

CAROLINE DA SILVA HYPPOLITO

Compilação e análise crítica sobre a utilização do DNA mitocondrial em casos forenses na população brasileira em comparação com outros países da América Latina e do mundo

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

Orientadora: Profa. Dra. Regina Maria Barretto Cicarelli

Araraquara
2017

FICHA CATALOGRÁFICA

H998c Hyppolito, Caroline da Silva
Compilação e análise crítica sobre a utilização do DNA
mitocondrial em casos forenses na população brasileira em
comparação com outros países da América Latina e do mundo /
Caroline da Silva Hyppolito. – Araraquara : [s.n.], 2017
76 f. : il.

Dissertação (mestrado) – Universidade Estadual Paulista,
Instituto de Química
Orientador: Regina Maria Barretto Cicarelli

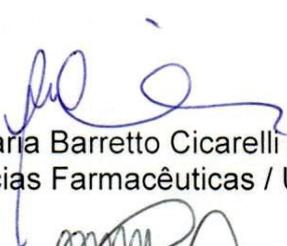
1. DNA. 2. Genética forense. 3. Haplotipos. 4. Variação
genética. 5. Sequenciamento de nucleotídeo. I. Título.

CAROLINE DA SILVA HYPOLITO

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

Araraquara, 14 de fevereiro de 2017.

BANCA EXAMINADORA



Prof^a Dr^a Regina Maria Barretto Cicarelli (Orientadora)
Faculdade de Ciências Farmacêuticas / UNESP / Araraquara - SP



Prof^a Dr^a Raquel Mantuaneli Scarel Caminaga
Faculdade de Odontologia / UNESP / Araraquara - SP



Dr^a Joyce Aparecida Martins Lopes Ferraz
Faculdade de Medicina / FAMERP / São José do Rio Preto - SP

DADOS CURRICULARES

IDENTIFICAÇÃO

Nome: Caroline da Silva Hyppolito

ENDEREÇO PROFISSIONAL

Instituto Médico Legal

Rua Professora Ana Ramos de Carvalho, 349, Nova Jaboticabal

Jaboticabal/SP

(16) 3202 8062

FORMAÇÃO ACADÊMICA

Bacharel em Biomedicina pela Universidade Federal do Triângulo Mineiro (UFTM).

ATUAÇÃO PROFISSIONAL

Auxiliar de Necropsia da Polícia Técnico-Científica do Estado de São Paulo

PARTICIPAÇÃO EM EVENTOS CIENTÍFICOS

- Apresentação do trabalho intitulado “**Papel do banco de dados de DNA mitocondrial nas investigações criminais**” no II Congresso Internacional sobre Bancos de Perfis Genéticos para fins de Persecução Criminal, realizado na Universidade do Vale do Rio Sinos (UNISINOS), em São Leopoldo (RS).

AGRADECIMENTOS

A Deus, pela oportunidade e força para trilhar esse caminho.

Ao Instituto de Química da UNESP de Araraquara, pelo ambiente amigável e repleto de recursos que proporciona. Aos funcionários da Seção Técnica de Pós Graduação, por estarem sempre dispostos a ajudar e a esclarecerem dúvidas.

À Profa. Dra. Regina Maria Barretto Cicarelli, pela orientação, pelos ensinamentos, correções, suporte no pouco tempo que lhe coube, incentivos, oportunidade e apoio na elaboração deste trabalho.

Aos professores e doutores que avaliaram o meu trabalho e contribuíram para o meu aprimoramento e crescimento.

Aos meus amigos do Laboratório de Paternidade, Danilo Braganholi, Fernanda Polverari, Isabela Brunelli e Juliana Martinez, pelos ensinamentos e conselhos compartilhados e disposição em me ajudar nas etapas difíceis. Aos demais amigos do laboratório, que me acolheram e fizeram parte da minha formação.

À minha mãe, família e amigos que me apoiaram e confortaram nos momentos difíceis, não permitindo que eu desanimasse e nem desistisse dos meus objetivos.

Ao CNPq pela bolsa concedida.

A todos que de alguma forma contribuíram na minha formação, o meu muito obrigado.

Resumo

O DNA mitocondrial é uma molécula circular, dupla fita, que se encontra na mitocôndria, uma organela responsável pela respiração e produção de energia da célula. Este DNA é herdado exclusivamente por via materna e possui polimorfismos que permitem diferenciar indivíduos, auxiliando nas investigações forenses e corroborando com outros exames de identificação humana. Devido à possibilidade de compatibilidade deste DNA entre indivíduos não relacionados por via materna, é necessário estimar as frequências haplotípicas utilizando bancos de dados populacionais. Com a finalidade de se compilar e avaliar a utilização do DNA mitocondrial no Brasil realizou-se um estudo sobre os haplótipos depositados no banco de dados do *European DNA Profiling Mitochondrial DNA Population Database* (EMPOP) e acessando artigos científicos sobre dados de populações brasileiras, observando os países participantes e seus respectivos dados depositados. Na avaliação dos dados do EMPOP, a América do Norte foi o continente com maior número de haplótipos depositados no banco, principalmente pela participação dos Estados Unidos. A América do Sul contém 3172 haplótipos, sendo 782 desse total correspondentes ao Brasil. Esses números parecem insuficientes para representar a população grande e miscigenada do país. Na literatura, encontrou-se um total de 2367 haplótipos, os quais foram organizados por haplogrupos e as regiões brasileiras correspondentes: Sudeste e Nordeste apresentaram-se similares, com predomínio de ancestralidade africana (44%), enquanto o Sul apresentou uma origem predominantemente europeia (68,4%) e a região Norte de nativo-americanos (58,8%). As regiões Centro-Oeste, Norte e Nordeste carecem de mais estudos, pois os dados referentes a essas regiões são muito escassos. A população brasileira demonstrou alta diversidade genética que reflete sua miscigenação, sendo que na avaliação da distância genética, as populações do Sul do Brasil ficaram mais próximas da população portuguesa, enquanto que as do Norte se aproximaram da africana. Os dados da literatura permitem concluir que, a matrilinearidade do Brasil é predominantemente de origem africana (37%), seguida da europeia (35%) e de nativo-americanos (28%). Além disso, a diferença entre as quantidades de dados brasileiros no banco e na literatura indica a dificuldade dos laboratórios em atender aos controles de qualidade exigidos para deposição de amostras no EMPOP. Para avaliação da complexidade da metodologia de análise do DNA mitocondrial, cinco amostras de referência foram sequenciadas para as regiões HV1, HV2 e HV3 e os resultados confirmaram as sequências informadas previamente. Isto permitiu verificar que, com a utilização de uma metodologia adequadamente padronizada para análise de DNA mitocondrial foi possível obter resultados satisfatórios, mesmo que o operador tenha pouca experiência na técnica. Assim, a utilização correta de uma metodologia padronizada permitirá atender aos controles de qualidade exigidos pelo EMPOP, ainda que a técnica apresente várias etapas críticas para a obtenção do sequenciamento de qualidade.

