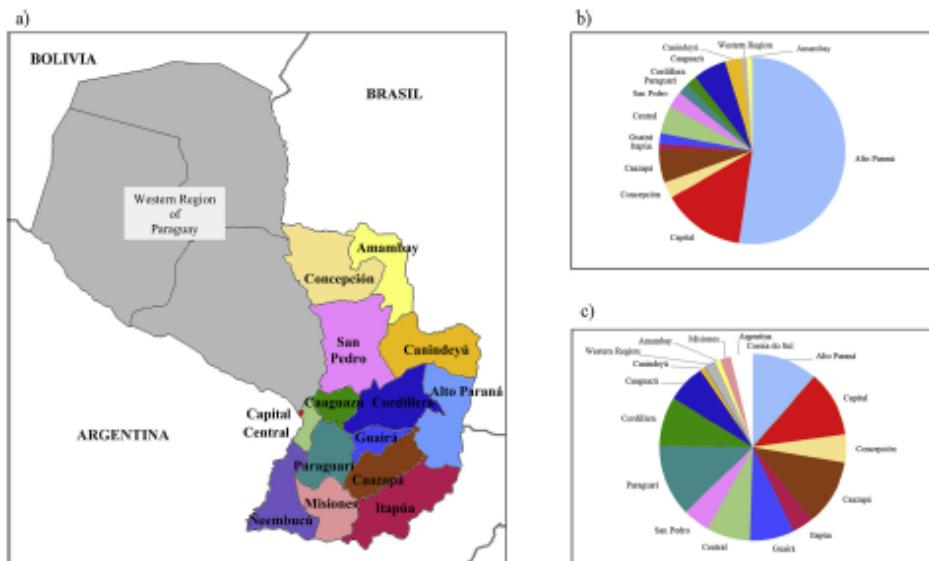


Fig. 1. a) Map of Paraguay and representation of the 17 departments. b) Distribution of participants birth place. c) Distribution of mother's participants birth place.

genotyping. Thus, it is now possible to analyse the complete mitochondrial molecule (=mitogenome) in a more efficient way with less time and money consumption.

Few studies have been published on the analysis of the maternal ancestry of Paraguay and, the existing, only include information from HVSI in native groups [5,6]. Dornelles et al. [5] sequenced the HVSI region (between positions 16024 and 16384) for 34 Ayoreo natives from Chaco. Approximately 80% of the samples belonged to haplogroup C and the remaining were from haplogroup D. Schmitt et al. [6] studied the HVSI segment in 64 individuals from the Aché ethnic group (Tupi-Guarani linguistic family) from Eastern Paraguay. In contrast with what was observed in the natives from the Western region, all haplotypes belonged to the haplogroups B (90%) and A (10%). Both population samples presented very low haplotype diversity (0.27 in the Ayoreo and 0.25 in the Aché).

With this work, we aimed to contribute to the characterization of the maternal genetic background of the admixed population of Alto Paraná, through the analysis of mitogenomes using the Precision ID mtDNA Whole Genome Panel and Ion S5 (Thermo Fisher Scientific) technology.

## 2. Material and methods

### 2.1. Population sample

A total of 105 samples were collected from unrelated male individuals living in Alto Paraná (Paraguay). Blood stains were stored in FTA paper and DNA was extracted using the Chelex® 100 (BioRad) protocol described by Walsh et al. [7]. The ethical principles of the 2000 Helsinki Declaration of the World Medical Association (<http://www.who.int/evidence/b3.htm>) were followed and participants signed a written informed consent. This study was approved by the Committee of Ethics in clinical research of the Social Prevision Institute, Asuncion, Paraguay.

### 2.2. Mitogenome typing

Library preparation for the entire mitogenome was performed with Precision ID mtDNA Whole Genome Panel and Ampliceq Library kit. In this approach two pools of 81 primers were used for mtDNA amplification. A total of 162 amplicons were generated for each sample, with an average size of 163 bp and an amplicon overlap between the 2 pools of 11 bp. Since we were dealing with non-degraded samples, the conservative method was used, and the two pools were amplified separately in 10 µl reactions that were afterwards mixed together. The target region amplification conditions were an initial enzyme activation at 99 °C during 2 min, followed by 18 cycles of denaturation at 99 °C for 15 s, annealing and extension at 60 °C for 4 min, and a final hold of 10 °C. After primer digestion with FuPa reagent, the IonCode adapters were ligated to each sample. For reaction clean-up and adapter dimers removal the AMPure XP reagent was used. The quantification was performed using the Ion Library TaqMan Quantitation kit (Thermo Fisher Scientific). To balance the sample input concentration, the samples were diluted to a concentration of 30 pM. Chips were loaded automatically with Ion Chef Instrument and sequenced using Ion S5 instrument (Thermo Fisher Scientific). Sequences were compared to the revised Cambridge Reference Sequence (rCRS) [8] with the software IGV v 2.4.10 [9,10] following the International Society for Forensic Genetics (ISPG) guidelines for mtDNA nomenclature [11]. The sequences were uploaded onto GenBank under accession numbers MH981823 - MH981927.

### 2.3. Haplotype assignment and statistical analysis

Haplotype assignment was performed on EMPOP mtDNA database v4/R11 [12], and haplotype frequencies were calculated by direct counting. Diversity indexes and genetic distances were calculated with Arlequin software [13]. Pairwise genetic distances were visualized in two-dimensional space using the multidimensional scaling (MDS) analysis included in the software STATISTICA, ver.8.0 ([www.statsoft.com](http://www.statsoft.com)), after replacing the negative  $F_{ST}$  values by zero. The maximum parsimony phylogenetic haplotype trees were obtained with mtPhyl

software (<https://sites.google.com/site/mtphyl/home>) and corrected manually based on PhyloTree build 17 [14]. Indels positions (except 249del, 290-291del and 8281-8289del), substitution at positions 16182, 16183 and 16519 and heteroplasmic positions were not considered for phylogenetic reconstruction. Moreover, for an easier and simpler visualization of the trees, only branches including samples from Paraguay were represented.

The data for mitogenomes in South America populations used in this study for comparison purposes are available on GeneBank and The 1000 Genomes Project (1000genomes) (<http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/>) [15] (see Supplementary Table S1).

### 3. Results and discussion

#### 3.1. Quality control

Mitogenome haplotypes were evaluated during EMPOP quality control [12] using logistic, plausibility and phylogenetic checks. All variants were identified as such when the coverage was > 10 (except in 7 samples with coverage between 5 and 10, in some positions). Five samples had coverage gaps, with no sequence information: sample 23,000,405 had a coverage gap between positions 10898–10962; sample 23,000,406 had a coverage gap between 14764–16767; sample 23,000,411 between positions 7793–7855; sample 23,000,513 between positions 285–287; sample 23,000,519 between positions 513–516 and 14024–14065; and sample 23,000,529 between positions 10425–10465. One sample had no information at position 286. Data was submitted to EMPOP and will be available for forensic searches under the accession number EMP00728.

#### 3.2. Genetic diversity

Haplotypes and haplogroups found in the 105 unrelated individuals living in Alto Paraná, Paraguay, for full mitogenome sequences (1–16569) are described in Supplementary Table S2. Since homopolymeric tracts can result in ambiguous haplotype readings, the diversity indexes were calculated excluding indels between positions 303–316, 512–525, 568–577, 16184–16194 and 5895–5900. The haplotype diversity (HD) obtained for the 105 samples was 0.9978 ( $\pm 0.002$ ), with a mean number of pairwise differences (MNPD) of 46.70 (Table 1). When considering the full sequences from the 105 Paraguayan mitogenomes, 84 were unique, 9 were shared between two individuals and three individuals shared the same haplotype. It is worth noting that in eight cases of shared haplotypes the individuals were born in different cities, although in one of these cases the mothers were born in the same city. In another case the two individuals and their mothers were born in the same city. For the three shared haplotypes, the individuals were born in different departments, but their mothers were born in the same city. The results obtained for a set of 25 autosomal STRs, excluded the possibility of first- and second-degree kinship among the individuals (data not shown).

For comparison purposes, values of haplotype diversity and mean number of pairwise differences were also calculated for the admixed populations from Peru [15] and Colombia [15] using the same parameters (Table 1). Paraguay presented the highest mean number of pairwise differences, although the haplotype diversity was higher in Peru. When only the native haplogroups in Paraguay were considered, the haplotype diversity was higher than for most native populations used for comparison (Table 1).

The high diversity found in Alto Paraná (for the total sample and for the subsample of native lineages) can be consequence of a high mobility to Paraguay that is known to have occurred after the War of Triple Alliance. Inside country displacements could also have contributed for the high diversity, especially concerning the native gene pool. In fact, detailed information provided by the individuals show that only about 50% were born in Alto Paraná, while the remaining were born in

different departments of the Eastern region of Paraguay (Fig. 1b). Also, mother's birth place information shows that only 11% of the mothers were born in Alto Paraná (Fig. 1c).

According to ISFG [16] and SWGDAM [17] guidelines for mtDNA comparisons in forensic casework, two or more differences (in the CR) are necessary for excluding non-maternally related individuals. When extending this rule to the mitogenomes, the exclusion power (mtCE) in the 105 Paraguayan samples is 0.9930 (including homopolymeric tracts). When only the CR is considered, the values of haplotype diversity and mtCE decrease (from 0.9978 to 0.9938 and from 0.9930 to 0.9877, respectively), with a higher number of shared haplotypes: 13 haplotypes were shared between 2 individuals, 3 haplotypes were shared between 3 individuals and 2 haplotypes were shared between 4 individuals.

#### 3.3. Haplotype classification using CR vs. mitogenome information

When considering the CR information alone, a different haplogroup classification is obtained in 26% of the samples (Table 2). Differences were due to three situations: (i) haplogroup assignment was only possible with coding region polymorphisms, for example, the positions 3547 G, 4977 C, 6473 T, 9950 C and 11,177 T are required for haplogroup B2 classification; (ii) the branch was changed because coding region positions were different from those required for the classification based on the CR, for example, one sample classified as C1d2a with CR turned out to be C1d1 due to the lack of 10,834 T and 593 C and the presence of 7697 A; (iii) although a CR mutation was missing, other coding region mutations allow a deeper classification. This is the example of sample 23,000,531 that does not have the position A263 but has 4242 T, 7013 A, 9557 T, 12,454 A required for C1b2 classification.

#### 3.4. Haplotype distribution in Alto Paraná

Approximately 86% of the haplotypes belong to the Native American haplogroups A, B, C and D. The European and African haplogroups have similar frequencies in the studied sample, each representing about 7% of the sequences (Supplementary Fig. S1).

Compared to the neighbouring admixed populations (Supplementary Fig. S1), the proportions of Native American and European maternal lineages in Alto Paraná are more similar to those found in populations from Chile. Although Paraguay has higher non-Native ancestry than Bolivia, it is less admixed than the other neighbouring countries, namely Brazil and Argentina. The frequency of African haplogroups in Paraguay can be considered relatively high, taking into account historical data that indicate the arrival of a low number of slaves during the colonial period. This African ancestry can most likely be explained by a recent immigration from Brazil, which became especially important after 1960.

#### 3.5. Native American lineages

All Native American lineages found in our sample (A2, B2, C1b, C1d, C1d1, D1 and D4h3a) were previously described as being "pan-American", since they are found across the double continent, except for one lineage (sample 23,000,402) belonging to haplogroup A+152 + 16,362 + 200. This haplogroup is not present in the EMPOP database, suggesting that this lineage is either local or rare in South America.

The only two studies on maternal lineages of native groups from Paraguay show that haplogroup C is highly prevalent in Ayoreo from Chaco and haplogroup B in the Aché from Eastern Paraguay [5,6]. These haplogroups have high frequencies in Alto Paraná (29% each). On the other hand, none of the Paraguayan mitogenomes were included in haplogroups D1j, D1g, B2i2 and C1b13, previously described as restricted to the southern cone of South America [18,19], showing no genetic continuity between these two regions.

**Table 1**

Haplotype diversities and mean number of pairwise differences on admixed and native populations from Alto Paranaíba, Peru, Ecuador and Colombia.

Country	Ethnic group	n	HD	MNPD	Reference
Paraguay	—	105	0.998 ± 0.002	46.70 ± 20.38	This study
Colombia	—	94	0.976 ± 0.008	40.76 ± 17.86	1000genomes [15]
Peru	—	86	1.000 ± 0.002	40.08 ± 17.58	1000genomes [15]
Paraguay	Admixed <sup>a</sup>	90	0.998 ± 0.002	41.63 ± 18.23	This study
Colombia	Ach-Piapoco	24	0.77	—	Arax et al. [21]
Colombia	Cetena	8	0.89	—	Arax et al. [21]
Colombia	Cocama	17	0.98	—	Arax et al. [21]
Colombia	Coreguaje	19	0.92	—	Arax et al. [21]
Colombia	Curripaco	17	0.88	—	Arax et al. [21]
Colombia	Deneño	17	0.98	—	Arax et al. [21]
Colombia	Guayabero	35	0.88	—	Arax et al. [21]
Colombia	Igua	17	0.96	—	Arax et al. [21]
Colombia	Kamenta	11	0.96	—	Arax et al. [21]
Colombia	Mur-Uritio	26	0.99	—	Arax et al. [21]
Colombia	Nukak	16	0.64	—	Arax et al. [21]
Colombia	Other-ET	10	1.00	—	Arax et al. [21]
Colombia	Pato	14	1.00	—	Arax et al. [21]
Colombia	Pira-Wamán	13	0.96	—	Arax et al. [21]
Colombia	Pumare	19	0.85	—	Arax et al. [21]
Colombia	Saliba	16	0.86	—	Arax et al. [21]
Colombia	Silvani	16	0.87	—	Arax et al. [21]
Colombia	Siona	17	0.82	—	Arax et al. [21]
Colombia	Sitom	10	1.00	—	Arax et al. [21]
Colombia	Tanimaka	10	0.73	—	Arax et al. [21]
Colombia	Tikuna	18	0.94	—	Arax et al. [21]
Colombia	Tuka-Tatuyo	10	0.98	—	Arax et al. [21]
Colombia	Yagua	13	0.92	—	Arax et al. [21]
Colombia	Yuro-Matapi	39	0.88	—	Arax et al. [21]
Ecuador	Several <sup>b</sup>	93	0.984 ± 0.006	35.08 ± 15.41	Brandini et al. [20]
Ecuador	Admixed <sup>c</sup>	114	0.999 ± 0.001	41.73 ± 18.24	Brandini et al. [20]

<sup>a</sup> only native haplogroups.**Table 2**

Haplotype classifications with mitogenomes and control region haplotypes.

Number of samples	Haplotype		
		Mitogenome	Control Region
7	B2	B4b	(i)
3	B2e	B4b	(i)
2	L1b1a12a	L1b1a12	(i)
2	H4a1a4a	R0	(i)
2	H1el4/HV+ 16311	R0	(i)
1	C1d1	C1d + 194	(i)
1	C1d1	C1d2a	(ii)
1	L1c1b	L1c1b/d	(i)
1	L3e1a1a	L3e1a1	(i)
1	C1d1	C1d + 194	(i)
1	U2e1e1	U2e1	(i)
1	C1d1b1	C1d1b	(i)
1	R2	R2b + 152	(ii)
1	H13a1a1	H82	(i)
1	C1b2	C1b	(iii)

Differences due to: (i) haplogroup assignment was more detailed with coding region polymorphisms; (ii) the branch was changed because coding region positions were different from those required for the classification based on CR; (iii) although a CR mutation was missing, other coding region mutations allow a deeper classification.

A phylogenetic analysis was performed for haplogroups A2, B2, C1b, C1d1, D1 and D4 (Supplementary Figs. S2–S6) using mitogenomes data from South American populations including samples from Alto Paranaíba (this study), Ecuador [20], Colombia [15,21], Peru [15], Venezuela and Brazil [22]. Extensive variation was observed inside all clades in accordance to previous findings from Brandini et al. Moreover, most samples from Alto Paranaíba are placed into new derived branches not present in the data included for comparison (see detailed information in Supplementary Figs. S2–S6).

The lineages A2 and A2 + 64 represent 21% of the samples from

Alto Paranaíba, a frequency similar to that found in Native Colombians (Table 3). A search on the EMPOP database, which comprises sequences from 2165 Native Americans and 4094 samples from the general American populations (results based upon release R11, from October 16<sup>th</sup>, 2013), place most matches in central and northern American populations. The subclade A2q with a frequency of 2% in our sample is also found at low percentages in natives and admixed populations from Colombia (1% and 6% respectively) (Table 3).

Haplotype B2 represent 29% of the native mitogenomes from Alto Paranaíba. The sub-lineages B2e, B2o and B2o1a have a total frequency of 5% and were also found in Native Colombians (Table 3). Altogether, the B2h, B2i1 and B2i3a sub-lineages have a 10% frequency but were absent in the compared South American mitogenomes. In EMPOP, the haplogroup B2h appears in Brazil, Argentina and Ecuador, while B2i1 is found in a single individual from North Argentina. No haplotype in EMPOP belongs to B2h3a. The remaining B2 samples do not belong to any known sub-lineage.

The C1 lineage has a high frequency in Alto Paranaíba (29%). Apart from one sample from haplogroup C1, the remaining are distributed among two main branches: C1b and C1d1. Approximately 15% of the samples from native Colombians belong to C1b haplogroup (Table 3). The C1b2 branch is present in 6% of Paraguayan samples and was also found in 1% of natives from Colombia (Table 3). The 85 C1b2 haplotypes deposited on EMPOP are virtually restricted to Dominican Republic. A similar proportion of haplogroups C1d1 and C1d1d was found in natives from Colombia, however, no haplotypes match the ones in the Paraguayan samples (Table 3; Supplementary Fig. S4b). The C1d1b1 branch was not present in the mitogenomes used for comparative analysis, although 8 C1d1b1 haplotypes from Argentina are deposited on EMPOP.

The macrohaplogroup D is the least frequent in Alto Paranaíba (~18%), with 10 samples (11%) belonging to haplogroup D1. This haplogroup has a frequency of 15% in the mitogenome samples from Peru. The D1e and D1f1 branches found in Paraguay were not present

**Table 3**  
Native American haplogroup frequencies in populations from Alto Paranaí, Peru, Ecuador and Colombia.

Haplogroup \ n	Alto Paranaí (this study) 90	Ecuador-natives [20] 93	Ecuador-admixed [20] 115	Colombia-native [21] 439	Colombia-admixed [15] 81	Peru-admixed [15] 81
A + IS2 + 16,362 + 200	1%	0%	0%	0%	0%	0%
A2	6%	3%	3%	7%	0%	5%
A2 + 64	16%	0%	0%	12%	1%	9%
A2d	2%	0%	0%	1%	6%	0%
Other A2	0%	6%	21%	11%	42%	4%
Total A	24%	0%	0%	32%	49%	17%
B2	12%	4%	3%	9%	1%	26%
B2b3a	2%	0%	0%	0%	0%	0%
B2e	3%	0%	0%	4%	0%	0%
B2h	4%	0%	0%	0%	0%	0%
B2l1	4%	0%	0%	0%	0%	0%
B2o	1%	0%	0%	1%	0%	0%
B2a1a	1%	0%	0%	1%	0%	0%
Other B2	0%	61%	43%	2%	40%	23%
Total B	29%	0%	0%	17%	41%	49%
C1	1%	0%	0%	0%	0%	0%
C1b	1%	5%	4%	15%	0%	1%
C1b2	6%	0%	0%	1%	0%	0%
C1c	0%	0%	2%	14%	0%	0%
C1d1	7%	0%	3%	9%	1%	0%
C1d1d	3%	1%	0%	2%	1%	0%
C1d1b1	1%	0%	0%	0%	0%	0%
Other C	0%	16%	10%	1%	5%	17%
Total C	29%	0%	0%	43%	7%	19%
D1	1%	0%	3%	4%	0%	15%
D1e	4%	0%	0%	0%	0%	0%
D1f1	1%	0%	0%	0%	0%	0%
D4b3a	1%	0%	1%	0%	0%	0%
Other D	0%	0%	0%	9%	2%	1%
Total D	18%	2%	7%	10%	2%	16%

in the remaining South American mitogenomes (Table 3). However, in EMPOP 18 D1e samples appear in North America, and 16 D1f1 samples are present mainly in Brazil and Argentina.

### 3.6. Genetic distances between South American populations

For a more in-depth understanding of the maternal genetic background of Paraguay,  $F_{ST}$  genetic distances were calculated between admixed samples from Alto Paranaí, Colombia and Peru [15]. Indel positions 309.xC, 315.xC, 523-524del, 524.xAC, 573.xC, 5899.xC and 16,193.xC were not considered in this comparison. Significant differences were observed in all pairwise comparisons. Although Alto Paranaí and Colombia have similar continental ancestry profiles (Supplementary Fig. S1), a much larger distance was observed between these two samples ( $F_{ST}=0.06283$ ; P-value < 0.000005) than between Alto Paranaí and Peru ( $F_{ST}=0.02640$ ; P-value = 0.00347). After excluding the non-native haplogroups, the genetic distance between Alto Paranaí and Colombia increased by 50% while the distance to Peru did not change significantly (Supplementary Table S3). These results reveal that the distance between Alto Paranaí and Colombia is mainly due to very different native backgrounds. Indeed, the haplogroups A and B are much more prevalent in Colombia (approximately 89%) than in Alto Paranaí. Moreover, there are many A and B sub-clades that are present at high frequencies in Colombia but are not present in the Paraguayan samples (namely, A2w and B2d).

The samples of native haplogroups in admixed populations were further compared against data from Northwestern Amazonian and Ecuadorian native populations (Supplementary Table S3). No clear linguistic or geographic pattern emerged from this analysis (Fig. 2), which is in accordance with the results from Arias et al. [21]. Although multiple causes may be behind the absence of a pattern, it is well known that genetic drift will have certainly contributed to the high distances found between groups and to the low values of diversity within

populations. However, it is worth noting that the population that is closest to Alto Paranaí is the Cocama ethnic group (Fig. 2), which belongs to the same linguistic family (Tupí-Guarani) as the natives that inhabit the Eastern region of Paraguay. In addition, concerning the native component of the admixed populations, the correlation between the geographic and genetic distances improves.