Palavras-chave: DNA mitocondrial - Brasil. Banco de dados. EMPOP.

Abstract

Mitochondrial DNA is a circular and double-stranded molecule found in the mitochondria, an organelle responsible for breathing and energy production of the cell. This DNA is exclusively inherited through maternal and has polymorphisms that allow differentiating individuals, it assists in forensic investigations and corroborates with other tests of human identification. Due to the possibility of compatibility of this DNA between unrelated individuals, it is necessary to estimate the haplotype frequencies through population databases. In order to compile and evaluate the use of mitochondrial DNA in Brazil, a study was carried out about the haplotypes deposited in the European DNA Profiling Mitochondrial DNA Population Database (EMPOP), and accessing scientific articles with data from Brazilian populations, verifying participating countries and their respective deposited data. In the evaluation of EMPOP data, North America was the continent with the highest number of data deposited in the bank, mainly by the participation of the United States. South America contains 3172 haplotypes, 782 of which correspond to Brazil. These numbers seem insufficient to represent the large and mixed population of this country. In the literature, a total of 2367 haplotypes were found, which were organized by haplogroups and the corresponding Brazilian regions: the Southeast and Northeast were similar, with predominance of African ancestry (44%), while the South had a predominantly European origin (68.4%) and the North of Native Americans (58.8%). The Center-West, North and Northeast regions need further studies, therefore, the data for these regions are very scarce. Brazilian population have shown high genetic diversity that reflects your miscegenation, and in the genetic distance evaluation, the populations of the South of Brazil are genetically closer to the Portuguese population, while those from the North are closer to the African ones.

According to literature data, it is possible to conclude that Brazil's matriarchy is predominantly African (37%), followed by European (35%) and Native American (28%). In addition, the difference between the Brazilian data in the database and the literature indicates the difficulty of the laboratories to meet the quality controls required for EMPOP deposition. To evaluate the complexity of the mitochondrial DNA analysis methodology, five reference samples had the HV1, HV2 and HV3 regions sequenced and the results confirmed the previously reported sequences. This allows to verify that with the use of a adequately standardized methodology for mitochondrial DNA analysis it is possible to obtain satisfactory results, even the operator having little experience in the technique. Therefore, the correct use of a standardized methodology, allows to meet the quality controls required by EMPOP, even the technique presents several critical steps to obtain quality sequencing.

Keywords: Mitochondrial DNA - Brazil. Database. EMPOP.

LISTA DE FIGURAS

Figura 1 – Estrutura e composição da mitocôndria	16
Figura 2 – Ilustração do DNA mitocondrial	17
Figura 3 – Exemplos de Polimorfismos de Nucleotídeo Único (SNP)	18
Figura 4 – Exemplo de heteroplasmia de posição	20
Figura 5 – Exemplo de heteroplasmia de comprimento	20
Figura 6 – Árvore filogenética mitocondrial apresentando os haplogrupos encontrados nas populações mundiais	23
Figura 7 – Distribuição mundial dos haplogrupos devido às migrações	24
Figura 8 – Mapa do Brasil indicando os principais Estados onde se estabeleceram os imigrantes	26

LISTA DE GRÁFICOS

Gráfico 1 – Porcentagem de haplótipos depositados no EMPOP por países Americanos	40
Gráfico 2 – Comparação da porcentagem de haplótipos fornecidos ao EMPOP entre continentes e subcontinentes	44
Gráfico 3 – Diversidade genética e tamanho da população (492 indivíduos) conforme o haplótipo estudado – HV1, HV2, HV3 e região controle.	45
Gráfico 4 - Número de haplótipos e tamanho da população (492 indivíduos) conforme o haplótipo estudado – HV1, HV2, HV3 e região controle	48
Gráfico 5 – Porcentagem da composição ancestral nas regiões brasileiras analisadas pelos dados da literatura	59
Gráfico 6 – Distâncias genéticas (<i>Fst</i>) das populações brasileiras e estrangeiras	67

LISTA DE TABELAS

Tabela 1 – Relação dos trabalhos brasileiros utilizados neste estudo	32
Tabela 2 – Iniciadores utilizados na PCR e no sequenciamento	35
Tabela 3- Número de haplótipos depositados no EMPOP por países da América do Sul	38
Tabela 4 – Número de haplótipos depositados no EMPOP por países da América Central	39
Tabela 5 - Número de haplótipos depositados por países da América do Norte no EMPOP	39
Tabela 6 - Número de haplótipos depositados por países da Europa no EMPOP	41
Tabela 7 – Número de haplótipos depositados no EMPOP por países da Ásia	42
Tabela 8 – Número de haplótipos depositados no EMPOP por países da África	43
Tabela 9 – Países com maior número de dados depositados no EMPOP	44
Tabela 10 – Diversidade Genética e tamanho da população estudada	46
Tabela 11 – Total de haplótipos em função da região analisada e do tamanho da população estudada	50
Tabela 12 – Quantidade de haplótipos classificadas em haplogrupos da região Sudeste avaliados na literatura	52
Tabela 13 - Quantidade de haplótipos classificadas em haplogrupos da região Sul avaliados na literatura	54
Tabela 14 - Quantidade de haplótipos classificadas em haplogrupos da região Nordeste avaliados na literatura	56
Tabela 15 - Quantidade de haplótipos classificadas em haplogrupos da região Norte avaliados na literatura	58

Tabela 16 – Comparação dos dados brasileiros avaliados na literatura com os depositados no EMPOP	60
Tabela 17 – Valores de <i>Fst</i> e <i>p value</i> das distâncias genéticas entre as populações	62
Tabela 18 – Valores de <i>Fst</i> e <i>p value</i> das distâncias genéticas entre as populações sulistas	63
Tabela 19 – Valores de <i>Fst</i> e <i>p value</i> das distâncias genéticas entre as populações do Norte	64
Tabela 20 - Valores de <i>Fst</i> e <i>p value</i> das distâncias genéticas entre as populações estudadas	65
Tabela 21 – Sequências analisadas e seus respectivos haplogrupos	68

LISTA DE ABREVIATURAS E SIGLAS

A	Adenina
C	Citosina
CODIS	<i>Combined DNA Index System</i>
DNA	Ácido desoxirribonucleico
EDNAP	European DNA Profiling Group
EMPOP	<i>EDNAP Mitochondrial DNA Population Database</i>
EUA	Estados Unidos da América
FBI	<i>Federal Bureau of Investigation</i>
G	Guanina
GHEP	Grupo de Línguas Espanhola e Portuguesa da <i>International Society for Forensic Genetics</i>
HV1	Região hipervariável 1
HV2	Região hipervariável 2
HV3	Região hipervariável 3
Indel	Inserção e/ou deleção de base
ISFG	<i>International Society for Forensic Genetics</i>
ISFH	<i>International Society of Forensic Haemogenetics</i>
IUPAC	<i>International Union of Pure and Applied Chemistry</i>
mtDNA	DNA mitocondrial
mt-MRCA	<i>Matrilineal most Recent Common Ancestor</i>
pb	pares de base
rCRS	<i>revised Cambridge Reference Sequence</i>
RIBPG	Rede Integrada de Bancos de Perfis Genéticos
RNA	Ácido ribonucleico
rRNA	RNA ribossômico
SNPs	Polimorfismos de Nucleotídeo Único
T	Tiamina
tRNA	RNA transportador

SUMÁRIO

1	INTRODUÇÃO	14
2	REVISÃO BIBLIOGRÁFICA	16
2.1	DNA mitocondrial	16
2.2	Polimorfismos De Nucleotídeo Único (SNPs)	18
2.3	Heteroplasmia	19
2.4	Utilização do DNA mitocondrial nas investigações forenses	21
2.5	Árvore Filogenética do DNA Mitocondrial	22
2.6	Distribuição de origens ancestrais pelo território brasileiro	24
2.7	Bancos de dados de DNA para finalidade forense	27
2.8	Banco de dados do EMPOP	28
3	OBJETIVOS	30
3.1	Objetivo geral	30
3.2	Objetivos específicos	30
4	METODOLOGIA	31
4.1	Análise das quantidades de dados no EMPOP	31
4.2	Tamanho amostral em relação à diversidade genética e número de haplótipos	31
4.3	Análise de dados brasileiros disponíveis na literatura	31
4.4	Cálculos estatísticos	33
4.5	Sequenciamentos de amostras referência (Promega e GHEP-ISFG)	33
4.5.1	AMOSTRAS	33
4.5.2	EXTRAÇÃO DE DNA	34
4.5.3	REAÇÃO EM CADEIA DA POLIMERASE (PCR)	34
4.5.4	PURIFICAÇÃO DOS PRODUTOS DA PCR	35
4.5.5	SEQUENCIAMENTO DO DNA	35
4.5.6	PURIFICAÇÃO E PRECIPITAÇÃO	36
4.5.7	ELETROFORESE CAPILAR	37
4.5.8	ANÁLISE DOS POLIMORFISMOS	37
4.5.9	CLASSIFICAÇÃO EM HAPLOGRUPOS	37
5	RESULTADOS E DISCUSSÕES	38
5.1	Quantidades de dados no EMPOP	38
5.2	Tamanho amostral, diversidade genética e total de haplótipos	45

5.3	Relação de haplogrupos distribuídos nos territórios brasileiros	51
5.3.1	SUDESTE	51
5.3.2	SUL	53
5.3.3	NORDESTE	55
5.3.4	NORTE	57
5.3.5	COMPARAÇÃO ENTRE OS DADOS DO BRASIL NO EMPOP E DA LITERATURA	60
5.4	Distâncias genéticas das populações brasileiras	60
5.4.1	SUDESTE	61
5.4.2	SUL	62
5.4.3	NORDESTE	63
5.4.4	NORTE	63
5.4.5	DISTÂNCIA GENÉTICA DAS POPULAÇÕES	64
5.5	Sequenciamento do mtDNA em amostras referência	68
6	CONCLUSÕES	70
	REFERÊNCIAS	72

1 INTRODUÇÃO

No âmbito forense, o DNA mitocondrial (mtDNA) possui propriedades que o destacam do DNA autossômico, como a grande quantidade de cópias por célula e alta resistência às condições degradantes do meio. Essas propriedades fazem do mtDNA uma ótima ferramenta para analisar amostras com DNA nuclear degradado ou ausente, como é observado em amostras de pelos e cabelos sem bulbo, ossadas, restos humanos carbonizados, dentre outros (PARSON et al., 2004).

O mtDNA é localizado na mitocôndria, uma organela responsável pela respiração e produção de energia da célula (CANN, 2001; SHOKOLENKO; ALEXEYEV, 2015). O mtDNA pode ser utilizado nas investigações forenses na identificação humana (SULLIVAN; HOPGOOD; GILL, 1992).