Finally, it should be noted that Alto Paranaí occupies an intermediate position between Amazonian and the remaining populations, which, to a certain extent, supports the theory about the Amazonian origin of the Tupí-Guarani [23,24] and, at the same time, shows the influence of other linguistic groups in its current genetic background.

### 4. Conclusions

The present work represents the first attempt to elaborate a genetic database from mitochondrial DNA markers on an admixed population from Paraguay. The results obtained revealed high haplotype diversity, showing an increased discrimination power when extending the analysis from the control region to the mitogenome. At the same time, the study of mitogenomes allowed refining haplogroups assignment, particularly inside B2 clades. Moreover, the construction of maximum parsimony trees from South American lineages, including published data and mitogenome sequences retrieved from the 1000 genome project, allowed disclosing a large number of previously unidentified subclades. Indeed, no matches were found between haplotypes from Alto Paranaí and the compared population samples.

According to historical data, during the European colonial period, Paraguay suffered lower non-native admixture than the surrounding countries of Argentina and Brazil, which is supported by its high proportion of Native maternal background. The prevalence of some African lineages can be explained either by an African input during the slave trade or by a later immigration after the war of Triple Alliance. Large differences in African maternal vs. paternal ancestries are usually

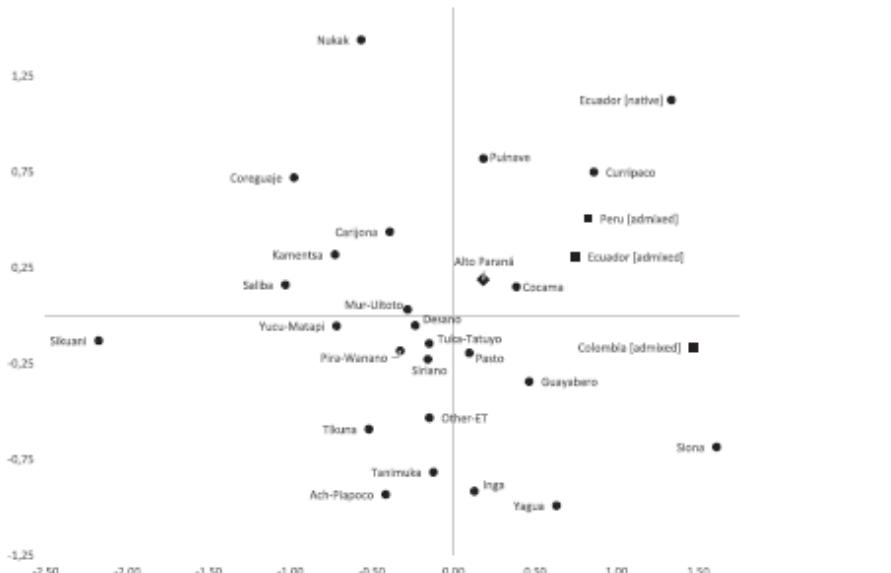


Fig. 2. MDS plot based on  $F_{ST}$  genetic distances between native mitogenomes from Alto Paranaí, Colombia, Ecuador and Peru.

observed in admixed populations from South America, due to a sex-biased mating during the colonial period. It is however expected that this difference will decrease if we assume a later African input in Paraguay, mediated mostly by men from an already admixed population (namely from Brazil) with higher African ancestry. Therefore, results on Y chromosome ancestry could be helpful to discern between the two possible scenarios.

Concerning the native maternal ancestry of Alto Paranaí, an intricate structure emerged from our results, pointing to a possible input of lineages from different sources, and not only from Tupí-Guarani groups. Nevertheless, the lack of mitogenome data from most South American populations prevents a thorough interpretation of the origin of most lineages that have been found in the population from Alto Paranaí.

#### Conflict of interests

None.

#### Acknowledgments

The authors would like to thank Leonardo Arias for kindly providing the full mitogenome data from Northwestern Amazonian samples. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brazil (CAPES) - Finance Code 001 and by a Short-Term Fellowship grant from the International Society for Forensic Genetics (ISPG). LG was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq (ref. 305330/2016-0).

#### Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.fsigene.2018.12.007>.

#### References

- [1] I. Jemmy, Y.J. Lin, *Paraguay (Cultures of the World)*, 2 edition, Benchmark Books, 2010.
- [2] M. Habibethwaite, *Paraguay (Bradt Travel Guide)*, 2 edition, (2014).
- [3] H.G. Warren, *Births of the Paraguayan Republic: The First Colonies Era, 1818–1904*, 1 edition, University of Pittsburgh Press, 1985.
- [4] P. Lambert (Eds.), *The Paraguay Reader: History, Culture, Politics*, 2012.
- [5] C.I. Domínguez, J. Battilana, N.J.R. Fagundes, L.B. Frethe, S.L. Bonatto, F.M. Salzano, Mitochondrial DNA and Alu insertions in a genetically peculiar population: the Ayoreo Indians of Bolivia and Paraguay, *Am. J. Hum. Biol.* 16 (2004) 479–488, <https://doi.org/10.1002/ajhb.20038>.
- [6] R. Schmitz, S.L. Bonatto, L.B. Frethe, V.C. Mischner, K. Hill, A.M. Hurtado, F.M. Salzano, Extremely limited mitochondrial DNA variability among the Aché Natives of Paraguay, *Am. Hum. Biol.* 19 (2004) 87–94, <https://doi.org/10.1002/093144600310001802063>.
- [7] P.S. Walsh, D.A. Metzger, R. Higuchi, Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material, *BioTechniques* 10 (1991) 505–513.
- [8] R.M. Andrew, L. Kubacka, P.F. Chinnery, R.N. Lightowler, D.M. Turnbull, N. Howell, Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA, *Nat. Genet.* 23 (1999) 147.
- [9] J.T. Robinson, H. Thorvaldsdóttir, W. Winckler, M. Guttmann, E.S. Lander, G. Getz, J.P. Meirov, Integrative genomics viewer, *Nat. Biotechnol.* (2011) 24–26.
- [10] H. Thorvaldsdóttir, J.T. Robinson, J.P. Moniz, Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration, *Brief. Bioinform.* (2013) 178–192.
- [11] W. Parsons, I. Gusmão, D.R. Hamz, Jr., Irwin, W.R., Mayr, N., Morling, E., Polakow, M., Pötz, a. Salas, P.M. Schneider, T.J. Parsons, DNA commission of the international society for forensic genetics revised and extended guidelines for mitochondrial DNA typing, *Forensic Sci. Int. Genet.* 13 (2014) 134–142, <https://doi.org/10.1016/j.fsigen.2014.07.010>.
- [12] W. Parsons, A. Diur, EMPOP-a forensic mtDNA database, *Forensic Sci. Int. Genet.* 1 (2007) 88–92, <https://doi.org/10.1016/j.fsigen.2007.01.018>.
- [13] L. Excoffier, G. Laval, S. Schneider, Arlequin (version 3.0), an integrated software package for population genetics data analysis, *Biol. Bioinform. Online.* 1 (2005) 47–50.
- [14] M. van Oven, M. Kayser, Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation, *Hum. Mutat.* 30 (2009) E386–94, <https://doi.org/10.1002/humu.20921>.
- [15] A. Auton, G.R. Abecasis, D.M. Altshuler, R.M. Durbin, A global reference for human genetic variation, *Nature* 526 (2015) 68–74, <https://doi.org/10.1038/nature15393>.
- [16] A. Carrasco, B. Bar, P. Lincoln, W. Mayr, N. Morling, B. Ohlson, P. Schneider, B. Budowle, B. Brinkmann, P. Gill, M. Holland, G. Tally, M. Wilson, DNA commission of the international society for forensic genetics: guidelines for

- mitochondrial DNA typing, *Forensic Sci. Int.* 110 (2000) 17–25 <http://www.ncbi.nlm.nih.gov/pmc/articles/1105405/>.
- [17] Scientific Working Group on DNA Analysis Methods, Interpretation Guidelines for MitochondrialDNA Analysis by Forensic DNA Testing Laboratories, 2013, pp. 1–23.
- [18] M. Bodner, U.A. Perego, G. Huber, I. Fendi, A.W. Rock, B. Zinnemann, A. Olivieri, A. Gómez-Carballa, H. Lancioni, N. Augerhofer, M.C. Bobillo, D. Conch, S.R. Woodward, A. Sale, A. Achilli, A. Torroni, H.-J. Bandelt, W. Pfenning, Rapid coastal spread of first Americans: new insights from South America's Southern Cone mitochondrial genomes, *Genome Res.* 22 (2012) 811–820, <https://doi.org/10.1101/g.131722.1111>.
- [19] M. de Saint Pierre, F. Gandini, U.A. Perego, M. Bodner, A. Gómez-Carballa, D. Conch, N. Augerhofer, S.R. Woodward, O. Semino, A. Sale, W. Pfenning, M. Montagna, A. Achilli, A. Torroni, A. Olivieri, Arrived of paleo-indians to the southern coast of South America: new clues from Mitogenomes, *PLoS One* 7 (2012) 1–9, <https://doi.org/10.1371/journal.pone.0051311>.
- [20] S. Brandini, P. Bergamaschi, M. Fernando Cerna, F. Gandini, F. Rustaroli, E. Bertolini, C. Oreda, L. Ferretti, A. Gómez-Carballa, V. Battaglia, A. Sale, O. Semino, A. Achilli, A. Olivieri, A. Torroni, The Paleo-Indian entry into South America according to mitogenomes, *Mol. Biol. Evol.* 35 (2018) 299–311, <https://doi.org/10.1093/molbev/msx267>.
- [21] I. Aris, C. Barberi, G. Barreto, M. Stenskaing, B. Pakendorf, High-resolution mitochondrialDNA analysis sheds light on human diversity, cultural interactions, and population mobility in Northwestern Amazonia, *Am. J. Phys. Anthropol.* 165 (2018) 238–255, <https://doi.org/10.1002/ajpa.23346>.
- [22] N.J.R. Segurado, R. Keine, R. Eickert, A.C.S. Valls, Ma.R. Bogo, F.M. Salzano, D.G. Smith, Wilson A.Silva Jr, M. Zago, A.K. Ribeiro-dos-Santos, S.E.R. Santos, M.L. Penna-Eder, S.L. Bonatto, Mitochondrial population genetics supports a single pre-Clavus origin with a coastal route for the peopling of the Americas, *Am. J. Hum. Genet.* 83 (2008) 583–592, <https://doi.org/10.1086/587071>.
- [23] W.A. Neves, D.V. Bernardo, M. Okamura, T. Fernanda De Almeida, A.M. Stenzel, Origin and dispersion of the Tupiguanan: what does cranial morphology say? *Bol. Do Mus. Para. Emílio Goeldi Ciências Humanas* 6 (2011) 95–122, <https://doi.org/10.1590/S1981-81222011000100007>.
- [24] H. Silverman, W. Isbell (Eds.), *Handbook of South American Archaeology*, Springer, 2008.

### **3.3 Caracterização do patrimônio genético de origem materna de indivíduos pertencentes ao grupo linguístico Ashaninka do Peru através da análise do mtDNA**

O manuscrito referente aos resultados do sequenciamento da região controle do mtDNA e dos mitogenomas de indivíduos Ashaninka residentes no Peru encontra-se em preparação (Artigo 4) e pretende-se submeter para publicação à revista PlosOne. O material suplementar está disposto no Apêndice E.

Os resultados preliminares obtidos do sequenciamento dos 20 mitogenomas da linhagem B2 + 16051G + 16129A + 152C dos indivíduos Ashaninka, Peru, foram divulgados em formato de pôster no *27th Congress of the International Society for Forensic Genetics* (Seoul, Coreia do Sul, Setembro 2017), com o título “*Tracing Arawakan Southern Dispersal: Clues from Mitogenome Sequencing of Southern Amerinds*” (Apêndice F).

## Artigo 4

Título traduzido: **Herança maternal do grupo nativo Ashaninka do Peru**

Resumo traduzido:

Os Ashaninka pertencem à família linguística Arawak e são o principal grupo indígena amazônico peruano. Estão, atualmente, dispersos ao longo de vários departamentos do país. Um grupo de 171 amostras da população Ashaninka, residente na região do Pasco, foi sequenciada para a região controle total do mtDNA (entre as posições 16024 e 576), através da metodologia de Sanger. Foi observada uma baixa diversidade halotípica, com todas as amostras pertencendo a haplogrupos nativo-americanos (A, B, C e D). Para uma compreensão mais profunda das migrações dos Arawak na América do Sul, os haplótipos B2 + 16051G + 16129A + 152C presentes na população Ashaninka foram sequenciados para a molécula completa, por MPS. A investigação da distribuição de linhagens B2 + 16051G + 152C em populações sul-americanas indicou uma possível rota de migração da família linguística Arawak em direção ao sul.

The maternal inheritance of the *Ashaninka native group from Peru*

F. Simão<sup>1</sup>, C. Xavier<sup>2</sup>, D.H. Tineo<sup>3</sup>, M. Bodner<sup>2</sup>, C. Strobl<sup>2</sup>, E. F. Carvalho<sup>1</sup>, W. Parson<sup>2,4</sup>

L. Gusmão<sup>1</sup>

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

<sup>2</sup>Institute of Legal Medicine, Medical University of Innsbruck, Innsbruck, Austria

<sup>3</sup>Instituto de Medicina Legal, Universidad Nacional Mayor de San Marcos, Lima, Peru

<sup>7</sup>Forensic Science Program, The Pennsylvania State University, PA, USA

## Abstract

The Ashaninka, which belong to the Arawak linguistic family, are the major Peruvian Amazonian native group and are currently spread across several departments of the country. A group of 171 samples from the Ashaninka population living in the Pasco region was sequenced for the total control region of the mtDNA (between positions 16024 and 576) using Sanger methodology. A low haplotype diversity was obtained and only native American haplogroups (A, B, C and D) were found. For a deeper understanding of early Arawak migrations across South America, the B2 + 16051G + 16129A + 152C haplotypes present in the Ashaninka population were sequenced for the complete mtDNA molecule using MPS technology. The investigation of the distribution of B2 + 16051G + 152C lineages in South American populations indicated a possible south migration route of the Arawakan linguistic family.

## Introduction

Like most South American populations, Peru is a multi-ethnic country, due to admixture processes between Amerindians, Europeans, Africans and Asians. The great majority of the native populations of Peru occupies plateau regions along the coast (HUNEFELDT, 2010). However, a few groups are fixed in the rainforest, distributed among four main linguistic families: Arawak, Carib, Tupi and Je. It is thought that the Arawak family was responsible for one of the major diasporas in South America, in pre-colonial times. Upon the arrival of the European colonizers, they were scattered across various regions of the subcontinent. Linguistic and anthropological studies situate western Amazonia (in the border between Peru and Brazil) as the most probable origin of the Arawakan

language (WALKER; RIBEIRO, 2011).. Afterwards Arawak are thought to have migrated alongside riverbanks throughout the continent, spreading their cultural and agricultural traditions. The Ashaninka belong to the Arawak linguistic family and are the largest native Amazonian group in Peru. They are currently distributed in several Peruvian departments: Ayacucho, Cusco, Huanuco, Junin, Pasco and Lima. According to the Peruvian sense of 2009, about 88,000 Ashaninka lived in the country at that time (HUNEFELDT; HARRIS, 2004; MCEWAN, 2006).

Few studies were performed to disclose the genetic composition of Ashaninka native populations. A small group of 20 samples were genotyped for mtDNA, 8 for Y-chromosomal markers and 14 were genotyped for autosomal STRs (DI CORCIA et al., 2017).

This study aims at contributing to a deeper understanding of this under-described population, by investigating the maternal lineages distribution in a large sample of Ashaninka people, living in Peru. An increase in the B2 haplogroup resolution, by sequencing the full mitogenome, will allow a better knowledge of the Ashaninka population structure as well as of their migratory movements at a minor scale.

## **Material and methods**

### Sampling

A total of 171 blood samples were collected from non-related individuals (55 men and 116 woman), belonging to 41 communities located in the banks of Pichis and Palcazu rivers, in the district of Puerto Bermúdez, Pasco region, Peru. Samples were extracted using either a phenol-chloroform protocol or a Chelex based protocol and quantified using an in-house real time qPCR method [6].

### Sanger sequencing

The control region of mitochondrial DNA (between positions 16024 and 576) was amplified in a final reaction volume of 25 µl, including: 15.8 µl of water, 2.5 µl buffer 10x, 2.5 µl BSA 2.5mg/ml, 2 µl dNTPs 10 mM, 0.3 µl of primer L15851 (5' ATC TCC CTA ATT GAA AAC AAA ATA CTC AAA 3' ) 10 µM, 0.3 µl of primer H639 (5'-GGG TGA TGT GAG CCC GTC TA -3') 10 µM, 0.2 µl of Advantage polymerase (Clotech) 5U/µl and 1 µl of extracted DNA (2-5 ng/µl). The PCR conditions were an initial denaturation at 95 °C for 2 min, followed by 35 cycles of 95 °C for 15 s, 56 °C for

30 s and 72 °C for 90 s, ending with a final extension at 72 °C for 10 min. The purification was performed with 2 µl of ExoSAP-IT (USB Affymetrix) for each 5 µl of amplified product. For sequencing, the BigDye Terminator v1.1 cycle Sequencing kit (Applied Biosystems) was used together with the primers L15971 (5' TTA ACT CCA CCA TTA GCA CC 3'), L15 (5' CAC CCT ATT AAC CAC TCA CG 3'), L16268 (5' CAC TAG GAT ACC AAC AAA CC 3'), H16 (5' TGA TAG ACC TGT GAT CCA TCG TGA 3') and H639 (5'- GGG TGA TGT GAG CCC GTC TA -3'). For a final volume of 10 µl, 2 µl of BigDye, 2 µl of Buffer, 0.3 µl of *primer* 10 µM and 2 µl of purified product were used. The thermocycling conditions were as follow: initial denaturation at 96 °C for 1 min, followed by 25 cycles of 95 °C for 15 s, 50 °C for 5 s and 60 °C for 4 min. For final cleaning, the sequences were purified using Sephadex (Illustra Sephadex DNA Grade; GE Healthcare). Separation and detection were achieved by capillary electrophoresis on an ABI 3500 (Applied Biosystems).

As recommended by Parson and Bandelt (PARSON; BANDELT, 2007), in cases where double coverage was not possible with the primers indicated above, sequencing was also performed with primers

F314 (5' CCG CTT CTG GCC ACA GCA CT 3'), R159 (5' AAA TAA TAG GAT GAG GCA GGA ATC 3'), R484 (5' TGA GAT TAG TAG TAT GGG AG 3') and F15851 (5' ATC TCC CTA ATT GAA AAC AAA ATA CTC AAA 3').

The CR mtDNA sequences were compared to the revised Cambridge Reference Sequence (rCRS) (ANDREWS et al., 1999) on the software Sequencher v.4.9 (Gene Codes), and the haplotype classification was carried out following the guidelines proposed by the International Society for Forensic Genetics (ISFG) (PARSON et al., 2014b)

### Mitogenome sequencing

Libraries were built using the Precision ID mtDNA Whole Genome Panel and Ampliseq Library kits (Thermo Fisher Scientific). The conservative method was used, and two pools were amplified in two 10 µl reactions that were afterwards mixed together. The amplification conditions were an initial enzyme activation at 99 °C during 2 min, followed by 18 cycles of denaturation at 99 °C for 15 s, annealing and extension at 60 °C for 4 min. The primers were digested with FuPA reagent. Afterwards the IonCode adapters were attached to each sample. The reaction was clean up and the dimers removed with AMPure XP reagent. The amplified products were quantified using the

Ion Library TaqMan Quantitation kit (Thermo Fisher Scientific) and samples were diluted to a concentration of 30 pM. Chips were loaded automatically with Ion Chef Instrument and sequenced using Ion S5 instrument (Thermo Fisher Scientific).

The haplotypes were obtained with the Ion Torrent Variant Caller v. 5.2.1.38 plugin (TFS), and further checked with IGV (Integrative Genome Viewer) software.

#### Haplogroup assignment and statistical analysis

The haplogroups were assigned according to Phylotree (build 17) [8] via EMPOP. The EMPOP tools “Network” and “EMPcheck” were used to detect potential errors in the dataset (PARSON; DÜR, 2007; ZIMMERMANN et al., 2014). Data were submitted to EMPOP for quality control (PARSON; DÜR, 2007).

Haplogroup frequencies were calculated by direct counting. The alignment was performed on Haplosearch (FREGEL; DELGADO, 2011) and adapted with the alignment correction of EMPOP (PARSON; DÜR, 2007). The statistical parameters of haplotype diversity were calculated using the Arlequin software (EXCOFFIER; LAVAL; SCHNEIDER, 2005).

## **Results and discussion**

#### mtDNA diversity

The control region sequences obtained in the analysis of Ashaninka population showed a low haplotype diversity ( $0.9121 \pm 0.0100$ ), with only 41 different haplotypes (Supplementary Table 1). This is in accordance with the values found for other isolated native South American populations. The most frequent haplotype was 16093C 16223T 16298C 16325C 16327T 73G 146C 249del 263G 290del 291del 309.1C 315.1C 489C 493G 523del 524del, observed in 30 samples.

#### Native haplogroups in Ashaninka

All major Native American lineages (except haplogroup X) were found in the sample. Haplogroup D1 was the most frequent with a percentage of 33 %, followed by haplogroups A2 (25 %), C1 (22 %) and B2 (20 %). With only 1 % frequency in our sample, haplogroup B4 was just found in 1 sample. Corroborating previous results, no clear linguistic or geographic patterns emerged when adding this new data to the already

available information from native groups from South America (ARIAS et al., 2018; SIMÃO et al., 2018).

#### Maternal vs. Paternal ancestries

In 2015, Tineo *et al.* (TINEO et al., 2015) studied a total of 58 samples from the Ashaninka linguistic group from Peru, for 27 Y chromosome specific STR markers included in the Y-filer Plus kit (Thermo Fisher). To determine the possible geographic origin of the paternal lineages, these samples were further genotyped for Y-SNPs. Most profiles were classified inside haplogroup Q, the most frequent in native-Americans. A single haplotype belonged to the African haplogroup E-M2\*(xU209, M191), and 4 were from European ancestry, namely from haplogroups R1b-P25, G-M201 and K- M9\*(xTat,92R7, M70) (unpublished data). From this group of samples, 55 overlap those studied for mtDNA markers, all belonging to native-American haplogroups. Although some native South American populations suffered a strong admixture with European and African individuals, others appear to have maintained their original genetic background. The Ashaninka population studied in this work, seems to have remained relatively isolated, with just a small input of non-native male lineages. This can be explained by its remote location in the tropical forest and a sex biased admixture, not detected in their maternal inheritance.

#### Tracing Arawakan Southern Dispersal

A previous study of the mtDNA control region of samples from Terena (Brazil) and Shipibo-Conibo (Peru) native groups allow detecting the presence of the B2+16051G+16360T+152C lineage in these populations (unpublished data). The deeper investigation of this particular lineage at the highest resolution possible (the full mtDNA genome) confirm that the lineages present in Peru and Brazil are very similar. The same lineage was detected along the Peruvian lowlands, the Titicaca basin and the Bolivian Moxos, when only the CR or partial CR was investigated (BERT et al., 2009). In the present study, a closely related B2 lineage was identified in the Peruvian Ashaninka (B2+16051G+ 16129A+152C), which was further sequenced for the full mtDNA genome. This lineage (full genome) was also found in other Peruvian samples, and previously described (only CR) in Matsiguenga and Yanesha tribes, all pre-Andine Arawakans (Fig. 1), indicating a regional diversification of this specific lineage. Particularly in western Amazonia/Bolivian Llanos and Wetlands, the B2 + 16051G +

16360T + 7786C +152C stands out as a possible reminiscence of an Arawakan southern migration. Characterized as riverine groups with developed watercraft techniques, a southern migration alongside the Ucayali and Urumbamba rivers until reaching Madre de Dios River and finally the Pantanal wetlands in Brazil would not be surprising.

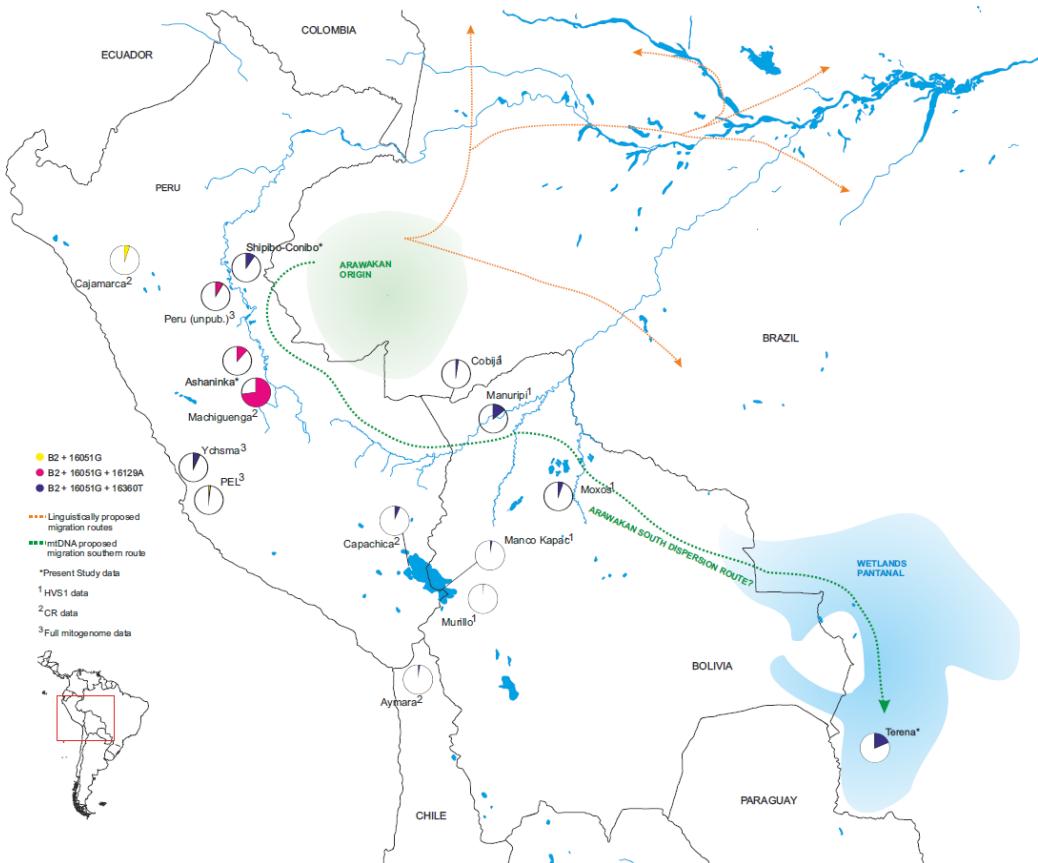


Fig. 1 B2+16051G (yellow), B2+16051G+16129A (pink), B2+16051G+16360T (dark blue) lineages distribution across the Peruvian lowlands, lake Titicaca, Bolivian Moxos and Brazilian Wetlands. Samples in bold and with a \* are the present study data, other groups have been gathered from literature [all of them]. Orange arrows represent the linguistically proposed migration routes by the Arawakan, the green arrow is the migration route proposed by the mtDNA results. Shaded green area is the linguistic proposed location of origin of the Arawak language, shaded blue area represents the wetlands covered area. Other populations' data retrieved from (AUTON et al., 2015; DE SAINT PIERRE et al., 2012b; SANDOVAL et al., 2013; TABOADA-ECHALAR et al., 2013; VALVERDE et al., 2016).

## Conclusions

According to historical data, the native populations living in the Amazonian rainforest have remained considerable isolated, with no or reduced European and African influxes. Our results corroborate the low genetic input of non-native maternal lineages in the Ashaninka native group. However, even if low, it is possible to find some European and

African paternal input. Further data on autosomal markers would help to clarify the extend of this non-native influence on the Ashaninka.

The study of mitogenomes for the B2+16051G+16360T+152C lineages help to identify a possible southern route of migration of the Arawakan natives and supports a western Amazonian origin of this linguistic family.

## References

- ANDREWS, R. M. et al. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. **Nature genetics**, v. 23, p. 147, 1999.
- ARIAS, L. et al. High-resolution mitochondrial DNA analysis sheds light on human diversity, cultural interactions, and population mobility in Northwestern Amazonia. **American Journal of Physical Anthropology**, v. 165, n. 2, p. 238–255, 2018.
- AUTON, A. et al. A global reference for human genetic variation. **Nature**, v. 526, n. 7571, p. 68–74, 2015.
- BERT, F. et al. Mitochondrial DNA diversity in the Llanos de Moxos: Moxo, Movima and Yuracare Amerindian populations from Bolivia lowlands. **Annals of Human Biology**, v. 31, n. 1, p. 9–28, 2009.
- DE SAINT PIERRE, M. et al. An alternative model for the early peopling of Southern South America revealed by analyses of three mitochondrial DNA haplogroups. **PLoS ONE**, v. 7, n. 9, 2012.
- DI CORCIA, T. et al. East of the Andes: The genetic profile of the Peruvian Amazon populations. **American Journal of Physical Anthropology**, v. 163, n. 2, p. 328–338, 2017.
- EXCOFFIER, L.; LAVAL, G.; SCHNEIDER, S. Arlequin (version 3.0); an integrated software package for population genetics data analysis. **Evol Bionform Online**, v. 1, p. 47–50, 2005.
- FREGEL, R.; DELGADO, S. HaploSearch: A tool for haplotype-sequence two-way transformation. **Mitochondrion**, v. 11, n. 2, p. 366–367, 2011.
- HUNEFELDT, C. **A Brief History of Peru**. 2 edition ed. Facts on File, 2010.
- HUNEFELDT, C.; HARRIS, B. **A Brief History of Peru**.
- MCEWAN, G. F. **The Incas: New Perspectives**.
- PARSON, W. et al. DNA Commission of the International Society for Forensic Genetics: Revised and extended guidelines for mitochondrial DNA typing. **Forensic science international. Genetics**, v. 13C, p. 134–142, 29 jul. 2014.
- PARSON, W.; BANDELT, H.-J. Extended guidelines for mtDNA typing of population data in forensic science. **Forensic science international. Genetics**, v. 1, n. 1, p. 13–9, mar. 2007.
- PARSON, W.; DÜR, A. EMPOP-a forensic mtDNA database. **Forensic science international. Genetics**, v. 1, n. 2, p. 88–92, jun. 2007.
- SANDOVAL, R. et al. The Genetic History of Indigenous Populations of the Peruvian and Bolivian Altiplano : The Legacy of the Uros. v. 8, n. 9, 2013.
- SIMÃO, F. et al. Defining mtDNA origins and population strati fi cation in Rio de Janeiro. v. 34, n. February, p. 97–104, 2018.

TABOADA-ECHALAR, P. et al. The Genetic Legacy of the Pre-Colonial Period in Contemporary Bolivians. v. 8, n. 3, 2013.

TINEO, D. H. et al. Genetic-characterization-of-27-Y-STR-loci-in-the-native-population-of-Ashaninka-from-Peru\_2015\_Forensic-Science-International-Genetics-Supplement-Seri.pdf. v. 5, p. 220–222, 2015.

VALVERDE, G. et al. Ancient DNA Analysis Suggests Negligible Impact of the Wari Empire Expansion in Peru ' s Central Coast during the Middle Horizon. p. 1–18, 2016.

WALKER, R. S.; RIBEIRO, L. A. Bayesian phylogeography of the Arawak expansion in lowland South America. n. January, p. 2562–2567, 2011.

ZIMMERMANN, B. et al. Improved visibility of character conflicts in quasi-median networks with the EMPOP NETWORK software. **Croatian Medical Journal**, v. 55, n. 2, p. 115–120, 2014.

## DISCUSSÃO FINAL E CONCLUSÕES

As populações da América do Sul são das mais heterogêneas do mundo. No entanto, muitas delas, carecem de estudos detalhados que descrevam a composição genética dos seus indivíduos. As populações do Rio de Janeiro, do Paraguai e do grupo nativo Ashaninka, analisadas na presente tese, têm histórias muito distintas, espelhadas na composição genética obtida pela análise do mtDNA.

Os primeiros ocupantes do território do Rio de Janeiro e do Este do Paraguai foram indígenas pertencentes à família linguística Tupi-Guarani. Já o grupo indígena Ashaninka pertence à família linguística Arawak. Esta diferença de origem linguística foi, de certa forma, observada pela distribuição dos haplogrupos nativos nos três grupos de amostras estudados. A linhagem D foi maioritária na população Ashaninka mas no Rio de Janeiro e no Paraguai o haplogrupo D está presente em menor proporção. Por outro lado, alguns haplogrupos encontrados no Rio de Janeiro e no Paraguai, não estão presentes no grupo Ashaninka (ou encontram-se em porcentagem muito baixa). É o caso do haplogrupo B4b que representa uma frequência de 22 % e 17 % das linhagens nativas do Rio de Janeiro e Paraguai, respectivamente, e está presente em apenas uma amostra das 171 de nativos do Peru.

A variação encontrada nas proporções de linhagens não-nativas nos três grupos pode estar associada, não só, à influência diferencial de colonizadores europeus no Brasil, Paraguai e Peru, mas também ao isolamento mantido pelos grupos nativos durante o período colonial, como é o caso dos Ashaninka do Peru.

Durante o período colonial, o Rio de Janeiro foi um dos principais portos marítimos da região sul do continente americano, por onde entrou um elevado número de colonizadores europeus e escravos africanos. Nos dias de hoje, é ainda evidente a elevada contribuição materna com origem nestes dois continentes, a par da preservação de um componente significativo de origem nativo-americana.

A população do Paraguai, apesar de ter sofrido influência europeia durante o período colonial, sempre foi um território consideravelmente isolado, com menor contribuição de populações europeias ou africanas. Este fato foi comprovado com os marcadores de mtDNA, onde a grande maioria dos haplótipos pertencem a linhagens maternas nativo-americanas e muito poucos pertencem a linhagens europeias ou

africanas. O estudo aprofundado por MPS de mitogenomas de indivíduos residentes no departamento do Alto Paraná revelou, de igual forma, a elevada contribuição nativa.

Por outro lado, a população Ashaninka do Peru manteve-se isolada na floresta amazônica, desde a chegada dos Espanhóis, com pouco ou nenhum contacto com outros grupos populacionais, não evidenciando, por isso, qualquer contribuição africana ou europeia.

A metodologia de Sanger, aplicada neste estudo para o sequenciamento da região controle do mtDNA, é amplamente utilizada em estudos populacionais e forenses. Trata-se de uma metodologia rápida e de baixo custo. Contudo, para uma análise mais aprofundada do genoma mitocondrial, a utilização deste método pode tornar-se moroso e com elevados custos, dado o grande número de reações por amostra que são necessárias efetuar para o sequenciamento completo da molécula de mtDNA. Desta forma, as tecnologias de sequenciamento massivo paralelo vieram agilizar o sequenciamento do mitogenoma. Dependendo do aparelho, com uma única corrida num instrumento de sequenciação de próxima geração é possível obter dezenas ou mesmo centenas de mitogenomas com elevada eficiência.

Neste trabalho, foi realizado um teste preliminar nas amostras dos indivíduos Ashaninka, com a amplificação de dois fragmentos sobrepostos para obtenção dos polimorfismos localizados na totalidade da mtDNA. No entanto, devido ao estado de conservação, as amostras de sangue encontravam-se degradadas, pelo que não foi possível realizar a amplificação nessas condições. A metodologia de MPS aplicada com recurso aos *kits Precision ID mtDNA Whole Genome Panel* e *Ampliseq Library* permitiu a amplificação de 162 segmentos mais curtos (média de 163 pbs), o que se verificou vantajoso, dada a baixa qualidade das amostras.

A análise de polimorfismos presentes na região codificante do mtDNA permite obter informação mais detalhada das linhagens maternas e, desta forma, desvendar aspectos sobre as migrações humanas e a constituição das populações. Para além disso, as metodologias de MPS permitem o estudo de amostras forenses em que a qualidade e/ou a quantidade de DNA são reduzidas.

No entanto, sendo que a metodologia de Sanger continua a ser amplamente utilizada em laboratórios forenses nos dias de hoje, a criação de bases de dados representativas de populações residentes nas diferentes regiões do mundo, com detalhamento da sua composição genética, é de extrema relevância.

Os resultados obtidos com a presente tese permitiram, por um lado, aprofundar o conhecimento existente sobre a ancestralidade materna de populações da América do Sul, destacando como os diferentes eventos históricos influenciaram a composição genética de 3 populações muito distintas. Por outro lado, obtiveram-se bases de dados robustas e detalhadas, que poderão ser utilizadas na rotina forense.

Em resumo:

- a) O estudo da população do Rio de Janeiro permitiu determinar o aporte, mais ou menos equitativo, de linhagens maternas de origem africana, nativo-americana e europeia, contribuindo para o aprofundamento e detalhamento da sua composição genética.
- b) O perfil genético africano do Rio de Janeiro demonstrou uma contribuição maioritária das regiões Sudoeste e Centro-Oeste de África, revelada pelas semelhanças encontradas com os haplogrupos presentes nestas regiões.
- c) Verificou-se a inexistência de um padrão genético entre as linhagens nativo-americanas, motivada por efeitos de deriva genética, o que impossibilitou a atribuição de uma origem étnica do componente nativo no Rio de Janeiro.
- d) Não foi possível realizar uma inferência sobre a origem das linhagens europeias no Rio de Janeiro, devido à baixa diferenciação genética entre populações de referência.
- e) Na análise da composição genética do Paraguai encontrou-se uma ancestralidade materna nativo-americana maioritária.
- f) A análise conjunta de marcadores autossômicos e do mtDNA revelaram processos de mistura recentes entre o Paraguai e outras populações já miscigenadas, mostrando a importância dos processos de miscigenação pós-coloniais na modulação do padrão genético atual do Paraguai.

- g) A implementação da metodologia MPS no estudo de amostras do Alto Paraná, Paraguai, permitiu confirmar o aumento de diversidade e obter uma maior precisão na classificação dos haplogrupos, comparativamente à análise da região controle.
- h) Os resultados do estudo das linhagens nativas do Alto Paraná indicaram um *input* não só de indivíduos pertencentes à família linguística Tupi-Guarani, como também a outros grupos linguísticos.
- i) O isolamento da população Ashaninka foi corroborado pela baixa diversidade haplotípica encontrada e pela inexistência de linhagens maternas não-nativas.
- j) Foi detectada uma possível rota migratória dos grupos Arawak para Sul do subcontinente Sul Americano, através da análise das linhagens B2 + 16051G + 152C e posterior comparação com outras populações sul-americanas.

## REFERÊNCIAS

- ALVES-SILVA, J. et al. The ancestry of Brazilian mtDNA lineages. **American journal of human genetics**, v. 67, n. 2, p. 444–61, ago. 2000.
- AMORIM, A.; BUDOWLE, B. **Handbook of Forensic Genetics: Biodiversity and Heredity in Civil and Criminal Investigation**. 1 ed. World Scientific Publishing Europe Ltd, 2016.
- ANA, A. P. et al. Contrasting admixture estimates in Rio de Janeiro obtained by different sampling strategies. **Forensic Science International: Genetics Supplement Series**, v. 6, n. July, p. e89–e91, 2017.
- ANDERSON, S. et al. Sequence and organization of the human mitochondrial genome. **Nature**, v. 290, p. 457–465, 1981.
- ANDREWS, R. M. et al. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. **Nature genetics**, v. 23, p. 147, 1999.
- ARAUJO, A. L. **African heritage and memories of slavery in Brazil and the South Atlantic world**. Cambria Press, 2015.
- ARIAS, L. et al. High-resolution mitochondrial DNA analysis sheds light on human diversity, cultural interactions, and population mobility in Northwestern Amazonia. **American Journal of Physical Anthropology**, v. 165, n. 2, p. 238–255, 2018.
- AUTON, A. et al. A global reference for human genetic variation. **Nature**, v. 526, n. 7571, p. 68–74, 2015.
- BANDELT, H.-J.; RICHARDS, M.; MACAULAY, V. **Human Mitochondrial DNA and the Evolution of Homo sapiens**. Springer, 2006.
- BÄR, W. et al. DNA Commission of the International Society for Forensic Genetics: guidelines for mitochondrial DNA typing. **Forensic science international**, v. 113, n. 4, p. 193–6, jan. 2000.
- BARBIERI, C. et al. Between Andes and Amazon: The genetic profile of the Arawak-speaking Yanomama. **American Journal of Physical Anthropology**, v. 155, n. 4, p. 600–609, 2014.
- BARBOSA, A. B. G. et al. Mitochondrial DNA control region polymorphism in the population of Alagoas State, North-Eastern Brazil. **Journal of Forensic Sciences**, v. 53, n. 1, p. 142–146, 2008.
- BEHAR, D. M. et al. A “Copernican” Reassessment of the Human Mitochondrial DNA Tree from its Root. **American Journal Of Human Genetics**, v. 90, n. 4, p. 675–684, 2012.
- BELLWOOD, P. **The Global Prehistory of Human Migration**. Wiley-Blackwell, 2014.

- BERNARDO, S. et al. MtDNA ancestry of Rio de Janeiro population, Brazil. **Molecular Biology Reports**, v. 41, n. 4, p. 1945–1950, 2014.
- BERT, F. et al. Mitochondrial DNA diversity in the Llanos de Moxos: Moxo, Movima and Yuracare Amerindian populations from Bolivia lowlands. **Annals of Human Biology**, v. 31, n. 1, p. 9–28, 2009.
- BETHELL, L. **The Cambridge World History of Slavery: Volume 3**.
- BODNER, M. et al. Rapid coastal spread of First Americans: Novel insights from South America's Southern Cone mitochondrial genomes. **Genome Research**, v. 22, n. 5, p. 811–820, 2012.
- BONATTO, S. L.; SALZANO, F. M. A single and early migration for the peopling of the Americas supported by mitochondrial DNA sequence data. **Proceedings of the National Academy of Sciences**, v. 94, n. 5, p. 1866–1871, 1997.
- BRANDINI, S. et al. The Paleo-Indian entry into South America according to mitogenomes. **Molecular Biology and Evolution**, v. 35, n. 2, p. 299–311, 2018.
- BROWN, W. M. Polymorphism in mitochondrial DNA of humans as revealed by restriction endonuclease analysis. v. 77, n. 6, p. 3605–3609, 1980.
- BUTLER, J. M. **Forensic DNA Typing**. Elsevier Academic Press, 2005.
- BUTLER, J. M. **Fundamentals of Forensic DNA Typing**. Academic Press, 2009.
- CATANESI, C. I. et al. Geographic Structure in Gran Chaco Amerindians Based on Five X-Chromosome STRs. **Human Biology**, v. 79, n. 4, p. 463–474, 2007.
- CEREZO, M. et al. Comprehensive Analysis of Pan-African Mitochondrial DNA Variation Provides New Insights into Continental Variation and Demography. **Journal of Genetics and Genomics**, v. 43, n. 3, p. 133–143, 2016.
- COHEN, R. **The Cambridge Survey of World Migration**. Cambridge University Press, 2010.
- CUMMINS, J. Mitochondrial DNA in mammalian reproduction. **Reviews of reproduction**, v. 3, n. 3, p. 172–82, set. 1998.
- DE F. FIGUEIREDO, R. et al. Male-specific contributions to the Brazilian population of Espírito Santo. **International Journal of Legal Medicine**, v. 130, n. 3, p. 679–681, 2016.
- DE SAINT PIERRE, M. et al. Arrival of Paleo-Indians to the Southern Cone of South America: New Clues from Mitogenomes. **PLoS ONE**, v. 7, n. 12, p. 1–9, 2012a.
- DE SAINT PIERRE, M. et al. An alternative model for the early peopling of Southern South America revealed by analyses of three mitochondrial DNA haplogroups. **PLoS ONE**, v. 7, n. 9, 2012b.
- DI CORCIA, T. et al. East of the Andes: The genetic profile of the Peruvian Amazon

populations. **American Journal of Physical Anthropology**, v. 163, n. 2, p. 328–338, 2017.

DORNELLES, C. L. et al. Mitochondrial DNA and Alu insertions in a genetically peculiar population: The Ayoreo Indians of Bolivia and Paraguay. **American Journal of Human Biology**, v. 16, n. 4, p. 479–488, 2004.