O mtDNA foi primeiramente sequenciado por Anderson e colaboradores em 1981 (ANDERSON et al., 1981), e essa sequência foi posteriormente revisada por Andrews e colaboradores em 1999. Esta sequência é utilizada como referência para se analisar outros perfis genéticos mitocondriais e é denominada de *revised Cambridge Reference Sequence* (rCRS) (ANDREWS et al., 1999).

A análise do mtDNA se baseia na identificação de polimorfismos por meio da comparação entre a sequência em análise e a rCRS. Os principais polimorfismos investigados no mtDNA são os Polimorfismos de Nucleotídeo Único (SNPs) encontrados na região controle ou não codificante. Ao comparar o mtDNA de uma amostra com o da vítima ou suspeito, pode-se obter três tipos de resultados: inclusão, quando todos os polimorfismos são compatíveis; indeterminado, quando apenas um polimorfismo difere entre as amostras; e exclusão, quando há mais de um polimorfismo diferente (SCIENTIFIC WORKING GROUP ON DNA ANALYSIS METHODS, 2013).

Devido à sua herança exclusivamente materna e sem recombinação, o DNA mitocondrial é transferido de forma única e inalterada (com exceção de mutações). Assim, os indivíduos relacionados maternamente possuem o mesmo mtDNA (GILES et al, 1980). Porém, há a possibilidade de encontrar indivíduos não relacionados por via materna com o mesmo mtDNA, pois, embora não possuam vínculo materno, possuem uma mesma origem ancestral (HOLLAND; PARSONS, 1999).

Para contornar este problema de compatibilidade entre indivíduos não aparentados, é necessário estimar a frequência dos perfis genéticos mitocondriais na população, por meio da contagem do número de vezes que os mesmos

aparecem na população estudada (SCIENTIFIC WORKING GROUP ON DNA ANALYSIS METHODS, 2013). Os pesquisadores podem depositar e compartilhar esses perfis em bancos de dados populacionais (PARSON et al., 2004).

O EMPOP é um banco de dados populacional de mtDNA criado pelo grupo *European DNA Profiling* (EDNAP), com a finalidade de compartilhar dados confiáveis. Estes dados podem ser acessados na plataforma online (www.empop.org); qualquer laboratório que participe dos exercícios colaborativos do controle de qualidade pode submeter seus dados populacionais. Atualmente, o EMPOP é o banco mais utilizado pelos laboratórios forenses ao redor do mundo (PARSON et al., 2004; PRIETO et al., 2011).

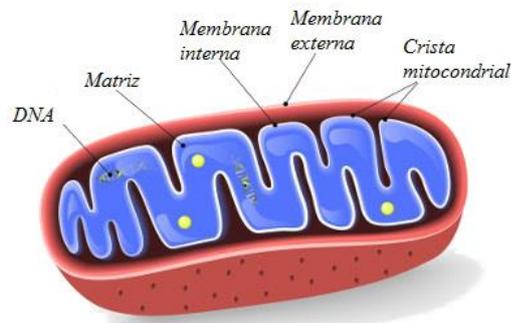
Com o propósito de avaliar a utilização do mtDNA nas investigações forenses no Brasil, neste estudo realizou-se um levantamento da quantidade de dados brasileiros contidos no EMPOP e aqueles publicados em artigos científicos. Por meio desse levantamento foi possível determinar a distribuição das origens ancestrais nas diferentes regiões brasileiras e comparar com outras populações mundiais. Este estudo apontou a importância de depositar e compartilhar perfis genéticos no banco de dados mitocondrial, para colaborar na obtenção de resultados estatísticos confiáveis a serem aplicados em investigações forenses.

2 REVISÃO BIBLIOGRÁFICA

2.1 DNA mitocondrial

A mitocôndria é uma organela encontrada em grande quantidade no citoplasma celular, cerca de 5000 por célula. Ela é composta por dupla membrana (interna e externa), pelas cristas mitocondriais que se formam das dobras da membrana interna e de uma matriz, onde se encontra o DNA mitocondrial (mtDNA) (figura 1). Esta organela é responsável pela respiração e pela produção de energia da célula, sendo a maioria das proteínas e RNAs, que atuam nessa função, codificada por genes do mtDNA (CANN, 2001; SHOKOLENKO; ALEXEYEV, 2015).

Figura 1 – Estrutura e composição da mitocôndria. A mitocôndria possui duas membranas, uma externa que a envolve, e uma interna que se dobra e forma as cristas mitocondriais. A matriz mitocondrial é envolta pela membrana interna e possui no seu interior moléculas como proteínas, RNAs e o DNA mitocondrial.



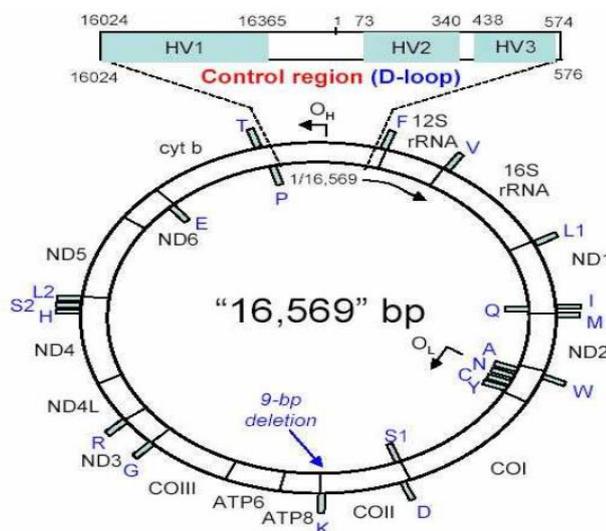
Fonte: Santos (2017).