E, M.-E. et al. Paraguayan population data on the fifteen STR loci included in the PowerPlex 16 kit. **Journal of Forensic Sciences**, v. 48, n. 1, p. 253–255, 2003.

EXCOFFIER, L.; LAVAL, G.; SCHNEIDER, S. Arlequin (version 3.0); an integrated software package for population genetics data analysis. **Evol Bioinform Online**, v. 1, p. 47–50, 2005.

FREGEL, R.; DELGADO, S. HaploSearch: A tool for haplotype-sequence two-way transformation. **Mitochondrion**, v. 11, n. 2, p. 366–367, 2011.

FRIDMAN, C. et al. Haplotype diversity in mitochondrial DNA hypervariable region in a population of southeastern Brazil. **International Journal of Legal Medicine**, v. 128, n. 4, p. 589–593, 2014.

GLOBAL, P. I.; II, P. Migration Profiles: Paraguay. 2004.

GOEBEL, T.; WATERS, M. R.; O'ROURKE, D. H. The Late Pleistocene dispersal of modern humans in the Americas. **Science**, v. 319, n. 5869, p. 1497–1502, 2008.

HEBBLETHWAITE, M. **Paraguay (Bradt Travel Guide)**. 2 edition ed.

HILL, J. D.; SANTOS-GRANERO, F. **Comparative Arawakan Histories: Rethinking Language Family and Culture Area in Amazonia**. University of Illinois Press, 2006.

HOLLAND, M. M.; MCQUILLAN, M. R.; O'HANLON, K. A. Second generation sequencing allows for mtDNA mixture deconvolution and high resolution detection of heteroplasmy. **Croatian Medical Journal**, v. 52, n. 3, p. 299–313, jun. 2011.

HOMBURGER, J. R. et al. Genomic Insights into the Ancestry and Demographic History of South America. **PLoS Genetics**, v. 11, n. 12, p. 1–26, 2015.

HUNEFELDT, C. **A Brief History of Peru**. 2 ed. Facts on File, 2010.

HUNEFELDT, C.; HARRIS, B. **A Brief History of Peru**.

IBORRA, F. J.; KIMURA, H.; COOK, P. R. The functional organization of mitochondrial genomes in human cells. **BMC biology**, v. 14, p. 1–14, 2004.

JAFFARY, N. E. **Gender, Race and Religion in the Colonization of the Americas**. 1. ed. [s.l.] Ashgate Publishing Limited, 2016.

JANNUZZI, J. et al. Genetic characterization of Rio de Janeiro for different Y-STR sets. **International Journal of Legal Medicine**, 2018.

- JERMYN, L.; LIN, Y. J. **Paraguay (Cultures of the World)**. 2 ed. Benchmark Books, 2010.
- JOTA, M. S. et al. New native South American Y chromosome lineages. **Journal of Human Genetics**, v. 61, n. 7, p. 1–11, 2016.
- JUST, R. S.; IRWIN, J. A.; PARSON, W. Mitochondrial DNA heteroplasmy in the emerging field of massively parallel sequencing. **Forensic Science International: Genetics**, v. 18, p. 131–139, 2015.
- KING, J. L. et al. High-quality and high-throughput massively parallel sequencing of the human mitochondrial genome using the Illumina MiSeq. **Forensic science international. Genetics**, v. 12, p. 128–35, set. 2014.
- KIVISILD, T. Maternal ancestry and population history from whole mitochondrial genomes. **Investigative genetics**, p. 1–10, 2015.
- KLEIN, H. S.; LUNA, F. V. **Slavery in Brazil**. 1. ed. Cambridge University Press, 2009.
- LAMBERT, P. (ED.). **The Paraguay Reader: History, Culture, Politics**.
- LINS, T. C. et al. Genetic Heterogeneity of Self-Reported Ancestry Groups in an Admixed Brazilian Population. **Journal of Epidemiology**, v. 21, n. 4, p. 240–245, 2011.
- LOCKARD, C. A. **Societies, Networks, and Transitions, Volume II: Since 1450: A Global History**. 3. ed. Cengage Learning, 2014.
- LODISH, H. et al. **Molecular Cell Biology**. New York, NY: W. H. Freeman, 2000.
- MACAULAY, V. et al. Single , Rapid Coastal Settlement of Asia Revealed by Analysis of Complete Mitochondrial Genomes. **Science**, v. 308, n. 2005, 2005.
- MCEWAN, G. F. **The Incas: New Perspectives**.
- MEADE, T. **A Brief History of Brazil**. 2nd. ed. Checkmark Books, 2009.
- MELTON, T. Mitochondrial DNA Heteroplasmy. **Forensic science review**, v. 16, n. 1, 2004.
- MESA, N. R. et al. Autosomal, mt{DNA}, and {Y}-chromosome diversity in {A}merinds: {P}re- and post-{C}olumbian patterns of gene flow in {S}outh {A}merica. **American Journal of Human Genetics**, v. 67, n. 5, p. 1277–1286, 2000.
- MESSINA, F. et al. Signs of continental ancestry in urban populations of Peru through autosomal STR loci and mitochondrial DNA typing. **PLoS ONE**, v. 13, n. 7, p. 1–18, 2018.
- MISHMARA, D. et al. Natural selection shaped regional mtDNA variation in humans. v. 100, n. 1, 2003.

- NELSON, D. L.; COX, M. M. **Lehninger Principles of Biochemistry**. 4. ed. W H Freeman & Co, 2004.
- OLIVEIRA, A. M. et al. Male lineage strata of Brazilian population disclosed by the simultaneous analysis of STRs and SNPs. **Forensic Science International: Genetics**, v. 13, p. 264–268, 2014.
- PAERREGAARD, K. **Peruvians Dispersed: A Global Ethnography of Migration**. Lexington Books, 2008.
- PAKENDORF, B.; STONEKING, M. Mitochondrial DNA and human evolution. **Annual review of genomics and human genetics**, v. 6, p. 165–83, jan. 2005a.
- PAKENDORF, B.; STONEKING, M. Mitochondrial DNA and human evolution. **Annual review of genomics and human genetics**, v. 6, p. 165–83, jan. 2005b.
- PALENCIA, L. et al. Mitochondrial DNA diversity in a population from Santa Catarina (Brazil): Predominance of the European input. **International Journal of Legal Medicine**, v. 124, n. 4, p. 331–336, 2010.
- PARSON, W. et al. DNA Commission of the International Society for Forensic Genetics: Revised and extended guidelines for mitochondrial DNA typing. **Forensic Science International: Genetics**, v. 13, p. 134–142, 2014a.
- PARSON, W. et al. DNA Commission of the International Society for Forensic Genetics: Revised and extended guidelines for mitochondrial DNA typing. **Forensic science international. Genetics**, v. 13C, p. 134–142, 29 jul. 2014b.
- PARSON, W. et al. Massively parallel sequencing of complete mitochondrial genomes from hair shaft samples. **Forensic Science International: Genetics**, v. 15, p. 8–15, 2015.
- PARSON, W.; BANDELT, H.-J. Extended guidelines for mtDNA typing of population data in forensic science. **Forensic science international. Genetics**, v. 1, n. 1, p. 13–9, mar. 2007.
- PARSON, W.; DÜR, A. EMPOP-a forensic mtDNA database. **Forensic science international. Genetics**, v. 1, n. 2, p. 88–92, jun. 2007.
- PARSONS, T. J. Mitochondrial DNA Genome Sequencing and SNP Assay Development for Increased Power of Discrimination. 2006.
- PENA, S. D. J. et al. The genomic ancestry of individuals from different geographical regions of Brazil is more uniform than expected. **PloS one**, v. 6, n. 2, p. e17063, jan. 2011.
- PEREIRA, L.; CUNHA, C.; AMORIM, A. Predicting sampling saturation of mtDNA haplotypes: An application to an enlarged Portuguese database. **International Journal of Legal Medicine**, v. 118, n. 3, p. 132–136, 2004.
- PEREIRA, R. et al. Straightforward inference of ancestry and admixture proportions through ancestry-informative insertion deletion multiplexing. **PLoS ONE**, v. 7, n. 1,

2012.

PRIETO, L. et al. The GHEP-EMPOP collaboration on mtDNA population data - A new resource for forensic casework. **Forensic Science International: Genetics**, v. 5, n. 2, p. 146–151, 2011.

RESQUE, R. et al. Male lineages in Brazil: Intercontinental admixture and stratification of the European background. **PLoS ONE**, v. 11, n. 4, p. 1–17, 2016.

RIBEIRO, J. et al. Analysis of 23 Y-STRs in a population sample from eastern Paraguay. **Forensic Science International: Genetics**, v. 37, n. xxxx, p. e20–e22, 2018.

ROBINSON, J. T. et al. Integrative Genomics Viewer. **Nature Biotechnology**, n. 29, p. 24–26, 2011.

ROTHBERG, J. M. et al. An integrated semiconductor device enabling non-optical genome sequencing. **Nature**, v. 475, n. 7356, p. 348–352, 2011.

ROTHHAMMER, F.; DILLEHAY, T. D. The Late Pleistocene Colonization of South America : An Interdisciplinary Perspective. p. 540–549, 2009.

SALAS, A. et al. The African diaspora: mitochondrial DNA and the Atlantic slave trade. **American journal of human genetics**, v. 74, n. 3, p. 454–65, mar. 2004.

SALAS, A. et al. Charting the Ancestry of African Americans. **The American Journal of Human Genetics**, v. 77, n. 4, p. 676–680, 2005.

SALAS, A. et al. A cautionary note on switching mitochondrial DNA reference sequences in forensic genetics. **Forensic Science International: Genetics**, v. 6, n. 6, p. 182–184, 2012.

SALOUM DE NEVES MANTA, F. et al. Revisiting the genetic ancestry of Brazilians using autosomal AIM-Indels. **PloS one**, v. 8, n. 9, p. e75145, jan. 2013.

SANCHES, N. M. et al. Mitochondrial DNA control region diversity in a population from Espírito Santo state, Brazil. **Molecular Biology Reports**, v. 41, n. 10, p. 6645–6648, 2014.

SANDOVAL, J. R. et al. The Genetic History of Peruvian Quechua-Lamistas and Chankas: Uniparental DNA Patterns among Autochthonous Amazonian and Andean Populations. **Annals of Human Genetics**, v. 80, n. 2, p. 88–101, 2016.

SANDOVAL, R. et al. The Genetic History of Indigenous Populations of the Peruvian and Bolivian Altiplano : The Legacy of the Uros. v. 8, n. 9, 2013.

SANGER, F.; NICKLEN, S. DNA sequencing with chain-terminating inhibitors. v. 74, n. 12, p. 5463–5467, 1977.

SANTOS, H. C. et al. A minimum set of ancestry informative markers for determining admixture proportions in a mixed American population: The Brazilian set. **European Journal of Human Genetics**, v. 24, n. 5, p. 725–731, 2016.

- SANTOS, N. P. C. et al. Assessing individual interethnic admixture and population substructure using a 48-insertion-deletion (INSEL) ancestry-informative marker (AIM) panel. **Human Mutation**, v. 31, n. 2, p. 184–190, 2010.
- SCHMITT, R. et al. Extremely limited mitochondrial DNA variability among the Aché Natives of Paraguay. **Annals of human biology**, v. 31, n. 1, p. 87–94, 2004.
- SCHURR, T. G. The peopling of the new world: Perspectives from Molecular Anthropology. **Annual review of anthropology**, p. 0–13, 2004.
- SIMÃO, F. et al. Defining mtDNA origins and population strati fi cation in Rio de Janeiro. v. 34, n. February, p. 97–104, 2018.
- SOARES, P. et al. Correcting for Purifying Selection: An Improved Human Mitochondrial Molecular Clock. **American Journal of Human Genetics**, v. 84, n. 6, p. 740–759, 2009.
- SUTOVSKY, P. et al. Ubiquitinated sperm mitochondria, selective proteolysis, and the regulation of mitochondrial inheritance in mammalian embryos. **Biology of reproduction**, v. 63, n. 2, p. 582–90, ago. 2000.
- TABOADA-ECHALAR, P. et al. The Genetic Legacy of the Pre-Colonial Period in Contemporary Bolivians. v. 8, n. 3, 2013.
- The Cambridge History of Latin America**. Cambridge University Press, 2008.
- THORVALDSDÓTTIR, H.; ROBINSON, J. T.; MESIROV, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. **Briefings in Bioinformatics**, n. 14, p. 178–192, 2013.
- TINEO, D. H. et al. Genetic-characterization-of-27-Y-STR-loci-in-the-native-population-of-Ashaninka-from-Peru\_2015\_Forensic-Science-International-Genetics-Supplement-Seri.pdf. v. 5, p. 220–222, 2015.
- TULLY, G. et al. Considerations by the European DNA profiling ( EDNAP ) group on the working practices , nomenclature and interpretation of mitochondrial DNA profiles. **Forensic science international**, v. 124, p. 83–91, 2001.
- VALVERDE, G. et al. Ancient DNA Analysis Suggests Negligible Impact of the Wari Empire Expansion in Peru ' s Central Coast during the Middle Horizon. p. 1–18, 2016.
- VAN OVEN, M.; KAYSER, M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. **Human mutation**, v. 30, n. 2, p. E386-94, fev. 2009.
- VOET, D.; VOET, J. G.; PRATT, C. W. **Fundamental of Biochemistry**. 4. ed. Wiley, 2012.
- VULLO, C. et al. Genetic polymorphism of 22 autosomal STR markers in Paraguay. **Forensic Science International: Genetics**, v. 25, p. e16–e17, 2016.
- W, P.; A, D. EMPOP: a forensic mtDNA Database. **Forensic science international**.

**Genetics**, v. 1, n. 2, p. 88–92, 2007.

WALKER, R. S.; RIBEIRO, L. A. Bayesian phylogeography of the Arawak expansion in lowland South America. n. January, p. 2562–2567, 2011.

WALSH, P. S.; METZGER, D. A.; HIGUCHI, R. Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. **BioTechniques**, v. 10, p. 506–513, 1991.

WARREN, H. G. **Rebirth of the Paraguayan Republic: The First Colorado Era, 1878-1904**. 1 edition ed. [s.l.] University of Pittsburgh Press, 1985.

YANG, N. N. et al. Contrasting Patterns of Nuclear and mtDNA Diversity in Native American Populations. **Annals of Human Genetics**, v. 74, n. 6, p. 525–538, 2010.

ZIMMERMANN, B. et al. Improved visibility of character conflicts in quasi-median networks with the EMPOP NETWORK software. **Croatian Medical Journal**, v. 55, n. 2, p. 115–120, 2014.

**ANEXO A – Extração orgânica de sangue total - mancha em papel FTA****Procedimento**

1 - Antes de iniciar o procedimento, limpar a capela de exaustão e marcar o código de cada amostra em 3 tubos.

2 - Cortar em pedaços uma porção do papel FTA contendo a mancha de sangue, com aproximadamente 1 cm<sup>2</sup> (variando de acordo com a intensidade da mancha), para um tubo de 1,5 mL.

3 - Adicionar 200 µL de SSC 1X e esperar 15 min para voltar a hidratar a mancha de sangue. Macerar com uma ponta de pipeta ou pistilo.

4 - Adicionar mais 800 µL de SSC 1X, agitar em vortex e esperar 15 min.,

5 - Centrifugar por 5 min a 13 224 g (Microcentrífuga CT-15000 Cientec, 12000 rpm) e descartar o sobrenadante (usando uma pipeta), deixando aproximadamente 30 µL.

**Nota:** Se o papel e o sobrenadante estiverem muito corados (muita hemoglobina) deve-se fazer uma nova lavagem com SSC 1X , repetindo os passos 4 e 5

6 – Adicionar 400 µL de Tampão de Extração, 25 µL de Proteinase K a 10 mg/mL e 20 µL de solução de DTT a 1M. Incubar a 37°C toda a noite.

7 - Adicionar 200 µL de Fenol/Clorofórmio/Álcool Isoamílico (25:24:1) e agitar em vortex até formar uma emulsão leitosa (10 segundos);

8 - Centrifugar 3 minutos a 13 224 g. Transferir a fase aquosa (superior) para outro eppendorf, usando uma ponteira de 200 µL cortada na extremidade. Descartar também a interface, que contém as proteínas.

9 - Adicionar 200 µL de Clorofórmio/Álcool Isoamílico (24:1) e agitar em vortex até formar uma emulsão leitosa (10 segundos).

10 - Centrifugar por 3 minutos a 13 224 g. Transferir a fase aquosa (superior), com ponteira de 200 µL cortada na extremidade, para outro eppendorf.

**Nota:** Se a fase inferior estiver muito corada ou se a interface ainda estiver difusa, deve-se

repetir os passos 9 e 10.

11 - Adicionar 35 µL de solução de Acetato de sódio a 3M e 1000 µL de Etanol a 100% (gelado). Agitar bem e colocar a -20°C a noite toda ou a -80°C durante 20-30 min.

13 - Centrifugar 15 minutos a 13 000 rpm, numa centrifuga refrigerada a 4 °C. Decantar muito cuidadosamente.

14 - Adicionar 1000 µL de Etanol 70% (gelado) e agitar.

15 - Centrifugar por 5 minutos a 13 224 g. Decantar muito cuidadosamente (escorrer em papel) e deixar secar.

16 - Ressuspender em 30 -50 µL de TE.

## APÊNDICE A - Artigo “*Contrasting admixture estimates in Rio de Janeiro obtained by different sampling strategies*”

Forensic Science International: Genetics Supplement Series 6 (2017) e89–e91



Contents lists available at ScienceDirect

Forensic Science International: Genetics Supplement Series

journal homepage: [www.elsevier.com/locate/fsigss](http://www.elsevier.com/locate/fsigss)



### Contrasting admixture estimates in Rio de Janeiro obtained by different sampling strategies

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

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

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

#### ARTICLE INFO

**Keywords:**  
Ancestry informative markers  
Population genetics  
Brazil  
Admixed populations

#### ABSTRACT

In this study, samples of non-related individuals from Rio the Janeiro State were genotyped for 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 and selected from paternity casework. 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. 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.

#### 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 [1]. 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 [3,4].

The ancestry of the Rio de Janeiro population was previous investigated by Manta et al. [5] using 46 AIM indels described by Pereira et al. [6], in a sample of 280 unrelated individuals born in the metropolitan region. In the present study 762 samples from the State of 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 the 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 the 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 the 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. [6]. Separation of the PCR amplified products and

\* Corresponding author at: DNA Diagnostic Laboratory (LDD), State University of Rio de Janeiro (UERJ), Brazil.  
E-mail address: [leonorgusmao@gmail.com](mailto:leonorgusmao@gmail.com) (L. Gusmão).

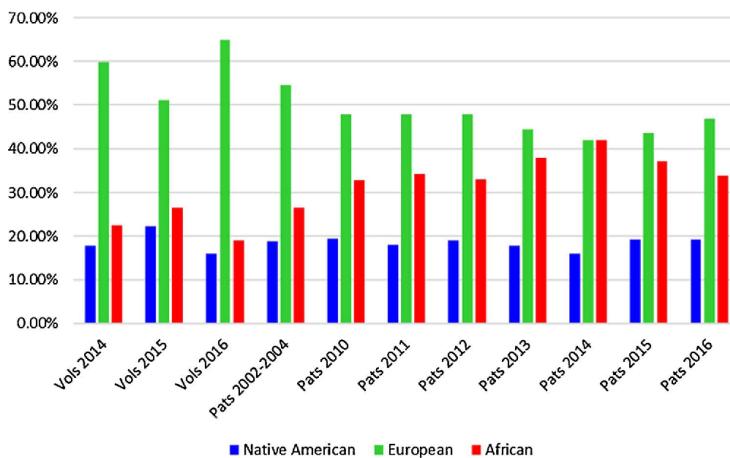


Fig. 1. Comparison of the average proportion of Native American, European and African ancestry, among non-related individuals within eleven sample sets (based on sampling criteria).

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 [7,8]. To estimate the ancestral membership proportions, a supervised analysis was performed using prior information on the geographic origin for the African, European and Native American reference samples (data from Pereira et al. [6]). 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 de Janeiro from 280 to 1042 individuals, there is a prevalence of the European ancestry, followed by an African contribution. This is in accordance with what was previously reported by Manta et al. [5], although, with the increment on sample size, the European contribution slightly increased (from 55% to 63%) and the African ancestry became lower (decreasing from 30% to 25%). The Native American background was the less represented, accounting for 12%, a value that is similar to that observed by Manta et al. [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 2014–2016, 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 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 Manta et al. [5], 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. This study points to the need of a more detailed study on populations from different regions, to better disclose the stratification of Rio de Janeiro State.

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. Salomé, R. Pereira, R. Viana, 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. Calafá, 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.

**Table S1.** Haplogroup and haplotypes found in a sample from 205 unrelated individuals living in Rio de Janeiro metropolitan region for mtDNA control region (16024-576)







**Table S2.** Point heteroplasmy in 8 samples from Rio de Janeiro population.

Sample	Point heteroplasmy
PRJ093	151Y
PRJ095	16147Y
PRJ099	207R
PRJ225	152Y
PRJ248	16311Y
PRJ285	199Y
PRJ384	237R
PRJ394	152Y

Y – mixture of C and T; R – mixture of A and G

**Table S3.** Pairwise *FST* values between admixed Brazilian populations for HVS-I region (below diagonal) and corresponding differentiation p-values (above diagonal)

	n	Rio de Janeiro	Santa Catarina	Sudoeste	Espirito Santo	São Paulo	Alagoas	São Paulo_Afro	Norte	Pernambuco	Minas Gerais	South
Rio de Janeiro	205	*	<b>0.0004</b>	0.1898	0.4701	0.4286	0.0480	<5E-06	0,0039	0.3372	0.1068	<b>0,0005</b>
Santa Catarina	80	0,0171	*	<b>0.0004</b>	<b>0.0004</b>	<5E-06	<5E-06	<5E-06	0,0114	0,0158	<b>0,0005</b>	0,1876
Sudeste	290	0,0011	0,0144	*	0,3791	0,1612	0,0303	<5E-06	0,0220	0,2414	0,0624	<b>0,0004</b>
Espirito Santo	95	-0,0003	0,0189	0,0003	*	0,8439	0,2210	<5E-06	0,0141	0,7917	0,3444	<b>0,0005</b>
São Paulo	142	0,0000	0,0191	0,0015	-0,0027	*	0,1420	<5E-06	0,0015	0,4304	0,0301	<b>0,0002</b>
Alagoas	167	0,0038	0,0307	0,0038	0,0017	0,0023	*	<5E-06	0,0043	0,0760	0,0172	<5E-06
São Paulo_Afro	120	0,0402	0,1024	0,0536	0,0326	0,0397	0,0372	*	<5E-06	<5E-06	<5E-06	<5E-06
Norte	48	0,0175	0,0160	0,0103	0,0149	0,0202	0,0195	0,0903	*	0,1035	0,0804	0,0378
Pernambuco	50	0,0009	0,0142	0,0022	-0,0041	0,0000	0,0073	0,0451	0,0093	*	0,6383	0,0157
Minas Gerais	99	0,0032	0,0174	0,0038	0,0007	0,0061	0,0078	0,0538	0,0080	-0,0024	*	0,0053
South	50	0,0274	0,0039	0,0221	0,0272	0,0319	0,0433	0,1119	0,0145	0,0192	0,0180	*

Significant values after Bonferroni's correction are indicated in red (p&lt;0.0009)

s.e.≤0,0053

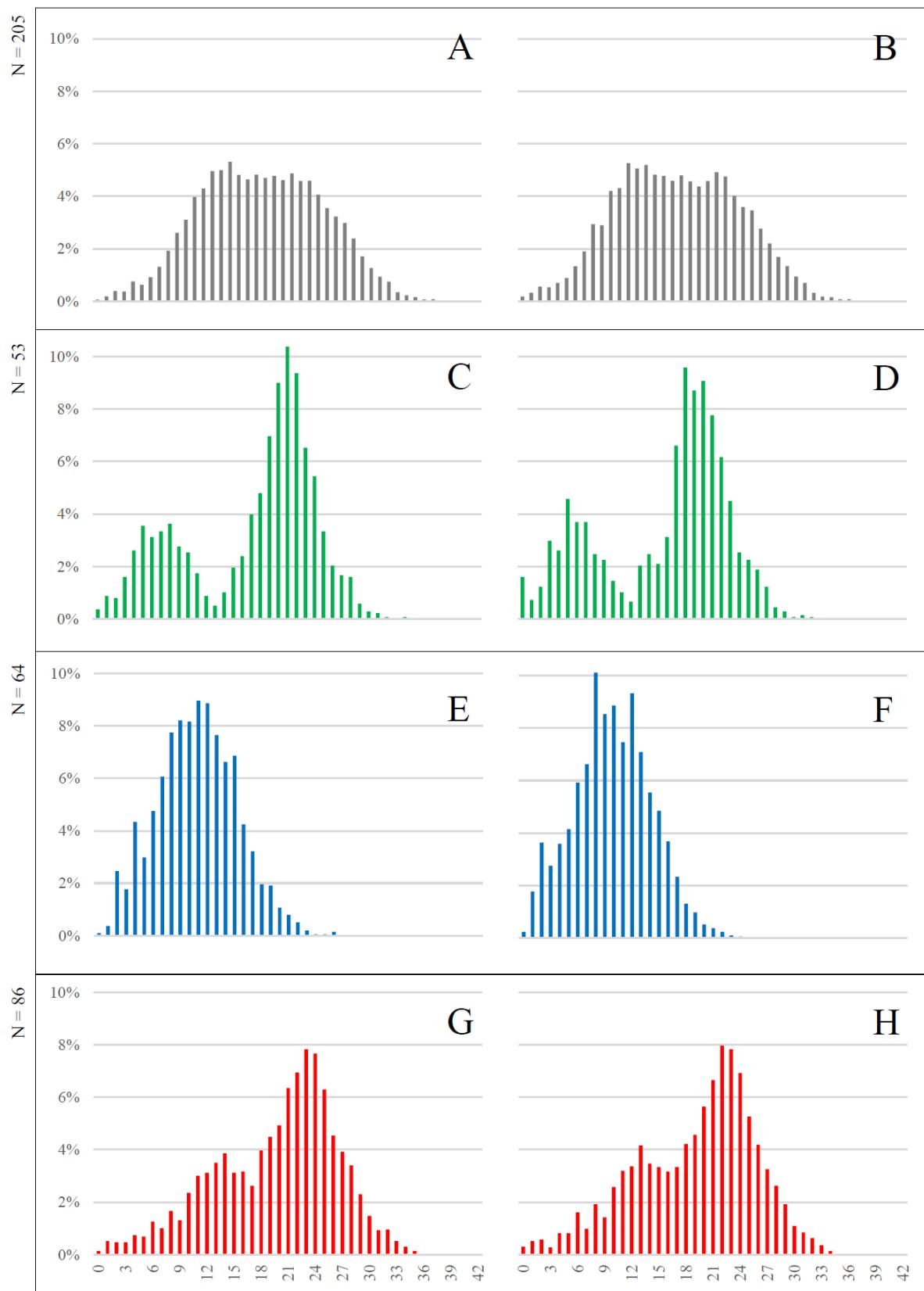
10100 permutations

**Table S4.** Matrix of pairwise *FST* genetic distances among European linages in Rio de Janeiro and 10 reference populations from Europe (below diagonal) and the corresponding differentiation p-values (above diagonal) obtained for 10100 permutations, for mtDNA control reg

	Rio de Janeiro	Portugal	Basque	Germany	Netherlands	Italy	Austria	Finland	Denmark	Greece	Spain
Rio de Janeiro	*	0.58984	0.04633	0.27779	0.01099	0.17127	0.15543	0.00218	0.44877	< 5E-06	0.20849
Portugal	-0.00093	*	0.00030	0.23423	< 5E-06	0.01782	0.05148	< 5E-06	0.01188	< 5E-06	0.00040
Basque	0.00702	0.01093	*	0.00119	< 5E-06	< 5E-06	< 5E-06	0.00010	0.00020	< 5E-06	0.03564
Germany	0.00171	0.00117	0.01358	*	0.63707	0.21740	0.39917	0.00010	0.55123	< 5E-06	0.00158
Netherlands	0.00757	0.00611	0.01929	-0.00079	*	0.00040	0.00099	< 5E-06	0.04307	< 5E-06	< 5E-06
Italy	0.00207	0.00212	0.01474	0.00116	0.00325	*	0.27364	< 5E-06	0.03109	< 5E-06	0.00010
Austria	0.00260	0.00180	0.01393	0.00016	0.00343	0.00040	*	< 5E-06	0.11939	< 5E-06	0.00010
Finland	0.01226	0.00838	0.01538	0.01241	0.01860	0.01488	0.01217	*	< 5E-06	< 5E-06	< 5E-06
Denmark	-0.00005	0.00342	0.01238	-0.00060	0.00189	0.00223	0.00143	0.01670	*	< 5E-06	< 5E-06
Greece	0.04739	0.03580	0.04751	0.03117	0.03260	0.04375	0.04117	0.04498	0.03620	*	< 5E-06
Spain	0.00171	0.00450	0.00362	0.00755	0.01022	0.00467	0.00504	0.01130	0.00682	0.04416	*

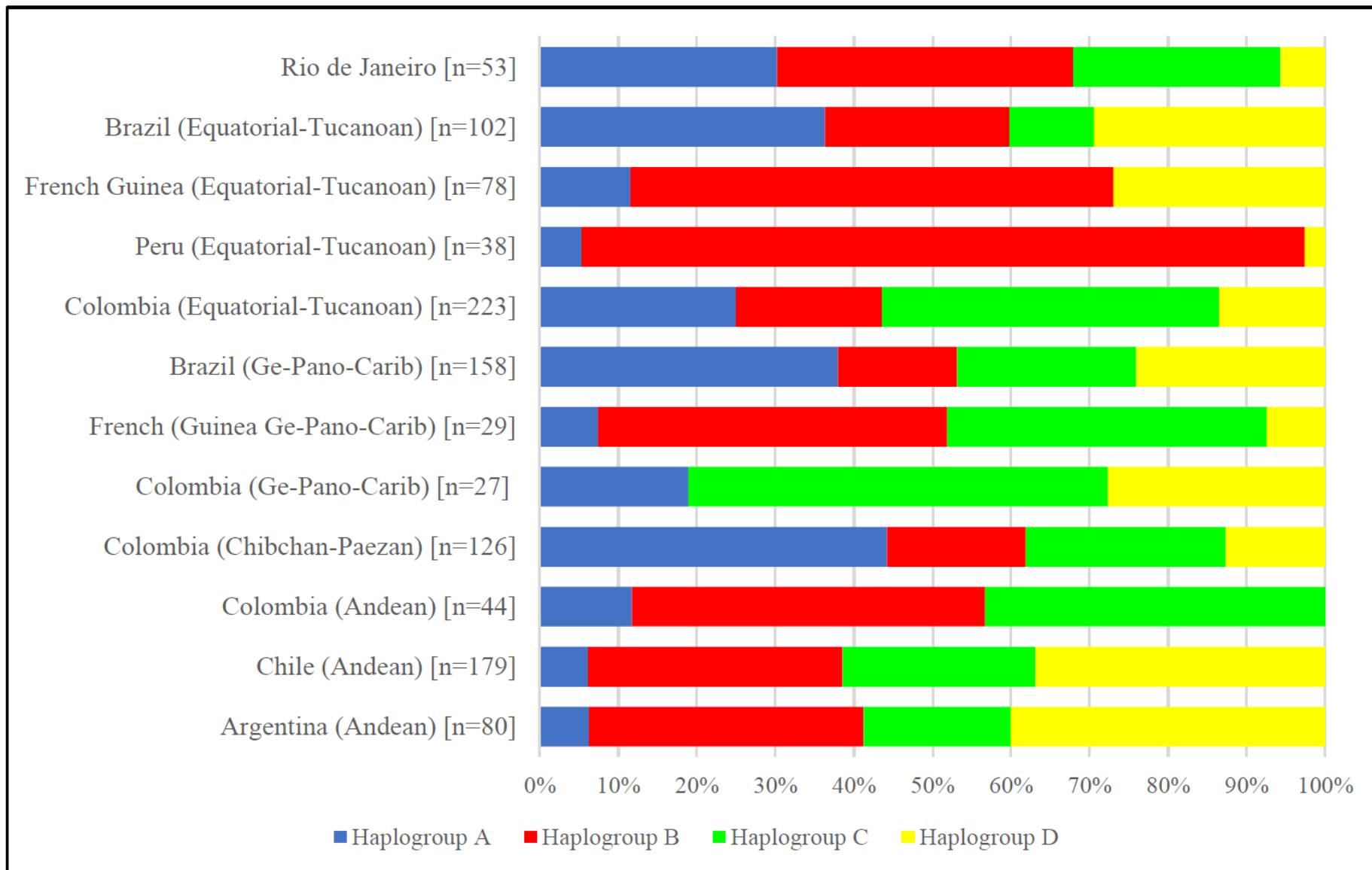
Note: significance level is 0.0009 (after applying Bonferroni's correction); s.e.≤0,0053

Fig S1 – Mismatch distribution for mtDNA control region of Rio de Janeiro samples



S1A: Total sample; S1B: Total sample, disregarding insertions (at positions 309, 315, 573 and 16193); S1C: Native American lineages; S1B: Native American lineages, disregarding insertions; S1C: Europeans lineages; S1B: Europeans lineages, disregarding insertions; S1C: African lineages; S1B: African lineages, disregarding insertions.

**Fig. S2.** Haplogroup distribution of the native American lineages in Rio de Janeiro and in reference Native American populations.



Sample ID	Haplotype																																				
2300000	16126C	16223T	16278T	16290T	16319A	16362C	57G	59C	64T	73G	146C	152C	153G	235G	263G	297G	315.1C	523del	524del	573.1C	573.2C	573.3C	573.4C	573.5C													
2300001	16092C	16182C	16183C	16189C	16217C	16241G	16304C	16519C	73G	103A	146C	152C	263G	309.1C	315.1C	523del	524del																				
2300002	16111T	16189C	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del																					
2300003	16142T	16179T	16223T	16295T	16325C	16362C	16497G	16519C	73G	263G	309.1C	315.1C	489C	524.1A	524.2C																						
2300004	16183C	16189C	16193.1C	16217C	16519C	73G	263G	309.1C	315.1C	499A	524.1A	524.2C																									
2300005	16223T	16241G	16301T	16342C	16362C	16519C	73G	106del	107del	108del	110del	111del	152C	263G	309.1C	315.1C	489C	523del	524del																		
2300007	16183C	16189C	16193.1C	16217C	16356C	16519C	73G	93G	195C	263G	309.1C	315.1C	499A																								
2300008	16183C	16189C	16217C	16519C	73G	263G	315.1C	533G																													
2300009	16111T	16189C	16223T	16290T	16319A	16362C	64T	73G	143A	146C	153G	204C	235G	263G	315.1C	523del	524del																				
2300010	16183C	16189C	16193.1C	16217C	16519C	73G	263G	309.1C	309.2C	315.1C	499A	524.1A	524.2C																								
2300011	16145A	16172C	16223T	16325C	16362C	73G	263G	315.1C	489C																												
2300012	16192T	16270T	73G	150T	263G	315.1C	533G																														
2300013	16111T	16145A	16172C	16189C	16217T	16319C	16223T	16265C	16278T	16286G	16294T	16311C	16360T	16519C	16527T	73G	151T	152C	182T	186A	189C	195C	198T	247A	263G	297G	309.1C	315.1C	316A	385G	471C	523del	524del				
2300014	16111T	16223T	16290T	16319A	16362C	16390A	16519C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del																				
2300015	16223T	16241G	16296T	16301T	16342C	16362C	73G	152C	263G	315.1C	489C	523del	524del																								
2300016	16111T	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	271T	315.1C	523del	524del																						
2300017	16183C	16189C	16193.1C	16217C	16519C	73G	263G	309.1C	309.2C	315.1C	499A	524.1A	524.2C																								
2300018	16168T	16182C	16183C	16189C	16217C	16249C	16312G	16344T	16519C	73G	152C	263G	271T	309.1C	309.2C	315.1C	499A																				
2300019	16183C	16189C	16193.1C	16217C	16519C	73G	263G	309.1C	309.2C	315.1C	499A	524.1A	524.2C	524.3A	524.4C																						
2300020	16183C	16189C	16193.1C	16217C	16519C	73G	263G	309.1C	309.2C	315.1C	499A	524.1A	524.2C																								
2300021	16183C	16189C	16217C	16519C	73G	103A	146C	152C	263G	279C	309.1C	315.1C	455.1T	499A																							
2300022	16223T	16298C	16299G	16325C	16327T	16519C	73G	249del	290del	291del	315.1C	489C	493G	523del	524del																						
2300023	16223T	16298C	16325C	16327T	16519C	73G	249del	290del	291del	315.1C	489C	493G	523del	524del																							
2300024	16233G	315.1C	523del	524del																																	
2300025	16223T	16325C	16326C	16519C	73G	263G	315.1C	533G																													
2300026	16223T	16325C	16362C	16519C	73G	146C	263G	315.1C	489C																												
2300027	16223T	16325C	16362C	16519C	73G	146C	263G	315.1C	489C																												
2300028	16223T	16241G	16296T	16301T	16342C	16362C	73G	152C	263G	315.1C	489C	523del	524del																								
2300029	16178C	16183C	16189C	16193.1C	16217C	16519C	73G	263G	315.1C	499A																											
2300030	16122T	16292T	16298C	16325C	16327T	16362C	73G	249del	263G	290del	291del	309.1C	315.1C	489C	493G	523del	524del																				
2300031	16145A	16182C	16183C	16189C	16217C	16519C	73G	263G	315.1C	499A	524.1A	524.2C	524.3A	524.4C	524.5A	524.6C																					
2300032	16129A	16173T	16183C	16189C	16193.1C	16519C	55.1T	57C	59C	71del	73G	263G	309.1C	309.2C	315.1C																						
2300033	16122T	16298C	16325C	16327T	16362C	73G	249del	263G	290del	291del	309.1C	315.1C	489C	493G	523del	524del																					
2300034	16178C	16182M	16183C	16189C	16193.1C	16217C	16519C	73G	263G	315.1C	499A																										
2300035	16183C	16189C	16193.1C	16223T	16278T	16519C	73G	153G	195C	225A	226C	263G	315.1C																								
2300036	16183C	16189C	16193.1C	16223T	16278T	16519C	73G	249del	263G	290del	291del	309.1C	315.1C	489C	493G	523del	524del																				
2300037	16145A	16182C	16183C	16189C	16217C	16294A	16519C	73G	263G	309.1C	309.2C	315.1C	484G	499A																							
2300038	16111T	16223T	16239A	16266T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del																				
2300039	16069T	16111T	16126C	16185A	16230T	16319A	16362C	57G	73G	146C	152T	152C	263G	297G	309.1C	315.1C	523del	524del																			
2300040	16182C	16183C	16189C	16217C	16316G	16519C	57G	73G	151T	260A	263G	309.1C	309.2C	315.1C	499A	523del	524del																				
2300041	16176T	16223T	16298C	16325C	16327T	73G	195C	203A	204C	249del	263G	290del	291del	309.1C	315.1C	489C																					
2300042	16183C	16189C	16217C	16468C	16519C	73G	263G	309.1C	309.2C	315.1C	499A	524.1A	524.2C																								
2300043	16111T	16145A	16223T	16288C	16290T	16319A	16362C	73G	146C	153G	235G	263G	309.1C	309.2C	315.1C	523del	524del																				
2300044	16127T	16234T	16278T	16366T	16519C	73G	185A	188G	228A	236G	295T	315.1C	462T	489C	523del	524del																					
2300045	16187A	16223T	16325C	16362C	73G	263G</td																															









**APÊNDICE D – Material suplementar do Artigo 3 “*Defining mtDNA origins and population stratification in Rio de Janeiro*”**

**Table S1.** Origin of mitogenomes used for comparison purposes in the present study.

Reference	n	Country	Etnich group
Arias et al. 2017	24	Colombia	Ach-Piapoco
Arias et al. 2017	9	Colombia	Admixed
Arias et al. 2017	8	Colombia	Carijona
Arias et al. 2017	17	Colombia	Cocama
Arias et al. 2017	19	Colombia	Coreguaje
Arias et al. 2017	17	Colombia	Curripaco
Arias et al. 2017	17	Colombia	Desano
Arias et al. 2017	35	Colombia	Guayabero
Arias et al. 2017	17	Colombia	Inga
Arias et al. 2017	11	Colombia	Kamentsa
Arias et al. 2017	26	Colombia	Mur-Uitoto
Arias et al. 2017	16	Colombia	Nukak
Arias et al. 2017	18	Colombia	Other ethnich group
Arias et al. 2017	10	Colombia	Other-ET
Arias et al. 2017	14	Colombia	Pasto
Arias et al. 2017	13	Colombia	Pira-Wanano
Arias et al. 2017	19	Colombia	Puinave
Arias et al. 2017	16	Colombia	Saliba
Arias et al. 2017	16	Colombia	Sikuani
Arias et al. 2017	17	Colombia	Siona
Arias et al. 2017	10	Colombia	Siriano
Arias et al. 2017	10	Colombia	Tanimuka
Arias et al. 2017	18	Colombia	Tikuna
Arias et al. 2017	10	Colombia	Tuka-Tatuyo
Arias et al. 2017	13	Colombia	Yagua
Arias et al. 2017	39	Colombia	Yucu-Matapi
Brandini et al 2017	208	Ecuador	Admixed
1000genomes	94	Colombia	Admixed
1000genomes	86	Peru	Admixed

**Table S2.** Haplogroups and haplotypes found in a sample from 105 unrelated individuals living in Alto Paraná, Paraguay, for mitogenomes.