O mtDNA é circular, constituído por dupla fita, sendo uma leve, rica em bases pirimídicas (T e C), e outra pesada, rica em purinas (A e G). Em 1981, Anderson e colaboradores sequenciaram completamente o genoma mitocondrial, que é composto por 16.569 pares de bases. Dessa forma, a denominação das posições de cada base se dá do número 1 ao 16.569. Essa sequência foi reanalisada e revisada por Andrews e colaboradores em 1999. Eles constataram que, na verdade, o mtDNA possui 16.568 pb, mas mantiveram a numeração original de 16.569 pb, indicando uma deleção na posição 3107. A sequência modificada foi denominada *revised Cambridge Reference Sequence* (rCRS). Atualmente, a rCRS é utilizada como uma

sequência referência para a análise de outras sequências do mtDNA (ANDREWS *et al.*, 1999).

O mtDNA apresenta duas regiões, uma denominada de região codificante, onde se encontram os 37 genes responsáveis pela síntese de RNAs transportadores (tRNA) e ribossômicos (rRNA), e de proteínas envolvidas na respiração celular e na produção e estoque de energia na célula (ANDERSON *et al.*, 1981). A outra região, denominada de D-loop, não-codificante ou controle, possui os sítios de replicação e transcrição da molécula, e inclui as regiões hipervariáveis: HV1 (da posição 16024 a 16365), HV2 (da posição 73 a 340) e HV3 (da posição 438 a 574) (figura 2). As regiões controle e codificante possuem marcadores de grande importância nas áreas forense e antropológica (BUDOWLE *et al.*, 1999; GREENBERG; NEWBOLD; SUGINO, 1983; LUTZ *et al.*, 2000).

Figura 2 – Ilustração do DNA mitocondrial. O DNA mitocondrial é composto por dupla fita circular, uma leve (O_L) e outra pesada (O_H). O mtDNA pode ser dividido em duas regiões distintas: a maior, denominada codificante, possui 37 genes responsáveis pela síntese de proteínas e RNAs que atuam na respiração celular; e a menor, chamada de controle (*Control region*), está representada na parte superior da figura, com cerca de 1200 pares de base, que pode ser dividida em três sítios altamente polimórficos (destacados em azul), denominados de HV1, HV2 e HV3. Devido à inserção e deleção existente ao longo do genoma nos diferentes indivíduos, nem sempre seu comprimento é de 16.569 pb.

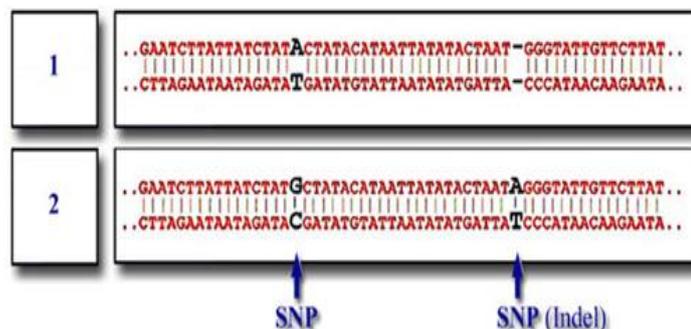


Fonte: Adaptado de Butler (2011).

2.2 Polimorfismos De Nucleotídeo Único (SNPs)

As variações genéticas são chamadas de polimorfismos quando presentes em uma frequência maior que 1% na população. No mtDNA, os polimorfismos mais frequentes são os de nucleotídeo único (*Single Nucleotide Polymorphism – SNP*) (figura 3): transições (trocas de bases púricas por púricas e pirimídicas por pirimídicas) e transversões (trocas de bases púricas por pirimídicas e vice-versa), em uma proporção de 40 transições para 1 transversão. Também é comum pequenas inserções e deleções nos trechos homo poliméricos (trechos poli-C) nas regiões entre as posições 302 e 310 e entre 16183 e 16184 (TULLY et al, 2001).

Figura 3 – Exemplos de Polimorfismos de Nucleotídeo Único (SNP). (1) Sequência de mtDNA sem variação e (2) sequência de mtDNA com SNP de substituição de A/T por G/C (SNP) e inserção (SNP – Indel) de par de base A/T.



Fonte: United States Department of Agriculture (2007).

As variações genéticas são determinadas por meio da comparação das sequências em análise com a rCRS (CARRACEDO et al., 2000). Dessa forma, as bases da sequência investigada que diferem da rCRS são os polimorfismos, e estes devem ser relatados de acordo com regras de nomenclatura.

A nomenclatura dos polimorfismos tem que estar em conformidade com as recomendações da IUPAC (*International Union of Pure and Applied Chemistry*) (SCIENTIFIC WORKING GROUP ON DNA ANALYSIS METHODS, 2013). As substituições são demonstradas pelo número da posição em que se encontram, seguida pela letra correspondente ao nucleotídeo modificado. Por exemplo, se na rCRS, na posição 73 (em HV2) há uma adenina (A), enquanto que na sequência estudada há uma guanina (G), denomina-se o SNP como 73G. As posições que não

são descritas entende-se como igual ao observado na rCRS (CARRACEDO et al., 2000).

As inserções são demonstradas pela posição, com a adição de um ponto decimal, seguida pelo número da inserção (“1” primeira inserção, “2” segunda inserção etc) e por último a letra correspondente ao nucleotídeo inserido (Ex: 309.1C, 309.2C). Nos casos de deleção, a mesma deve ser indicada pelo “DEL”, “del” ou “-“, relatando a posição onde ocorreu tal mutação, como por exemplo, a deleção de uma citosina na posição 574 (574DEL). A letra N é usada quando qualquer um dos quatro nucleotídeos pode ser encontrado na referida posição, enquanto que em outras misturas heteroplásmicas, utilizam-se outras letras recomendadas pela IUPAC, como o R na possibilidade de A e G, e o Y na possibilidade de C (citosina) e T (timina) (PARSON et al., 2014).