Sample ID #1 ALL	GenBank ID	Haplotype	n	Haplotype																													
23000337	MH981823	C1b	1	45G	73G	113T	249del	263G	290del	291del	315..1C	489C	493G	523del	524del	750G	1438G	2706G	3552A	4715G	4736C	4769G	5054C	5095C	6249A	7028T	7196A	8134C					
23000338	MH981824	A2+(64)	1	64T	73G	146C	153G	210G	235G	263G	315..1C	523del	524del	663G	739T	750G	1438G	1736G	2706G	4248C	4769G	4824G	6340T	7028T	8027A	8794T	8860G	9210G					
23000339	MH981825	A2+(64)	1	64T	73G	93G	146C	153G	185A	235G	263G	315..1C	338T	523del	524del	573..1C	663G	750G	1438G	1719A	1736G	2706G	3438A	4248C	4769G	4824G	7028T	8027A	16223T	16325C	16327T		
23000340	MH981826	B2e	1	73G	263G	315..1C	499A	524..1A	524..2C	750G	827G	1438G	2706G	3547G	4769G	4820A	4977C	6119T	6473T	7028T	8281del	8282del	8283del	8284del	8285del	8286del	8287del	8288del					
23000341	MH981827	B2b3a	1	73G	152C	263G	271T	315..1C	499A	750G	827G	1438G	2706G	3547G	3918A	4232C	4769G	4820A	4977C	5899..1C	6473T	6755A	7028T	8281del	8282del	8283del	8284del	8285del	8286del	8287del	8288del		
23000342	MH981828	D1f1	1	73G	131C	263G	315..1C	489C	750G	1438G	2092T	2706G	3010A	4769G	4883T	5178A	7028T	8414T	8701G	8860G	9078C	9540C	10398G	10400T	10873T	11719A	11770C						
23000343	MH981829	C1d1d	1	73G	194T	249del	263G	290del	291del	315..1C	489C	523del	524del	750G	1438G	2706G	3552A	4715G	4769G	7028T	7196A	7697A	8584A	8701G	8860G	9540C	9545G	10398G					
23000344	MH981830	C1b	1	73G	249del	263G	290del	291del	315..1C	489C	493G	523del	524del	750G	1438G	2393T	2706G	3197C	3552A	4715G	4769G	7028T	7196A	8584A	8701G	8860G	9540C	9545G					
23000345	MH981831	A2q	1	64T	73G	146C	153G	235G	238G	263G	315..1C	523del	524del	750G	1438G	1736G	2706G	3816G	4188G	4248C	4769G	4824G	7028T	7581C	8027A	8794T	8860G	11719A					
23000346	MH981832	B2b3a	1	73G	152C	263G	271T	315..1C	499A	750G	827G	1438G	1719A	2706G	3547G	3918A	4232C	4769G	4820A	4977C	5899..1C	6473T	6755A	7028T	8281del	8282del	8283del	8284del	8285del	8286del	8287del	8288del	
23000347	MH981833	B2h	1	73G	204Y	263G	315..1C	499A	524..1A	524..2C	750G	827G	1438G	2706G	3547G	4769G	4820A	4977C	6473T	7028T	8281del	8282del	8283del	8284del	8285del	8286del	8287del	8288del					
23000348	MH981834	B2e	1	73G	263G	315..1C	499A	524..1A	524..2C	524..3A	750G	1438G	2706G	3547G	4769G	4820A	4947C	4977C	6119T	6473T	7028T	8281del	8282del	8283del	8284del	8285del	8286del	8287del	8288del				
23000349	MH981835	D1	1	73G	146C	263G	315..1C	489C	750G	1438G	1943G	2092T	2706G	3010A	3591A	4769G	4883T	5178A	6305T	7028T	8414T	8701G	8860G	9540C	9554A	10398G	10400T	10873C					
23000350	MH981836	L1c1a2	1	73G	89C	93G	95C	152C	182T	186A	189C	236C	247A	263G	297G	315..1C	316A	523del	524del	750G	769A	825A	1018A	1438G	2395del	2706G	2755G	2758A					
23000351	MH981837	C1	1	73G	143A	249del	263G	290del	291del	315..1C	489C	523del	524del	750G	986C	1438G	1462A	1842G	2706G	3531A	3552A	4691T	4715G	4769G	5003T	7028T	7196A	7673G					
23000352	MH981838	B2h	1	73G	241..1T	241..2A	241..3A	263G	315..1C	499A	524..1A	524..2C	750G	827G	1438G	2706G	3547G	4769G	4820A	4977C	6473T	7028T	8281del	8282del	8283del	8284del	8285del	8286del	8287del	8288del			
23000353	MH981839	C1b2	1	72C	73G	249del	290del	291del	315..1C	489C	493G	523del	524del	750G	1438G	2706G	3552A	3915A	4191G	4242T	4715G	4769G	7013A	7028T	7196A	8584A	8633G	8701G					
23000354	MH981840	A2	1	73G	146C	153G	235G	263G	315..1C	523del	524del	663G	750G	1438G	1736G	2706G	3733R	4248C	4769G	4824G	7028T	8794T	8860G	9096C	10463C	11719A	12007A						
23000355	MH981841	L3e2a1b1	1	73G	150T	195C	198T	263G	315..1C	499A	750G	1438G	2352C	2706G	4769G	4823C	6413C	7028T	7170T	7867T	8701G	8860G	9540C	10398G	10819G	10873C	11719A	12705T					
23000356	MH981842	J1c2e2	1	73G	185A	188G	228A	263G	295T	315..1C	462T	489C	523del	524del	709A	750G	1438G	2706G	3010A	4216C	4232C	4769G	7028T	8404C	8860G	10398G	11251G	11719A					
23000357	MH981843	C1b2	1	72C	73G	249del	290del	291del	315..1C	489C	493G	523del	524del	750G	1438G	2706G	3552A	3915A	4191G	4242T	4715G	4769G	7013A	7028T	7196A	8584A	8633G	8701G					
23000358	MH981844	B2	1	57G	73G	146C	151T	260A	263G	315..1C	499A	523del	524del	750G	827G	1438G	2706G	3372C	4769G	4820A	4977C	6473T	7028T	7830A	8281del	8282del	8283del						

23000359	MH981845	A2+(64)	1	64T 12007A 73G 146C 153G 210G 235G 263G 315.1C 523del 524del 663G 739T 750G 1438G 1736G 2706G 4248C 4769G 4824G 7028T 8027A 8794T 8860G 9210G 11719A	12705T 14766T 15300C 15326G 16111T 16223T 16290T 16319A 16362C 16519C 16223T 16325C 16362C
23000360	MH981846	D1e	1	73G 11248G 146C 263G 273T 315.1C 489C 750G 1438G 2092T 2706G 3010A 3316A 4769G 4883T 5178A 5773A 7028T 8414T 8701G 8860G 8934T 9540C 10398G 10400T 10873C	11719A 12705T 14668T 14766T 14783C 15043A 15301A 15326G 16223T 16325C 16362C
23000361	MH981847	C1d1	1	73G 9545G 194T 195C 249del 263G 290del 291del 315.1C 489C 523del 524del 750G 1438G 2706G 3547G 4769G 4820A 4977C 5186G 5375T 5895T 6473T 7028T 8281del 8282del 8283del 8284del 8285del 8286del 8287del	10398G 10400T 10873C 11090T 11719A 11914A 12501A 12705T 12879C 13263G 14318C 14766T 14783C 15043A 15326G 15535T 15884A 16189C 16217C 16519C
23000362	MH981848	B2	1	73G 8288del 263G 315.1C 499A 573.1C 750G 827G 1438G 2706G 3547G 4769G 4820A 4977C 5186G 5375T 5895T 6473T 7028T 8281del 8282del 8283del 8284del 8285del 8286del 8287del	8289del 8860G 9500C 11177T 11719A 13125T 13590A 14766T 15326G 15535T 15884A 16189C 16217C 16519C
23000363	MH981849	L1b1a12a	1	73G 3308C 152G 182T 185T 195C 198T 247A 263G 315.1C 357G 523del 524del 709A 710C 750G 769A 825A 1018A 1462A 1738C 2352C 2706G 2758A 2768G 2885C	3594T 3666A 3693A 4104G 4769G 5036G 5046A 5393C 5655C 6548T 6827C 6989G 7028T 7055G 7146G 7256T 7389C 7521A 7867T 8248G 8468T 8655T 8701G 8860G 16223T
23000364	MH981850	C1d1	1	73G 9347G 150T 194T 249del 263G 290del 291del 315.1C 489C 523del 524del 750G 1438G 2706G 3537G 3552A 4715G 4769G 6915A 7028T 7196A 7697A 8584A 8701G 8860G	9540C 10398G 10400T 10754G 10873C 11719A 11914A 12705T 13015C 13263G 14318C 14766T 14783C 15043A 15326G 15487T 15886T 16051G 16209C 16223T 16298C 16325C
23000365	MH981851	L2b1b	1	73G 3594T 4104G 4158G 4370C 4767G 4769G 4944G 5027T 5331A 5814C 6026A 6629G 6713T 6852A 6896C 7028T 7055G 7146G 7256T 7521A 7624A 8080T 8206A 8387A 8701G 8860G 9221G 16311C 16362C 16390A	10115C 10398G 10828C 10873C 11719A 11944C 12236A 12705T 12948G 13590A 13650T 13924T 14059G 14766T 15110A 15217A 15301A 15326G 16114A 16129A 16153A 16213A 16223T 16278T
23000366	MH981852	A2	1	73G 12007A 146C 150T 152C 182T 183G 195C 198T 204C 263G 315.1C 523del 524del 663G 750G 1438G 1736G 2706G 4248C 4769G 4824G 7028T 8027A 8794T 8860G 9064A 9096C 10463C 11719A	12705T 14127C 14766T 15326G 15951G 16111T 16126C 16223T 16259T 16290T 16319A 16362C 16519C
23000367	MH981853	A2+(64)	1	64T 11719A 12007A 73G 146C 153G 235G 263G 315.1C 523del 524del 663G 750G 1438G 1736G 2706G 4248C 4769G 4824G 7028T 8027A 8794T 8860G 9064G 9096C 10463C 11272G	120705T 14766T 15326G 16111T 16223T 16274R 16290T 16319A 16362C
23000368	MH981854	L1c1b	1	73G 2758A 2885C 3594T 3954T 3666A 3796T 3834A 3843G 4104G 4589C 4688C 4696G 508G 553C 5951G 6071C 6252G 6713T 6852A 6896C 7028T 7055G 7146G 7256T 7521A 7624A 8080T 8206A 8387A 8701G 8860G 9221G 13789C 13874C 14000A 14148G 14178C 14356T 14393G 14554G 14560A 14766T 14911T 15326G 16086C 16129A 16187T 16189C 16223T 16293G 16311C 16360T 16519C	8468T 8619T 8655T 8701G 8860G 9072G 9540C 9861C 10084C 10321C 10398G 10586A 10688A 10810C 10861C 10873C 11719A 11899C 12681C 12705T 12810G 13105G 13485G 13506T 13650T
23000369	MH981855	B2i1	1	73G 8287del 263G 315.1C 430C 485C 499A 570G 827G 961C 1438G 2706G 3547G 4769G 4820A 4935R 4977C 6272G 6473T 7028T 8281del 8282del 8283del 8284del 8285del 8286del	8288del 8289del 8736C 8860G 9950C 11177T 11719A 13590A 14766T 15326G 15535T 16189C 16217C 16311C 16519C
23000370	MH981856	D1	1	73G 14783C 15043A 15326G 16223T 16252C 16362C 16519C	15043A 15326G 16111T 16223T 16274R 16290T 16319A 16362C
23000371	MH981857	B2o1a	1	73G 151T 152C 182T 186A 189C 195C 198T 247A 263G 297G 315.1C 316A 416A 508G 523del 524del 750G 769A 825A 1018A 1291C 1438G 1535de 1736G 2359G 2706G 4248C 4769G 4824G 7028T 8027A 8794T 8860G 9064G 9096C 10468T 11767T	1573A 15856A 18860G 95950C 11177T 11719A 13476G 13590A 14500G 14766T 15326G 15535T 16092C 16145A 16189C 16217C 16519C
23000372	MH981858	H4a1a4a	1	195C 263G 315.1C 523del 524del 750G 1438G 3992T 4024G 4769G 5004C 8269A 8860G 9123A 9615C 14365T 14569A 14582G 15326G 16233G	263G 315.1C 523del 524del 750G 1438G 3992T 4024G 4769G 5004C 8269A 8860G 9123A 9615C 14365T 14569A 14582G 15326G 16233G
23000373	MH981859	B2	1	73G 8287del 152C 263G 315.1C 460C 499A 570G 827G 1438G 2706G 3547G 4769G 4820A 4977C 5417A 6473T 7028T 7609C 8281del 8283del 8284del 8285del 8286del	8288del 8289del 8860G 9950C 11177T 11719A 12880C 14094C 14766T 15326G 15535T 16189C 16217C 16466G 16468C 16519C
23000374	MH981860	A2+(64)	1	64T 12705T 13926C 14766T 15326G 16084A 16111T 16215G 16223T 16290T 16319A 16362C 16519C	15043A 15326G 16111T 16223T 16274R 16290T 16319A 16362C
23000375	MH981861	H1e1a4	1	263G 315.1C 709A 750G 1438G 3010A 4769G 5460A 7196T 8512G 8860G 1113C 11371G 14902T 15326G 16311C 16519C	11719A 12007A 146C 263G 273T 315.1C 489C 750G 1438G 2092T 2706G 3010A 3316A 4769G 4883T 5178A 5773A 7028T 8414T 8701G 8860G 8934T 9540C 10398G 10400T 10873C
23000376	MH981862	D1e	1	73G 11248G 146C 263G 273T 315.1C 489C 750G 1438G 2092T 2706G 3010A 3316A 4769G 4883T 5178A 5773A 7028T 8414T 8701G 8860G 8934T 9540C 10398G 10400T 10873C	11719A 12705T 14668T 14766T 14783C 15043A 15326G 16223T 16325C 16362C
23000377	MH981863	C1b2	1	73G 9557T 10398G 10400T 10873C 11719A 11914A 12454A 12705T 13263G 14318C 14766T 15326G 15535T 16189C 16217C 16519C	14766T 15326G 16111T 16223T 16274R 16290T 16319A 16362C
23000378	MH981864	B2i1	1	73G 8287del 315.1C 430C 485C 499A 750G 827G 961C 1438G 2706G 3547G 4769G 4820A 4853R 4977C 6272G 6473T 7028T 8281del 8282del 8283del 8284del 8285del 8286del	8288del 8289del 8736C 8860G 9950C 11177T 11719A 13590A 14766T 15326G 15535T 16189C 16217C 16311C 16519C
23000379	MH981865	B2	1	57G 8284del 73G 151T 152C 207A 211G 263G 315.1C 499A 523del 524del 750G 827G 1438G 2706G 3372C 3547G 4769G 4820A 4977C 5417A 6473T 7028T 7609C 8281del 8282del 8283del 8284del 8285del 8286del	8285del 8286del 8860G 9950C 10364A 11177T 11719A 13590A 14527G 14766T 15326G 15535T 16093C 16189C 16217C 16316G 16360T 16519C
23000380	MH981866	J1c2e2	1	73G 12612G 13708A 14766T 14798C 15326G 15452A 16069T 16126C 16278T 16366T 16519C	185A 188G 228A 263G 295T 315.1C 462T 489C 523del 524del 709A 750G 1438G 2352C 2706G 3438A 3918A 4769G 6221C 6587T 7028T 8404C 8860G 10398G 11251G 11719A
23000381	MH981867	L3e1a1a	1	73G 11719A 12705T 14152G 14212C 14766T 15301A 15326G 15670C 15942C 16185T 16223T 16311C 16327T	150T 185A 189G 200G 263G 315.1C 750G 1438G 2352C 2706G 3438A 3918A 4769G 6221C 6587T 7028T 8404C 8860G 9540C 10398G 10819G 10873C 11332G

23000382	MH981868	C1b	1	73G 249del 263G 290del 291del 315.1C 316A 489C 493G 523del 524del 750G 1438G 2706G 3552A 4715G 4769G 4823C 6060C 7028T 7196A 8584A 8701G 8860G 9540C 9545G 9797C 10313C 10398G 10400T 10873C 11719A 11914A 12705T 13239T 13263G 14318C 14766T 14783C 15043A 15301A 15326G 15487T 16223T 16298C 16325C 16327T	16362C
23000383	MH981869	C1b	1	73G 249del 263G 290del 291del 315.1C 489C 493G 523del 524del 750G 1438G 2706G 3206T 3552A 4715G 4769G 7028T 7196A 8584A 8701G 8860G 9540C 9545G 10398G 10400T 10873C 11719A 11914A 12705T 13239T 13263G 14318C 14766T 14783C 15043A 15301A 15326G 15487T 16223T 16298C 16325C 16327T	
23000384	MH981870	A2q	1	64T 73G 146C 153G 235G 238G 263G 315.1C 523del 524del 750G 1438G 1736G 2706G 3816G 4188G 4248C 4769G 4824G 7028T 7581C 8027A 8794T 8860G 11719A 12007A 12705T 14766T 15326G 16111T 16209C 16223T 16290T 16301T 16319A	
23000385	MH981871	B2h	1	73G 263G 315.1C 499A 524.1A 524.2C 750G 827G 1438G 2706G 3547G 4769G 4820A 4977C 6473T 7028T 8281del 8282del 8283del 8284del 8285del 8286del 8287del 8288del 8860G 9950C 11177T 11719A 11821G 13590A 14094C 14410A 14766T 15326G 15535T 16189C 16217C 16468C 16519C	
23000386	MH981872	A2+(64)	1	64T 73G 146C 153G 235G 263G 271T 315.1C 523del 524del 663G 750G 1438G 1736G 2706G 4248C 4769G 4824G 7028T 7118G 8027A 8794T 8860G 10237C 11719A 12007A 12705T 12858T 14364A 14766T 15326G 16111T 16223T 16266T 16290T 16319A 16362C	
23000387	MH981873	D1e	1	73G 146C 263G 273T 315.1C 489C 750G 1438G 2092T 2706G 3010A 3316A 4769G 4883T 5178A 5773A 7028T 8414T 8701G 8860G 8934T 9540C 10398G 10400T 10873C 11248G 11719A 12705T 14484C 14668T 14783C 15043A 15301A 15326G 16223T 16325C 16362C	
23000388	MH981874	A2+(64)	1	64T 73G 146C 153G 235G 263G 315.1C 523del 524del 663G 750G 1438G 1736G 2706G 3397G 4248C 4769G 4824G 7028T 8027A 8794T 8860G 11719A 12007A 12705T 13926C 14766T 15326G 16084A 16111T 16215G 16223T 16290T 16319A 16362C	
23000389	MH981875	B2	1	57G 73G 151T 152C 207A 211G 263G 315.1C 499A 523del 524del 750G 827G 1438G 2706G 3372C 3547G 4769G 4820A 4977C 6473T 7028T 7664A 8281del 8282del 8283del 8284del 8285del 8286del 8287del 8288del 8860G 9950C 11177T 11719A 13590A 14527G 14766T 15326G 15535T 16093C 16189C 16217C 16316G 16360T 16519C	
23000390	MH981876	A2	1	73G 146C 153G 235G 263G 315.1C 523del 524del 663G 750G 1438G 1736G 2706G 4248C 4769G 4824G 7028T 8027A 8794T 8860G 9548A 9899C 11719A 12007A 12705T 13569C 14687G 14766T 15326G 16111T 16223T 16235G 16290T 16319A 16362C 16512C 16544C	
23000391	MH981877	HV+16311	1	263G 315.1C 750G 1438G 2706G 4769G 7028T 8860G 9801A 10205T 10920T 15326G 15514C 16311C 16519C 152G 182T 185T 195C 198T 247A 263G 315.1C 357G 523del 524del 709A 710C 750G 769A 825A 1018A 1462A 1738C 2352C 2706G 2758A 2768G 2885C 3308C 3594T 3666A 3693A 4104G 4769G 5036G 5046A 5393C 5655C 6548T 6827C 6989G 7028T 7055G 7146G 7256T 7389C 7521A 7867T 8248G 8468T 8655T 8701G 8860G 9540C 10398G 10688A 10810C 10873C 11002G 11719A 12519C 12705T 13105G 13506T 13789C 13880A 14178C 14203G 14560A 14766T 14769G 15115C 15326G 16126C 16187T 16189C 16223T	
23000392	MH981878	L1b1a12a	1	73G 152C 182T 185T 195C 198T 247A 263G 315.1C 357G 523del 524del 709A 710C 750G 769A 825A 1018A 1462A 1738C 2352C 2706G 2758A 2768G 2885C 15264T 16270T 16278T 16293G 16311C 16400T 16519C	
23000393	MH981879	D1	1	73G 210G 228A 263G 315.1C 489C 750G 1438G 2092T 2706G 3010A 4769G 4883T 5178A 6779G 7028T 8414T 8701G 8860G 9540C 10398G 10400T 10873C 11150A 11719A 12705T 14668T 14766T 14783C 15043A 15106A 15301A 15326G 16223T 16293T 16298C 16325C 16362C	
23000394	MH981880	A2+(64)	1	64T 73G 143A 146C 153G 235G 263G 315.1C 456T 523del 524del 663G 750G 1438G 1736G 2706G 2850C 3498T 4216C 4248C 4769G 4824G 7028T 8027A 8794T 8860G 8939C 8989C 11016A 11719A 12007A 12705T 13708A 14766T 15326G 15968C 16111T 16223T 16290T 16319A 16362C	
23000395	MH981881	C1d1	1	73G 194T 249del 263G 290del 291del 315.1C 489C 523del 524del 547R 750G 1438G 2706G 3338C 3552A 4048A 4652T 4715G 4769G 4955T 7028T 7196A 7354C 7697A 8584A 8701G 8860G 9540C 9545G 10398G 14040T 10873C 11167G 11719A 11914A 12705T 12930G 13263G 1418C 14766T 14783C 15034G 15043A 15301A 15326G 15487T 16051G 16223T 16287T 16298C 16311C 16325C 16327T	
23000396	MH981882	D4h3a	1	73G 152C 263G 315.1C 489C 523del 524del 750G 1438G 2242C 2706G 3010A 3336C 3396C 3644C 4025T 4025T 4769G 4883T 5048C 5178A 6285A 7028T 8414T 8701G 8860G 8946G 9458T 9540C 10256C 10398G 10400T 10873C 11719A 12705T 13135A 14668T 14766T 14783C 15043A 15301A 15326G 16223T 16241G 16296T 16301T 16342C 16362C	
23000397	MH981883	A2+(64)	1	64T 73G 143A 146C 153G 235G 263G 315.1C 456T 523del 524del 663G 750G 1438G 1736G 2706G 2850C 3498T 4216C 4248C 4769G 4824G 7028T 8027A 8794T 8860G 8939C 8989C 11016A 11719A 12007A 12705T 13708A 14766T 15326G 15968C 16111T 16223T 16234Y 16290T 16319A 16362C	
23000398	MH981884	A2	1	73G 146C 153G 235G 263G 315.1C 523del 524del 663G 750G 1438G 1736G 2706G 4248C 4769G 4824G 7028T 8027A 8794T 8860G 9548A 9899C 11719A 12007A 12705T 13569C 14687G 14766T 15326G 16111T 16223T 16235G 16290T 16319A 16362C 16512C 16544C	
23000399	MH981885	B2iI	1	34A 73G 263G 315.1C 430C 485C 499A 524.1A 524.2C 750G 827G 961C 1438G 2706G 3547G 4769G 4820A 4977C 6272G 6473T 7028T 8281del 8282del 8283del 8284del 8285del 8286del 8287del 8288del 8860G 9950C 11177T 11719A 13590A 14766T 15326G 15535T 16179del 16189C 16217C 16311C 16519C	
23000400	MH981886	D1	1	73G 146C 263G 315.1C 489C 750G 980C 1438G 2092T 2706G 3010A 3591A 4769G 4883T 5178A 6305T 7028T 8414T 8701G 8860G 9540C 9554A 10398G 10400T 10873C 11719A 12705T 13989T 14668T 14766T 14783C 15043A 15301A 15326G 15682G 16223T 16325C 16519C	
23000401	MH981887	D1e	1	73G 146C 263G 273T 315.1C 489C 750G 1438G 2092T 2706G 3010A 3316A 4769G 4883T 5178A 5773A 7028T 8414T 8701G 8860G 8934T 9540C 10398G 10400T 10873C 11248G 11719A 12705T 14484C 14668T 14783C 15043A 15301A 15326G 16223T 16325C 16362C	
23000402	MH981888	A+152+16362+200	1	73G 152C 200G 235G 263G 315.1C 523del 524del 663G 750G 1438G 1736G 2706G 4248C 4769G 4824G 7028T 8794T 8860G 11719A 12705T 14766T 15326G 16223T 16256T 16290T 16319A 16362C	
23000403	MH981889	C1b	1	73G 249del 263G 290del 291del 315.1C 489C 493G 523del 524del 750G 930A 1438G 2706G 3552A 4715G 4769G 6125G 7028T 7196A 8584A 8701G 8860G 9540C 9545G 10398G 10400T 10873C 11719A 11914A 12705T 13263G 14318C 14766T 14783C 15043A 15301A 15326G 15487T 16187T 16223T 16298C 16325C 16327T	
23000404	MH981890	C1d1d	1	73G 194T 249del 263G 290del 291del 315.1C 489C 523del 524del 750G 1438G 1736G 2706G 3552A 4715G 4769G 7028T 7196A 7697A 8584A 8701G 8860G 9540C 9545G 10398G 10400T 10873C 11719A 11914A 12705T 13263G 14318C 14766T 14783C 15043A 15301A 15326G 15487T 16051G 16172C 16223T 16298C 16325C 16327T	