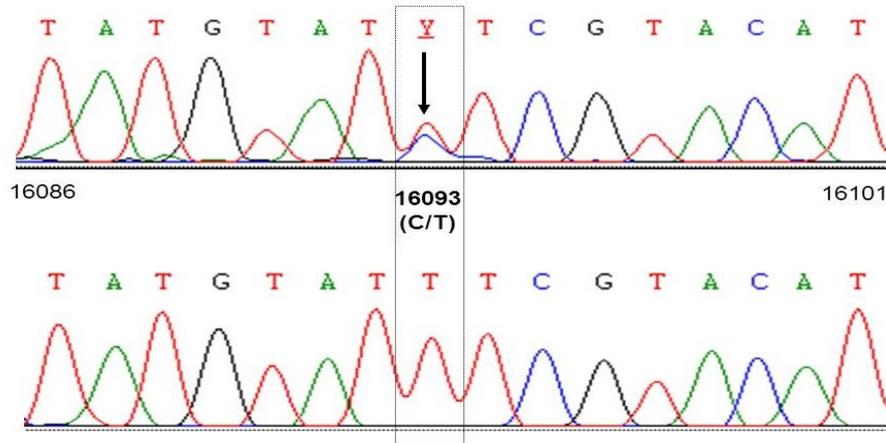
A maioria dos laboratórios que utilizam tipagem do mtDNA baseiam-se nos SNPs presentes na sequência de nucleotídeos da região controle. O mtDNA é herdado de forma única e sem recombinação, exclusivamente da mãe. Por se tratar de um genoma que advém apenas de um dos genitores, o mtDNA é denominado de haplótipo. O haplótipo possui polimorfismos ou marcadores que permitem identificar geneticamente um indivíduo (BUTLER, 2011; GILES et al, 1980).

2.3 Heteroplasmia

Devido à sua natureza altamente suscetível a mutações, há a possibilidade de existirem dois ou mais tipos de mtDNA em uma mesmo indivíduo. Este fenômeno é denominado de heteroplasmia e pode ser encontrada em uma única célula, em um mesmo tecido ou em tecidos diferentes (SALAS; LAREU; CARRACEDO, 2001).

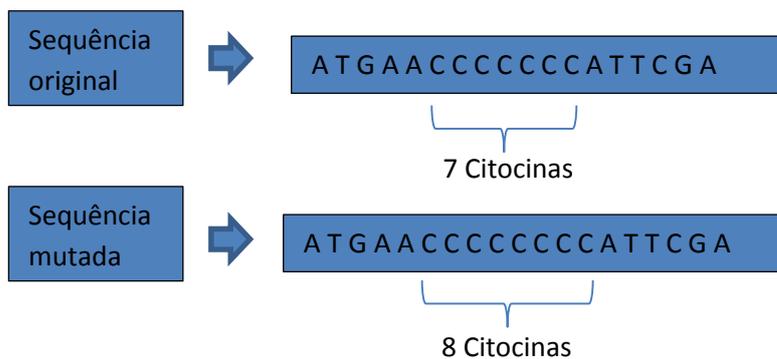
No sequenciamento, a heteroplasmia de posição é representada pela presença de duas bases ocupando um mesmo *locus* (figura 5), enquanto que a heteroplasmia de comprimento é caracterizada pela inserção ou deleção de bases na sequência (figura 6), frequentemente observada nas regiões homopolímeras de citosina, na posição 16184 – 16193 da HV1, 303 – 315 de HV2 e 568 – 573 de HV3 (PANETO, 2010).

Figura 4 – Exemplo de heteroplasmia de posição. O eletroferograma inferior apresenta um nucleotídeo T na posição 16093. O eletroferograma superior apresenta heteroplasmia de posição demonstrada pela sobreposição dos nucleotídeos C/T na mesma posição.



Fonte: Adaptado de Butler (2011).

Figura 5 – Exemplo de heteroplasmia de comprimento. Inserção de um nucleotídeo C em um trecho homopolimérico, resultando em uma sequência com 8 citosinas. Dessa forma, a sequência original (7 citosinas) difere da mutada (8 citosinas) pelo comprimento.



Fonte: da autora.

A heteroplasmia pode gerar resultados ambíguos dificultando a interpretação dos casos, por exemplo, quando uma amostra apresenta o fenômeno e a outra não. Por outro lado, a heteroplasmia pode aumentar a sensibilidade para discriminar e identificar indivíduos, quando os pontos de heteroplasmia se encontram na mesma posição no mtDNA das amostras comparadas (HOLLAND; PARSONS, 1999; IVANOV et al., 1996).

2.4 Utilização do DNA mitocondrial nas investigações forenses

A identidade é um conjunto de características que permitem individualizar pessoas, animais, vegetais e objetos. A identificação engloba os processos destinados a definir uma identidade. Portanto, a identificação humana tem por objetivo a diferenciação entre pessoas, ou seja, determinar as características que tornam cada indivíduo único (BENFICA; VAZ, 2015).

No âmbito forense, são utilizados diversos processos destinados à identificação humana, como: a papiloscopia, que é o estudo das impressões papilares dos dedos, mãos e pés e que não se repetem nem mesmo no próprio indivíduo; a odontologia forense, que estuda a conformação da arcada dentária; a antropologia forense, que analisa restos mortais para determinar o sexo, a estatura, idade e características de lesões ou doenças que levaram o indivíduo à morte (SOUZA, 2013); e os exames de DNA, que se baseiam em marcadores genéticos dos cromossomos autossômicos, sexuais e mtDNA.

A utilização do mtDNA é vantajosa devido ao fato de uma única célula possuir várias cópias do mesmo mtDNA e pela conformação circular do mesmo, o que aumenta as chances de que algumas cópias de mtDNA irão perdurar em ambientes altamente desfavoráveis e em amostras com pouco ou nenhum DNA nuclear presente (ossadas, pelos, cabelos sem bulbo etc). Portanto, a análise dessa molécula permite agregar informações nas investigações e corrobora com outros exames (PARSONS; COBLE, 2001; BENFICA; VAZ, 2015).

Além disso, devido a sua herança exclusivamente materna, falta de recombinação e resistência ao meio, o mtDNA pode ser essencial nos casos de desaparecimento, desastres naturais ou acidentes, comparando-se o genótipo da mãe ou de um parente que possua relação materna com o indivíduo a ser identificado, ou até mesmo com o DNA retirado dos pertences pessoais da vítima, como as escovas de dentes e de cabelo (BUDOWLE et al., 1999; LUTZ et al., 2000; PARSON et al., 2004; PRIETO et al., 2011). Além disso, segundo o perito criminal federal, Dr. Jorge Marcelo de Freitas (comunicação pessoal)¹, o mtDNA também pode ser utilizado na identificação de espécies animais nos crimes de contrabando.

¹ FREITAS, J. M. de. Utilização do DNA mitocondrial e de bancos de dados nas investigações criminais pelos laboratórios forenses do Brasil. Araraquara, São Paulo, Laboratório de Paternidade da UNESP, 12 nov. 2015. Entrevista realizada pela autora.

De acordo com o *Scientific Working Group on DNA Analysis Methods* (2013), as análises forenses de mtDNA baseiam-se na compatibilidade entre os polimorfismos encontrados no mtDNA da amostra retirada da cena onde ocorreu o fato com o do suspeito ou vítima. Todos os polimorfismos devem ser compatíveis entre as amostras para ter uma inclusão ou apresentar vínculo materno. Um polimorfismo diferente gera um resultado inconclusivo e acima de duas diferenças classifica-se o resultado como uma exclusão.

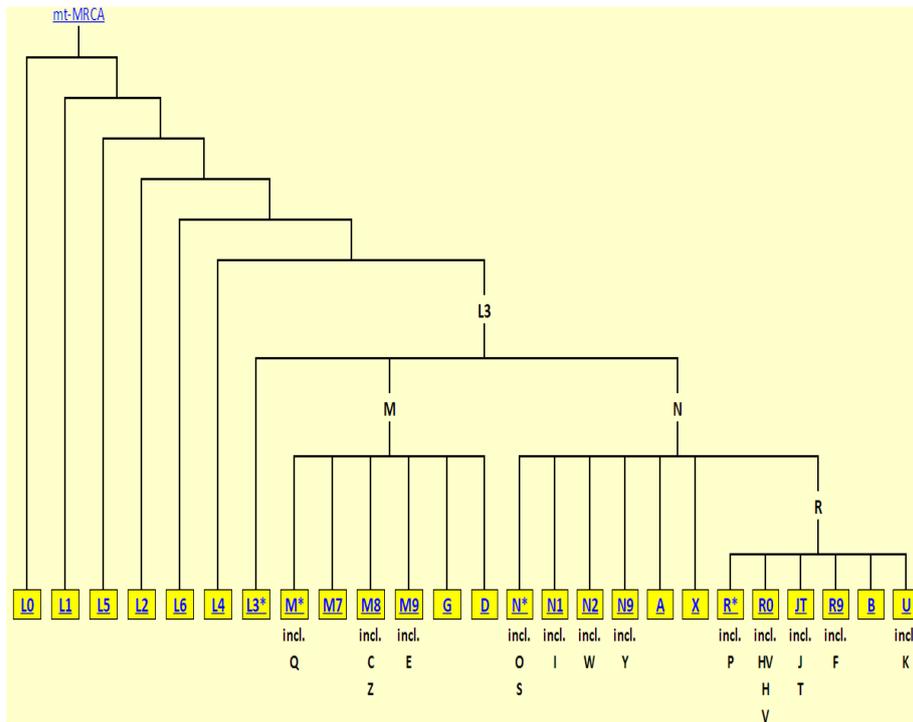
Porém, em casos de inclusão, há a possibilidade de dois indivíduos não relacionados apresentarem haplótipos iguais e, portanto, necessita-se estimar a frequência dos haplótipos na população em questão. Para isso, há a necessidade de se realizar um levantamento dos haplótipos de uma população e depositá-los em um banco de dados, de forma que os laboratórios possam compartilhar e utilizar as frequências e divergências genéticas para aumentar a confiabilidade de seus resultados (PRIETO et al., 2011).

2.5 Árvore Filogenética do DNA Mitocondrial

O DNA mitocondrial primordial ou *Matrilíneal Most Recent Common Ancestor* (mt-MRCA), que deu origem a todos os haplótipos da Terra, surgiu na África há cerca de 150 a 200 mil anos atrás. Embora a primeira ancestral com esse genoma não fosse a única mulher da época, seu mtDNA foi o mais bem sucedido e o único que sobreviveu, dando origem aos haplótipos mitocondriais de todas as pessoas do mundo atual. O mtDNA é transferido de mãe para filhos de forma quase inalterada, porém, devido à sua natureza suscetível a mutações, ocorrem alterações inofensivas no mtDNA ao longo do tempo, levando ao surgimento de novos polimorfismos genéticos nas populações (CANN; STONEKING; WILSON, 1987).

Dessa forma, por meio de várias mutações ao longo dos anos, o mt-MRCA deu origem a novos mtDNAs, que por sua vez também deram origem a novos haplótipos. Os haplótipos são organizados em haplogrupos de acordo com sua origem ancestral e similaridade (figura 6). Os haplogrupos possuem nomenclatura alfanumérica, como os haplogrupos A2, B, C1 e D, por exemplo (VAN OVEN; KAYSER, 2008).

Figura 6 – Árvore filogenética mitocondrial apresentando os haplogrupos encontrados nas populações mundiais.



Fonte: Van Oven (2016).

A migração dos povos africanos, em paralelo com a evolução do mtDNA e a sua estabilização em novos territórios resultaram em uma distribuição geográfica dos haplogrupos que correspondem aos das populações humanas (figura 7) (BARBOSA, 2008).

são os indígenas ou americanos nativos, os europeus, os africanos e, em menor proporção, os asiáticos (ANTUNES, [entre 1998 e 2008]). Cada uma dessas ancestralidades possuem haplogrupos característicos e, portanto, permitem relacionar a população estudada com a população de origem (BUDOWLE et al., 1999; CARRACEDO et al., 2000; PARSON et al., 2014).