23000407	MH981893	C1b	1	73G 10398G	249del 10400T	263G 10873C	290del 11719A	291del 11914A	315..1C 12705T	489C 13263G	493G 14318C	523del 14766T	524del 14783C	750G 15043A	1438G 15301A	2393T 15326G	3552A 15487T	4715G 16126C	4769G 16223T	7028T 16298C	7196A 16325C	8584A 16327T	8701G 16327T	8860G 16327T	9540C 16327T	9545G 16327T			
23000408	MH981894	B2	1	73G 8287del	93G 8288del	146C 8860G	195C 9950C	263G 11177T	315..1C 11719A	489C 11884G	750G 14766T	827G 14791T	1438G 15326G	2076G 15535T	3438A 16189C	3547G 16217C	4136G 16519C	4769G 16737T	4820A 16737T	4977C 16737T	6473T 16737T	7028T 16737T	8281del 16737T	8282del 16737T	8283del 16737T	8284del 16737T	8285del 16737T	8286del 16737T	
23000409	MH981895	H4a1a4a	1	152C	195C	263G	315..1C	523del	524del	750G	1438G	3992T	4024G	4769G	5004G	8269A	8860G	9123A	9615C	14365T	14569A	14582G	15326G	16233G					
23000410	MH981896	U2e1c1	1	73G 11719A	152C 12308G	217C 13246T	263G 13272A	315..1C 13734C	340T	508G	750G	1438G	8111G	2524C	2706G 16051G	3720G 16129C	4769G 16519C	5390G	5426C	5840T	6045T	6152C	7028T 16298C	7151T	8860G 16325C	10876G 16327T	11176A 16327T	11467G 16327T	
23000412	MH981898	C1d1	1	73G 9545G	194T 10398G	195C 10400T	249del 10783C	290del 11190T	315..1C 11719A	489C 12051A	523del	524del	750G	1438G	3552A 15043A	4715G 15301A	4769G 15326G	4820A 15487T	4976G 16051G	7028T 16093C	7196A 16223T	7697A 16223T	8584A 16223T	8588C 16223T	8701G 16223T	8860G 16223T	9540C 16223T	9545G 16223T	
23000413	MH981899	A2+(64)	1	64T	73G 12705T	146C 12858T	153G 14364A	235G 14766T	263G 15326G	315..1C 16111T	523del	524del	750G	1438E	1736G	2706G	4248C	4769G	4824G	7028T	7853A	8027A	8794T	8860G	10237C	11719A	12007A		
23000415	MH981900	C1d1	1	73G 8701G	194T 8860G	249del 9540C	263G 10398G	290del 10400T	315..1C 10873C	489C 11617G	523del	524del	750G	1438G	2706G	3338C	3552A	4048A	4652T	4715G	4769G	4955T	7028T	7196A	7354C	7697A	8584A		
23000416	MH981901	D1	1	73G 11719A	207A 12127A	263G 12705T	315..1C 14766T	489C 14783C	507A 15043A	1438G 15301A	2902T	2706G	3010A	3504C	4769G	4883T	5178A	5178A	7028T	8065A	8414T	8701G	8860G	9509C	9540C	10398G	10400T	10873C	
23000417	MH981902	C1d1	1	73G 10400T	194T 10873C	249del 11719A	263G 12693G	290del 13149G	315..1C 13263G	489C 13731G	523del	524del	750G	1438G	2706G	3552A	4715G	4769G	7028T	7196A	7697A	8584A	8701G	8860G	9540C	9545G	10398G	10398G	16519C
23000418	MH981903	B2i1	1	73G 8288del	263G 8289del	315..1C 8736C	430C 8860G	485C 9950C	499A 11177T	750G 11719A	827G 11590A	961C 14766T	1438G 15326G	2706G	3547G	4769G	4820A	4977C	6272G	6473T	7028T	8281del	8282del	8283del	8284del	8285del	8286del	8287del	
23000419	MH981904	C1d1b1	1	73G 8860G	194T 10398G	195C 10400T	249del 10873C	263G 11719A	315..1C 12075T	489C 12051A	523del	524del	750G 13149C	1438C 14783C	2706G	3553T	4715C 16189C	4769G 16217C	4899G 16319C	15043A 16519C	15301A 16519C	15326G 16519C	15487T 16051G	16092C 16223T	16223T 16298C	16223T 16325C	16223T 16519C	16223T 16519C	
23000510	MH981905	B2h	1	73G 8860G	263G 9950C	315..1C 11177T	499A 11719A	524..1A 11821G	524..2C 13590A	750G 14094C	827G 14410A	1438G 14766T	2706G	3547G	4769G	4820A	4977C	6473T	7028T	8281del	8282del	8283del	8284del	8285del	8286del	8289del			
23000511	MH981906	A2+(64)	1	64T 12705T	73G 13788A	146C 14766T	153G 15236G	235G 16111T	263G 16223T	315..1C 16290T	523del	524del	663G 16319A	750G 16362C	1438G 16519C	2706G	4248C	4769G	4824G	7028T	7918T	8027A	8794T	8860G	11719A	12007A	12354C		
23000512	MH981907	D1	1	73G 11638T	195C 11719A	263G 12705T	315..1C 14668T	489C 14767T	509C 15043A	1438G 15301A	2092T	2706G	3010A	4769G	4883T	5178A	6216C	7028T	8414T	8701G	8860G	9540C	9716C	10398G	10400T	10873C	11150A		
23000514	MH981909	B2e	1	73G 8288del	263G 8289del	315..1C 8860G	499A 9950C	524..1A 11177T	524..2C 11719A	750G 13590A	827G 14049T	986A 14766T	1438G 15326G	2706G	3547G	4769G	4820A	4977C	6119T	6473T	7028T	8281del	8282del	8283del	8284del	8285del	8286del	8287del	
23000515	MH981910	D1	1	73G 14783C	263G 15043A	315..1C 15326G	489C 16145A	750G 16172C	1438G 16223T	2092T	2706G	3010A	4769G	4883T	5178A	7028T	8414T	8701G	8860G	9540C	10398G	10400T	10873C	11171A	12705T	13557G	14668T	14766T	
23000517	MH981911	B2	1	73G 8288del	103A 8289del	146C 8860G	152C 9950C	455..1T 11177T	499A 11719A	509G 13590A	827G 13594G	1438G 14470C	2706G	3547G	4769G	4820A	4977C	6473T	7028T	8281del	8282del	8283del	8284del	8285del	8286del	8287del			
23000518	MH981912	D1	1	73G 14766T	263G 14783C	315..1C 15043A	489C 15326G	750G 16223T	2092T	2706G	3010A	4769G	4883T	5178A	6261A	7028T	8414T	8701G	8860G	9540C	10398G	10400T	10873C	11171A	11830C	12705T	14668T		
23000520	MH981914	C1d1d	1	73G 8860G	194T 9007G	249del 9540C	263G 10398G	290del 10400T	315..1C 10873C	489C 11488G	523del	524del	750G 11719A	1416G 11914A	1438G 12050T	2706G	3552A	4715G	4769G	6221C	7028T	7196A	7269A	7697A	8584A	8701G	8749C		
23000521	MH981915	H13a1a1	1	150T	257G	263G	315..1C	507A	1438G	2259T	4745G	4769G	7337A	8860G	13326C	13680T	14872T	15326G	16189C	16220G									
23000522	MH981916	B2	1	57G 8285del	73G 8286del	151T 8287del	260A 8288del	263G 8860G	315..1C 9950C	499A 11177T	523del	524del	750G 11719A	1438G 12780G	2706G 13590A	3372C 14766T	3547G 15326G	4769G 16189C	4820A 16217C	4977C 16316G	6473T 16519C	7028T	7830A	8281del	8282del	8283del	8284del	8285del	8286del
23000523	MH981917	D1	1	73G 14766T	263G 14783C	315..1C 15043A	489C 15326G	750G 16223T	1438G 16325C	2092T	2706G	3010A	4769G	4883T	5178A	6261A	7028T	8414T	8701G	8860G	9540C	10398G	10400T	10873C	11171A	12705T	14668T		
23000524	MH981918	C1b	1	73G 9540C	249del 10398G	263G 10873C	315..1C 11719A	489C 11914A	493G 12075T	523del	524del	750G 13263G	1438G 14318C	2706G	3552A	4715G	4769G	6221C	7028T	7196A	7269A	7697A	8584A	8701G	8860G				
23000525	MH981919	A2+(64)	1	64T 12007A	73G 12705T	146C 13708A	235G 13678C	315..1C 14605G	523del	524del	663G 16111T	750G 16223T	1438G 16362C	1736G	2706G	2850C	4216C	4769G	4824G	7028T	8027A	8794T	8860G	11719A					
23000526	MH981920	C1b2	1	72C 8860G	73G 9540C	249del 10400T	290del 10873C	315..1C 11719A	489C 11914A	523del	524del	750G 16130A	1438G 14545A	2706G	3552A	3915A	4191G	4242T	4715G	4769G	7013A	7028T	7196A	8584A	8633G	8701G			

23000527	MH981921	B2	1	73G 263G 315.1C 499A 750G 827G 1438G 2706G 3547G 4769G 4820A 4977C 6473T 7028T 8281del 8282del 8283del 8284del 8285del 8286del 8287del 8288del 8289del 8790A 8860G 9950C 11177T 11719A 13590A 14766T 15326G 15535T 16129A 16178C 16189C 16217C 16519C
23000528	MH981922	B2o	1	57G 73G 151T 260A 263G 307T 314del 315del 464G 499A 523del 524del 750G 827G 1438G 2706G 3372C 3547G 4769G 4820A 4977C 5585A 6473T 7028T 7830A 8281del 8282del 8283del 8284del 8285del 8286del 8287del 8288del 8289del 8701G 8860G 9540C 9545G 8288del 8289del 8860G 9540C 9545G
23000530	MH981924	C1b	1	73G 249del 263G 290del 291del 315.1C 489C 493G 523del 524del 750G 1438G 2706G 3552A 4715G 4736C 4769G 5054C 7028T 7196A 8584A 8701G 8860G 9540C 9545G 10398G 10400T 10873C 11719A 11914A 12705T 13263G 14318C 14766T 14783C 15043A 15301A 15326G 15487T 16223T 16298C 16325C 16327T 10398G 10400T 10873C 11719A 11914A 12705T 13263G 14318C 14553T 14766T 14783C 15043A 15301A 15326G 15487T 16223T 16298C 16325C 16327T
23000531	MH981925	C1b2	1	73G 249del 263G 290del 291del 315.1C 489C 493G 523del 524del 750G 1438G 2706G 3552A 4242T 4715G 4769G 7013A 7028T 7196A 8584A 8701G 8860G 9540C 9545G 9557T 10398G 10400T 10873C 11719A 11914A 12705T 13263G 14318C 14766T 14783C 15043A 15301A 15326G 15487T 16223T 16298C 16325C 16327T 9557T 10398G 10400T 10873C 11719A 11914A 12705T 13263G 14318C 14553T 14766T 14783C 15043A 15301A 15326G 15487T 16223T 16298C 16325C 16327T
23000532	MH981926	B2	1	73G 263G 315.1C 499A 750G 827G 1438G 2706G 3547G 4769G 4820A 4977C 6473T 7028T 8281del 8282del 8283del 8284del 8285del 8286del 8287del 8288del 8860G 9950C 11177T 11719A 12188C 13590A 14766T 15326G 15535T 16178C 16217C 16519C
23000533	MH981927	B2	1	73G 247A 263G 315.1C 499A 750G 827G 1438G 2706G 3547G 4769G 4820A 4977C 6473T 7028T 7055G 8281del 8282del 8283del 8284del 8285del 8286del 8287del 8288del 8289del 8860G 9128C 9950C 11177T 11719A 12007A 12245C 13590A 14766T 14962T 15326G 15535T 16189C 16217C 16519C
#! 1-10898 10962-16569				
23000405	MH981891	C1b	1	73G 249del 263G 290del 291del 315.1C 489C 493G 523del 524del 750G 930A 1438G 1888R 2706G 3552A 4715G 4736C 4769G 6125G 7028T 7196A 8584A 8701G 8860G 9540C 9545G 10398G 10400T 10873C 11719A 11914A 12705T 13263G 14318C 14766T 14783C 15043A 15301A 15326G 15487T 16187T 16223T 16298C 16325C 16327T
#! 1-14764 14767-16569				
23000406	MH981892	A2+(64)	1	64T 73G 146C 153G 235G 263G 315.1C 408A 523del 524del 573.1C 663G 750G 1438G 1736G 2706G 4248C 4769G 4824G 7028T 8027A 8794T 8843C 8860G 9482C 11172G 11719A 12007A 12705T 14040A 14502C 15326G 16104T 16111T 16223T 16290T 16319A 16362C 16519C
#! 1-7793 7855-16569				
23000411	MH981897	A2+(64)	1	64T 73G 146C 153G 189G 217C 235G 263G 315.1C 489C 523del 524del 663G 750G 1438G 1736G 2706G 2766T 4248C 4769G 4824G 6842C 6875A 7028T 7756T 8027A 8794T 8860G 10235C 11719A 12007A 12705T 14020C 14561G 14766T 15326G 16111T 16185T 16223T 16290T 16310A 16319A 16362C 16519C
#! 1-285 287-16569				
23000513	MH981908	C1b	1	73G 249del 263G 290del 291del 315.1C 489C 493G 523del 524del 750G 1438G 2706G 3221R 3552A 4715G 4736C 4769G 5054C 7028T 7196A 8584A 8701G 8860G 9540C 9545G 10398G 10400T 10873C 11719A 11914A 12705T 13263G 14318C 14766T 14783C 15043A 15301A 15326G 15487T 16136C 16223T 16298C 16325C 16327T
#! 1-513 516-10424 10465-16569				
23000519	MH981913	A2	1	73G 146C 153G 235G 263G 315.1C 523del 524del 663G 750G 1438G 1736G 2706G 4248C 4769G 4824G 7028T 8027A 8794T 8860G 9096C 11719A 12007A 12705T 14127C 14766T 15326G 15951G 16111T 16126C 16223T 16259T 16290T 16319A 16327T 16362C 16519C
#! 1-10425 10465-16569				
23000529	MH981923	D1	1	73G 263G 315.1C 489C 750G 1438G 2092T 2706G 3010A 4769G 4883T 5178A 6261A 7028T 8414T 8701G 8860G 9540C 10398G 10400T 10873C 11719A 11830Y 12705T 14668T 14766T 14783C 15043A 15301A 15326G 16223T 16325C 16362C 16519C

**Table S3.** Pairwise  $F_{ST}$  values between populations from Alto Paraná\*, Ecuador\*, Peru\* and Colombia\* (below diagonal) and corresponding differentiation p-values (above diagonal)

	Yucu-Matapi	Curripaco	Ach-Piapoco	Carijona	Desano	Other-ET	Pira-Wanano	Siriano	Tanimuka	Tuka-Tatuyo	Siona	Coreguaje	Sikuani	Guayabero	Saliba	Mur-Uitoto	Puinave	Nukak	Pasto	Kamentsa	Inga	Tikuna	Cocama	Yagua	Ecuador admixed	Ecuador native	Colombia	Peru	Alto Pará	
Yucu-Matapi	*	0.00069	0.00752	0.28680	0.16246	0.08514	0.1875	0.28888	0.27829	0.15048	<0.00005	0.03455	0.00168	0.00079	0.03841	0.03326	0.01049	0.02079	0.05504	0.19058	0.01337	0.06920	0.01693	0.00257	<0.00005	<0.00005	<0.00005	<0.00005	0.00050	
Curripaco	0.17363	*	0.00079	0.09474	0.02386	0.01020	0.02386	0.06653	0.01148	0.04257	0.01020	0.00406	0.00020	0.03089	0.00109	0.00356	0.13622	0.02069	0.03653	0.00851	0.00168	0.00050	0.28017	0.00683	0.21315	0.09811	0.00970	0.18800	0.03524	
Ach-Piapoco	0.09707	0.20299	*	0.06247	0.09078	0.24483	0.11484	0.27720	0.06564	0.11771	0.00059	0.00297	0.00446	0.01901	0.01733	0.00851	0.00198	0.00059	0.08316	0.04366	0.08960	0.08405	0.02287	0.03455	<0.00005	<0.00005	<0.00005	<0.00005	0.00059	
Carijona	0.01406	0.09417	0.10865	*	0.41154	0.10237	0.29314	0.36086	0.17721	0.27453	0.00337	0.46669	0.04534	0.09435	0.18761	0.14117	0.28948	0.20503	0.15256	0.44223	0.04643	0.08385	0.21760	0.02891	0.03168	0.00703	0.00059	0.02198	0.16721	
Desano	0.02229	0.10734	0.04975	-0.00160	*	0.48134	0.96466	0.96476	0.11395	0.91852	0.00079	0.06653	0.00604	0.06564	0.14949	0.57153	0.06970	0.02020	0.68894	0.35284	0.10326	0.18523	0.20236	0.03277	0.00683	0.00040	0.00050	0.00406	0.22523	
Other-ET	0.05952	0.17673	0.01691	0.08900	-0.01135	*	0.58489	0.74735	0.16781	0.69785	0.00762	0.01406	0.01564	0.09356	0.09494	0.25453	0.02208	0.00307	0.57460	0.10949	0.32423	0.30898	0.10593	0.12751	0.00663	0.00010	0.00554	0.00327	0.09435	
Pira-Wanano	0.02195	0.12453	0.04779	0.01698	-0.04719	-0.02597	*	0.89645	0.11534	0.98485	0.00218	0.04772	0.01208	0.07039	0.12791	0.65518	0.04554	0.01723	0.61588	0.26938	0.12009	0.28007	0.15127	0.04208	0.00733	<0.00005	<0.00005	0.00129	0.00366	0.19127
Siriano	0.00941	0.09818	0.01083	0.00288	-0.05609	-0.04798	-0.05510	*	0.25403	0.78428	0.01782	0.08366	0.01968	0.23295	0.02267	0.41788	0.09415	0.04761	0.71298	0.28760	0.32373	0.36383	0.33581	0.12187	0.03406	0.00129	0.01139	0.02218	0.33442	
Tanimuka	0.01089	0.18158	0.09371	0.08098	0.05595	0.05006	0.06046	0.01871	*	0.14801	0.01960	0.01950	0.00614	0.05099	0.02455	0.06108	0.03198	0.02030	0.15325	0.07336	0.16632	0.14424	0.09801	0.07296	0.00525	0.00020	0.00436	0.00079	0.01980	
Tuka-Tatuyo	0.03489	0.10917	0.05841	0.02490	-0.05005	-0.04155	-0.07106	-0.05236	0.05880	*	0.00842	0.04346	0.01129	0.13484	0.76131	0.07871	0.02831	0.74191	0.21305	0.15088	0.27106	0.23928	0.06574	0.03416	0.00069	0.00495	0.02416	0.34066		
Siona	0.25203	0.14569	0.20643	0.25781	0.18742	0.17524	0.20256	0.15166	0.16990	0.18258	*	<0.00005	<0.00000	0.00248	<0.00000	0.00010	0.00079	<0.00000	0.02228	0.00010	0.02030	<0.00000	0.03109	0.02079	0.00545	0.0010	0.04693	0.00178	0.00030	
Coreguaje	0.06316	0.20667	0.17645	-0.00952	0.05644	0.15802	0.08432	0.07590	0.16367	0.10148	0.34479	*	0.00564	0.00178	0.07455	0.00772	0.03792	0.04613	0.01257	0.29720	0.00327	0.00802	0.01168	0.00109	0.00010	<0.00005	<0.00005	0.00010	0.00099	
Sikuani	0.16938	0.34558	0.21771	0.15401	0.16981	0.24256	0.18152	0.19098	0.27693	0.21303	0.45298	0.17829	*	0.00238	0.00762	0.00040	0.00030	0.00119	0.00337	0.01574	0.00386	0.00040	<0.00005	<0.00005	<0.00000	<0.00000	<0.00000	<0.00000	0.00005	
Guayabero	0.10977	0.08700	0.07826	0.08093	0.05360	0.06179	0.06166	0.02838	0.09732	0.06082	0.14809	0.16862	0.18777	*	0.00448	0.00396	0.00792	0.00149	0.01028	0.01802	0.02574	0.01109	0.10118	0.02079	0.00010	<0.00005	0.00020	0.00020	0.00257	
Saliba	0.06227	0.21893	0.12670	0.03449	0.03134	0.07258	0.04020	0.02605	0.13731	0.04736	0.32151	0.05462	0.16568	0.13469	*	0.03772	0.01544	0.01049	0.05217	0.25443	0.00931	0.04029	0.01406	0.00337	0.00010	<0.00005	<0.00005	0.00010	0.00257	
Mur-Uitoto	0.04673	0.13610	0.09539	0.04379	-0.0174	0.01507	-0.01746	-0.00226	0.07527	-0.02847	0.20579	0.10403	0.20620	0.09018	0.05983	*	0.01208	0.00238	0.61394	0.12910	0.21305	0.02505	0.08356	0.06257	0.00495	0.00404	<0.00005	<0.00000	0.00050	0.08088
Puinave	0.06649	0.04213	0.17573	0.01736	0.06246	0.15148	0.09075	0.07240	0.14049	0.08369	0.22156	0.09836	0.27310	0.11045	0.13279	0.09508	*	0.06158	0.04404	0.08336	0.03007	0.00634	0.21731	0.00832	0.02921	0.01129	0.00010	0.02891	0.06257	
Nukak	0.09218	0.16523	0.25252	0.04791	0.12216	0.24380	0.14981	0.14394	0.20058	0.15834	0.37938	0.09175	0.28618	0.19183	0.15014	0.15123	0.09771	*	0.00752	0.07158	0.00020	0.00188	0.02069	0.00059	<0.00005	<0.00005	<0.00000	0.00010	0.00139	
Pasto	0.06003	0.09336	0.05726	0.05133	-0.02175	-0.02122	-0.02350	-0.03403	0.04776	-0.03903	0.11219	0.13318	0.23277	0.04580	0.07651	-0.01569	0.08674	0.18612	*	0.11187	0.25374	0.17087	0.33264	0.10019	0.06287	0.00069	0.01109	0.03336	0.40451	
Kamentsa	0.02419	0.18675	0.11536	-0.00779	0.00229	0.07744	0.01344	0.01068	0.11477	0.03014	0.31223	0.01108	0.18224	0.01269	0.01843	0.03801	0.08603	0.10656	0.05940	*	0.05059	0.10682	0.06356	0.01386	0.00168	0.00030	0.00010	0.00149	0.04495	
Inga	0.10373	0.18957	0.05491	0.13221	0.04497	0.00381	0.04419	0.00217	0.04261	0.04432	0.11739	0.20464	0.28746	0.08061	0.15340	0.07708	0.17928	0.28392	0.01521	0.11273	*	0.11029	0.05039	0.08316	0.00099	<0.00005	0.00238	0.00010	0.00752	
Tikuna	0.04470	0.20943	0.05333	0.08236	0.02119	0.00684	0.00892	-0.00183	0.04770	0.01501	0.21957	0.13068	0.21403	0.10459	0.08980	0.03943	0.14944	0.20194	0.02817	0.05967	0.04639	*	0.02990	0.04158	0.00030	<0.00005	<0.00005	0.000683		
Cocama	0.08395	0.01146	0.09278	0.02773	0.02109	0.05910	0.03388	0.00563	0.07438	0.01914	0.09076	0.12370	0.25421	0.04198	0.11865	0.05042	0.02215	0.13523	0.00347	0.08158	0.07187	0.08562	*	0.06346	0.48104	0.02525	0.03425	0.25007	0.47579	
Yagua	0.15715	0.16699	0.10886	0.15925	0.09279	0.05959	0.09684	0.05790	0.09726	0.08363	0.12250	0.24699	0.34000	0.10026	0.20958	0.13019	0.17750	0.30000	0.05860	0.19112	0.06236	0.09741	0.07359	*	0.00198	0.00020	0.00960	0.00099	0.00881	
Ecuador admixed	0.14159	0.01175	0.15015	0.08647	0.07797	0.11647	0.09345	0.06999	0.12733	0.07385	0.08894	0.17318	0.28192	0.08207	0.17266	0.09102	0.04828	0.16849	0.04235	0.13917	0.11725	0.14990	-0.00533	0.11440	*	0.00396	0.00010	0.23651	0.00446	
Ecuador native	0.25007	0.02927	0.28485	0.17323	0.20075	0.27058	0.22841	0.20462	0.26102	0.21037	0.19107	0.25999	0.39609	0.18390	0.28731	0.21588	0.08422	0.22900	0.17465	0.25184	0.26473	0.29030	0.06593	0.23983	0.03201	*	<0.00005	0.00574	<0.00005	
Colombia	0.22857	0.09564	0.18540	0.21758	0.16455	0.15882	0.17663	0.13569	0.17491	0.15861	0.05836	0.30599	0.38631	0.12063	0.28732	0.18481	0.16731	0.31342	0.10298	0.26470	0.13260	0.21008	0.06181	0.11742	0.05850	0.13237	*	<0.00000	<0.00000	
Peru	0.16349	0.01286	0.18117	0.10267	0.09224	0.14157	0.11148	0.09241	0.16294	0.08936	0.11992	0.19099	0.31055	0.10220	0.18829	0.10216	0.05247	0.18244	0.05893	0.15786	0.15274	0.18032	0.00828	0.14330	0.00268	0.03402	0.07755	*	0.00036	
Alto Paraná	0.07009	0.05029	0.09207	0.03478	0.00928	0.04219	0.01675	0.00574	0.08095	0.00348	0.12784	0.010521	0.21176	0.05936	0.09078	0.02110	0.03450	0.12217	-0.00143	0.05676	0.07399	0.07274	-0.00469	0.09179	0.02793	0.10865	0.09882	0.03005	*	

Significant p-values after Bonferroni's correction are indicated in red (p&lt;0.0001)

s.e.&lt;0.0053

**APÊNDICE E** - Material suplementar Artigo 4 - The maternal inheritance of the *Ashaninka native group from Peru*

## Supplementary Table 1 – Haplotypes from Ashaninka, Peru

Sample ID	Haplotype																			
AD003	16051G	16129A	16183C	16189C	16217C	16519C	73G	152C	263G	309.1C	315.1C	499A								
AD005	16093C	16223T	16298C	16325C	16327T	73G	146C	249del	263G	290del	291del	309.1C	315.1C	489C	493G	523del	524del			
AD006	16362C	16519C	60A	73G	263G	309.1C	309.2C	315.1C	489C											
AD008	16093C	16223T	16298C	16325C	16327T	73G	146C	249del	263G	290del	291del	309.1C	315.1C	489C	493G	523del	524del			
AD009	16111T	16223T	16239A	16266T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	309.2C	315.1C	523del	524del		
AD010	16111T	16223T	16239A	16266T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	309.2C	315.1C	523del	524del		
AD011	16093C	16223T	16298C	16325C	16327T	73G	146C	249del	263G	290del	291del	309.1C	315.1C	489C	493G	523del	524del			
AD014	16093C	16223T	16298C	16325C	16327T	73G	146C	249DEL	263G	290DEL	291DEL	309.1C	315.1C	489C	493G	523DEL	524DEL			
AD017	16093C	16223T	16298C	16325C	16327T	73G	146C	249DEL	263G	290DEL	291DEL	309.1C	315.1C	489C	493G	523DEL	524DEL			
AD018	16111T	16223T	16239A	16266T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	309.2C	315.1C	523del	524del		
AD019	16325C	16362C	16519C	60A	73G	263G	309.1C	319.2C	315.1C	489C										
AD021	16093C	16223T	16298C	16325C	16327T	73G	146C	249DEL	263G	290DEL	291DEL	309.1C	315.1C	489C	493G	523DEL	524DEL			
AD024	16111T	16129A	16223T	16239A	16266T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del		
AD032	16051G	16129A	16183C	16189C	16193.1C	16217C	16519C	73G	152C	263G	309.1C	309.2C	315.1C	499A						
AD033	16051G	16129A	16183C	16189C	16193.1C	16217C	16519C	73G	152C	263G	309.1C	309.2C	315.1C	499A						
AD037	16111T	16223T	16290T	16319A	16399G	16519C	64T	73G	146C	153G	182T	200G	235G	263G	315.1C	523Del	524Del			
AD038	16111T	16223T	16239A	16266T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	309.2C	315.1C	523del	524del		
AD040	16325C	16362C	16519C	60A	73G	263G	309.1C	309.2C	315.1C	489C										
AD041	16111T	16223T	16290T	16319A	16399G	16519C	64T	73G	146C	153G	182T	200G	235G	263G	315.1C	523Del	524Del			
AD042	16325C	16362C	16519C	60A	73G	263G	309.1C	309.2C	315.1C	489C										
AD043	16325C	16362C	16519C	60A	73G	263G	309.1C	309.2C	315.1C	489C										
AD045	16223T	16258G	16259T	16302G	16325C	16362C	16519C	73G	263G	315.1C	489C									
AD047	16223T	16258G	16259T	16302G	16325C	16362C	16519C	73G	263G	315.1C	489C									
AD048	16325C	16362C	16519C	60A	73G	263G	309.1C	309.2C	315.1C	489C										
AD049	16223T	16258G	16259T	16302G	16325C	16362C	16519C	73G	263G	315.1C	489C									
AD050	16111T	16223T	16290T	16319A	16399G	16519C	64T	73G	146C	153G	182T	200G	235G	263G	315.1C	523Del	524Del			
AD052	16223T	16258G	16259T	16302G	16325C	16362C	16519C	73G	263G	315.1C	489C									
AD053	16093C	16223T	16298C	16325C	16327T	73G	146C	249del	263G	290Del	291Del	309.1C	315.1C	489C	493G	523Del	524Del			
AD054	16093C	16223T	16298C	16325C	16327T	73G	146C	249del	263G	290Del	291Del	309.1C	315.1C	489C	493G	523Del	524Del			
AD056	16093C	16223T	16298C	16325C	16327T	73G	146C	249del	263G	290del	291del	309.1C	315.1C	489C	493G	523del	524del			
AD057	16111T	16192T	16223T	16239A	16266T	16290T	16319A	16362C	16519C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del	
AD059	16051G	16129A	16183C	16189C	16193.1C	16217C	16519C	73G	152C	263G	309.1C	309.2C	315.1C	499A						
AD060	16093C	16223T	16298C	16325C	16327T	73G	146C	249del	263G	290del	291del	309.1C	315.1C	489C	493G	523del	524del			
AD061	16051G	16129A	16183C	16189C	16193.1C	16217C	16519C	73G	152C	263G	309.1C	315.1C	499A							
AD063	16223T	16258G	16259T	16302G	16325C	16362C	16519C	73G	263G	315.1C	489C									
AD064	16051G	16129A	16183C	16189C	16217C	16519C	73G	152C	263G	309.1C	315.1C	499A								
AD065	16111T	16223T	16290T	16319A	16362C	73G	103A	146C	152C	153G	183G	235G	263G	309.1C	315.1C	456T	523del	524del		
AD067	16111T	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	315.1C	340T	523del	524del					
AD068	16051G	16129A	16183C	16189C	16217C	16519C	73G	152C	263G	309.1C	315.1C	499A								
AD069	16093C	16183C	16189C	16193.1C	16217C	16362C	16519C	56.1A	57C	73G	150T	217C	263G	309.1C	315.1C	499A	514T			
AD071	16223T	16258G	16259T	16302G	16325C	16362C	16519C	73G	263G	315.1C	489C									
AD072	16111T	16144C	16223T	16266T	16290T	16319A	16362C	16437C	16519C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	356.1C	499A	523del
AD073	16223T	16325C	16362C	73G	263G	309.1C	315.1C	489C												
AD074	16111T	16223T	16290T	16319A	16399G	16519C	64T	73G	146C	153G	182T	200G	235G	263G	315.1C	523del	524del			
AD075	16093C	16223T	16298C	16325C	16327T	73G	146C	249del	263G	290del	291del	309.1C	315.1C	489C	493G	523del	524del			
AD077	16111T	16223T	16290T	16319A	16399G	16519C	64T	73G	146C	153G	182T	200G	235G	263G	315.1C	523del	524del			
AD080	16126C	16223T	16298C	16311C	16325C	16327T	16519C	73G	249del	263G	290del	291del	309.1C	315.1C	489C	493G	523del	524del		
AD081	16111T	16223T	16290T	16319A	16399G	16519C	64T	73G	146C	153G	182T	200G	235G	263G	315.1C	523del	524del			
AD082	16093C	16223T	16298C	16325C	16327T	73G	146C	249del	263G	290del	291del	309.1C	315.1C	489C	493G	523del	524del			







## APÊNDICE F - Pôster “Tracing Arawakan Southern Dispersal: Clues from Mitogenome Sequencing of Southern Amerinds”

# Tracing Arawakan Southern Dispersal: Clues from Mitogenome Sequencing of Southern Amerinds



C. Xavier<sup>1</sup>, F. Simão<sup>2</sup>, M. Bodner<sup>1</sup>, C. Strobl<sup>1</sup>, D. Silva<sup>2</sup>, C. Barletta<sup>3</sup>, A. Casas-Vargas<sup>4</sup>, W. Usaqué<sup>4</sup>, D.H. Tineo<sup>5</sup>  
L. Gusmão<sup>2,6</sup> and W. Parson<sup>1,7</sup>

<sup>1</sup>Institute of Legal Medicine, Medical University of Innsbruck, Innsbruck, Austria

<sup>2</sup>DNA Diagnostic Laboratory (LDD), State University of Rio de Janeiro (UERJ), Rio de Janeiro, Brazil

<sup>3</sup>Laboratorio de Genética Humana, Universidad Nacional Mayor de San Marcos, Lima, Peru

<sup>4</sup>Group of Population Genetics and Identification, Genetics Institute, National University of Colombia, Bogotá, Colombia

<sup>5</sup>Instituto de Medicina Legal, Universidad Nacional Mayor de San Marcos, Lima, Peru

<sup>6</sup>Institute of Pathology and Immunology of the University of Porto (IPATIMUP), Porto, Portugal

<sup>7</sup>Forensic Science Program, The Pennsylvania State University, PA, USA

catarina.gomes@i-med.ac.at



P03-202  
170320

## Introduction

South America's colonization and internal migration routes remain unsolved issues for the scientific community. The wide variety of different environmental settings and the amplitude of climate variations at the time of colonization might explain the big variation of different tribes and groups in South America, in contrast to its northern counterpart [1].

The Arawakan are the largest and most widespread linguistic family in South America, being distributed at time of European arrival from the Caribbean Islands to the Brazilian wetlands. The Arawakan were also a distinct Amazonian group which would stand out from other Amazonian tribes by their advanced socio-cultural societies, the agricultural and fishing economic practices as well as complex commercial and political alliances [1, 2]. Linguistic and anthropological studies situate western Amazonia (in the border between Amazonian Peru and Brazil) as the most probable origin of the Arawakan language (Fig.1, [3]). Afterwards Arawakan are thought to have migrated alongside riverbanks throughout the continent, spreading their cultural and agricultural traditions, naming their diaspora the South America's Neolithic (Fig.1, [3]).

For this study we analysed samples from three different Arawakan communities: the Wayúu in northern Colombia, the Ashaninka in pre-Andine Peru and the Terena in the Brazilian wetlands. Furthermore we gathered a sample set from a Shipibo-Conibo tribe, inhabiting the shores of the Ucayali river in Peruvian Amazon. The Shipibo are a Panoan speaking tribe which engaged in several warfare encounters and intermarriage with Arawakans (Piro) living in the same region [4]. All samples have been typed for the full control region of the mitochondrial DNA (mtDNA), the information of the found lineages allowed a triage of samples to undergo the full mitogenome sequencing.

A deeper study of under-described populations as well as the increase of the study's resolution allows a better comprehension of the Arawakan migratory movements at a minor scale as well as increase the size of a forensic and population mtDNA database - EMPOP [5].

## Materials & Methods

The study was approved by Ethics Commissions. Sampling was performed under informed consent. All samples were extracted using a phenol-chloroform protocol or a Chelex based protocol and quantified using an in house real time qPCR method [6]. Libraries were built using the Precision ID mtDNA Whole Genome Panel (Thermo Fisher Scientific) and sequenced using the IonChef and IonS5 system (TFS). A validation study of this method is presented in P01-23.

The sequences were compared to the revised Cambridge Reference Sequence (CRS) [7] using the Ion Torrent Variant Caller v. 5.2.1.38 plugin (TFS) and all these positions have been posteriorly checked with IGV (Integrative Genome Viewer). Haplotypes were determined by two separate analysts and haplogroups were assigned according to PhyloTree (build 16) [8]. Furthermore, all haplotypes have been submitted to EMPOP4 (P06-422) for alignment and haplogroup check.

In order to understand the distribution of the lineages found in this study and close neighbors, literature research and database queries were done.

## Discussion

The previous sequencing of the full CR of the mtDNA helped us detecting one lineage (B2+16051G+16360T+152C) that was shared between the Terena and the Shipibo-Conibo (Fig.2). The deeper investigation of this particular lineage at the highest resolution possible (the full mtDNA genome) verifies the lineages present in Peru and Brazil are very similar (Fig.2). A thorough look into literature identified the existence of the same lineage in Peru (Fig.1). If only the CR or partial CR is investigated the same lineage appears along the Peruvian lowlands, the Titicaca basin and the Bolivian Moxos.

Furthermore, we identified a closely related lineage present in the Peruvian Ashaninka (B2+16051G+16129A+152C) that was further sequenced for the full mtDNA genome (Fig.2). The same lineage (full genome) was found in other Peruvian samples and previously described (only CR) in Matsiguenga and Yanesha tribes, all pre-Andine Arawakans (Fig.1). This might indicate a regional diversification of this specific lineage which also agrees with the linguistic differentiation between pre-Andine Arawakan and Proto-Piro-Apurina-Baure-Ignaciano Arawakan (western Amazonian lowlands and Southern Arawakans) in which Terena and the neighbor groups of Shipibo (the Piro) are included.

Other lineages have been chosen for further full mtDNA genome sequencing, either because they have not yet been described in the literature or presented a typical motive in their geographical panorama. For example, B2d lineages in Colombia have been identified previously and seem to be restricted to northwestern South America and Mesoamerica territories [9].

These results reinforce the fact that intense trading systems and alliances with neighbor groups with strong intermarriage policies have shaped the Arawakan genetic panama. However, particularly in western Amazonia /Bolivian Llanos and Wetlands the B2+16051G + 16360T + 7786C + 152C stands out as a possible reminiscence of an Arawakan southern migration. Characterized as riverine groups with developed watercraft techniques, a southern migration alongside the Ucayali and Urumbamba rivers until reaching Madre de Dios River and finally the Pantanal wetlands in Brazil would not be surprising.

## Results

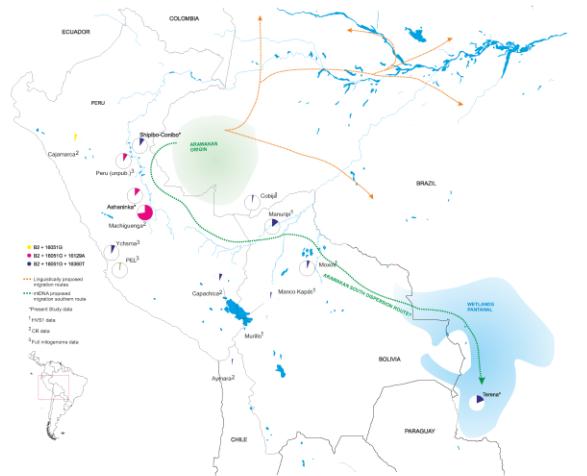


Fig. 1 B2+16051G (yellow), B2+16051G+16129A (pink), B2+16051G+16360T (dark blue) lineages distribution across the Peruvian lowlands, lake Titicaca and Bolivian Moxos. Shaded area is the linguistic proposed migration route by the Arawakan. Green arrow is the linguistic proposed migration route by the mtDNA results. Shaded green area is the linguistic proposed location of origin of the Arawakan language, shaded blue area represents the wetlands covered area. Other populations' data retrieved from [10-15].

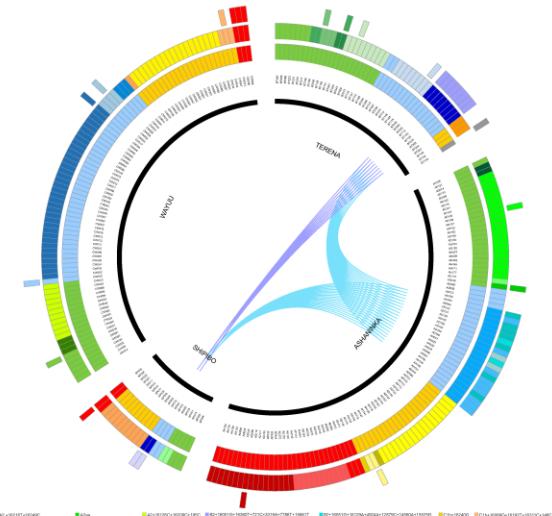


Fig. 2 Graphical representation of the mtDNA lineages found in the four studied groups. Inner black circular segments represents each of the groups. Inner graphic shows the composition of mtDNA lineages for the four major Native American haplogroups and L2. Middle graphic shows variation within each macrohaplogroup and outer graphic depicts the samples chosen for full mtGenome sequencing and diversity found. Links between sample groups reflect close lineage(s) of B2+16051G in dark (+16360T+7786T) and light blue (+16129C). Plot designed using Circos [16].

## References & Acknowledgements

- [1] Xavier C et al (2014) PLoS ONE 9(10):e110454
- [2] Hill JD & Slatkin M (2000) University of Illinois Press
- [3] Walker RS & Sorenson GR (2002) University of Illinois Press
- [4] Laharp D (1970) Thèmes & Thodes
- [5] Parson W & Díaz A (2007) Int J Legal Genet 51:28-34.
- [6] Andrews RM et al (1999) Nature Genetics 23:147.
- [7] Van Oven M et al (2009) Hum Mutat 30(2):386-394 www.phylotree.org
- [8] Xavier C et al (2014) PLoS ONE 9(10):e110454
- [9] Sandvol S et al (2013) PLoS ONE 8(9)
- [10] The 1000 Genomes Project (2015) Nature 526(7657):68-74
- [11] Valverde G et al (2016) PLoS ONE 11(6)
- [12] Tito RS et al (unpub)
- [13] Tito RS et al (2013) PLoS ONE 8(3)
- [14] de Saint Pierre M et al (2012) PLoS ONE 7(9)
- [15] de Saint Pierre M et al (2012) PLoS ONE 7(9)
- [16] Krywinski M et al. (2009) Genome Res 19:1639-1645

C. Xavier had, at the time of analysis, a FCT grant (SFRH/BD/90873/2012) from Fundação para a Ciência e Tecnologia (FCT). M. Bodner was supported by Theodor Körner Fonds zur Förderung von Wissenschaft und Kunst. The authors would like to thank Bettina Zimmermann, Gabriela Huber and Simone Nagl for technical help and analysis of the Sanger sequencing data previous to this work.