Durante o período colonial (século XVI), a miscigenação da população ocorreu de forma intensa, com o relacionamento, muitas vezes à força, dos colonos portugueses com as escravas africanas e indígenas, além do relacionamento entre os próprios índios e pretos. Durante esse período, o número de indígenas variava de 2 a 10 milhões, e cerca de 6 milhões de africanos foram trazidos ao Brasil para a escravidão (ANTUNES, [entre 1998 e 2008]).

Os imigrantes portugueses formaram o grupo mais numeroso, com cerca de 1.785.000 indivíduos, uma vez que, de 1500 a 1808, apenas estes podiam entrar livremente no Brasil, além de não sofrerem as restrições aplicadas aos outros imigrantes após a oficialização da imigração em 1808. Ocuparam principalmente os Estados de São Paulo e Rio de Janeiro (ANTUNES, [entre 1998 e 2008]).

Em 1819, os suíço-alemães foram os primeiros imigrantes a chegarem ao Brasil, com a abertura dos portos, sendo os fundadores da cidade de Nova Friburgo, no Rio de Janeiro. Os alemães chegaram logo depois, em 1824, habitando boa parte do Rio Grande do Sul, e nas proximidades de São Paulo, Espírito Santo e Rio de Janeiro (ANTUNES, [entre 1998 e 2008]).

A imigração espanhola é muito antiga, tendo um movimento desde o período colonial até o ano de 1970, fixando-se, predominantemente, nas áreas urbanas do Sul e Sudeste, principalmente em São Paulo (ANTUNES, [entre 1998 e 2008]).

Os imigrantes italianos formam o segundo maior grupo depois dos portugueses. Tiveram sua maior movimentação no período de 1887 a 1914, localizando-se na parte centro-norte do Rio Grande do Sul, em Santa Catarina, Espírito Santo e com predominância em São Paulo (ANTUNES, [entre 1998 e 2008]).

Os eslavos, oriundos da Pôlonia, Rússia Branca e Ucrânia, chegaram ao Brasil a partir de 1875, e fixaram-se no Paraná e no Rio Grande do Sul. Os sírios, libaneses, árabes, palestinos, armênios e turcos migraram em menor quantidade e se estabeleceram na Amazônia e em diversas outras cidades, no período de 1860 a 1890 (ANTUNES, [entre 1998 e 2008]).

2.7 Bancos de dados de DNA para finalidade forense

Desde a criação do primeiro banco de dados na Inglaterra, em 1990, pode-se observar um aumento dos números de casos forenses que puderam ser solucionadas com a utilização das informações do banco. De 2001 a 2014, o banco de dados do Reino Unido possibilitou a identificação de mais de 471 mil criminosos, principalmente estupradores e assassinos em série (GODINHO, 2014).

Em 2012, a presidenta Dilma Roussef sancionou a lei 12.654, que determina o cadastro nos bancos de dados dos perfis genético de criminosos que praticarem crimes hediondos ou crimes violentos contra a pessoa. Tal medida auxiliaria, principalmente, a identificação de criminosos reincidentes. Dessa forma criou-se no Brasil a Rede Integrada de Bancos de Perfis Genéticos (RIBPG). Esta rede é baseada no CODIS (*Combined DNA Index System*), um *software* criado pelo FBI, que une dados forenses com a informática, possibilitando a associação de um perfil genético armazenado nele com o de uma amostra retirada da cena do crime. O CODIS e a RIBPG são, portanto, redes integradas, onde os dados de perfis genéticos são submetidos e compartilhados por vários laboratórios forenses participantes (GODINHO, 2014).

Embora a Rede Integrada de Bancos de Perfis Genéticos seja útil em compartilhar e comparar os perfis genéticos obtidos das cenas de crimes em nível nacional, tal sistema não armazena sequências de mtDNA.

Diante da importância previamente discutida sobre a utilização do DNA mitocondrial como auxílio nas análises de amostras com pouco ou nenhum DNA nuclear, faz-se necessário o desenvolvimento de bancos de dados brasileiros de mtDNA para auxiliarem nas investigações em que tais amostras são utilizadas.

Diferente do RIBPG e do CODIS, alguns bancos de dados de mtDNA, que serão descritos adiante, possuem a função de armazenar os haplótipos de uma dada população, e não de criminosos. Dessa forma, estes bancos permitem estimar a frequência dos haplótipos e a probabilidade de combinação genética entre indivíduos não relacionados. Isto auxilia na discriminação entre indivíduos sem vínculo materno, na identificação humana e em outras áreas abordadas pela antropologia.

A frequência haplotípica é obtida pelo método de contagem, através da fórmula $p=x/N$, no qual “p” é a frequência, “x” é o número de vezes que um

determinado haplótipo foi contado no banco de dados e “N” é o total de haplótipos contidos no banco (SCIENTIFIC WORKING GROUP ON DNA ANALYSIS METHODS, 2013).

Os laboratórios de análise forense geralmente utilizam um conjunto de dados da sua população local. Porém, há um uso limitado desses dados, que muitas vezes não estão à disposição de toda a comunidade forense. Além disso, esses dados podem conter erros ou ambiguidades, o que aponta a necessidade de coletar, revisar e organizar esse conjunto de dados, de forma a minimizar os erros, aumentar a acessibilidade e padronizar os procedimentos de análises dos haplótipos do mtDNA (PRIETO et al., 2011).

2.8 Banco de dados do EMPOP

Em 1999, na conferência da *International Society of Forensic Haemogenetics* (ISFH) em San Francisco, o grupo *European DNA Profiling* (EDNAP) se comprometeu em apoiar o desenvolvimento de um banco de dados online de mtDNA contendo dados de alta qualidade de várias populações (PARSON et al., 2004). Assim, foi criado o EMPOP (*EDNAP Mitochondrial DNA Population Database*), que é um projeto colaborativo entre vários laboratórios forenses ao redor do mundo, com o objetivo de aumentar a quantidade de dados confiáveis de mtDNA populacional e que possam ser disponibilizados via internet (www.empop.org) (PARSON et al., 2004; PRIETO et al., 2011).

Os haplótipos armazenados no EMPOP são submetidos a revisões minuciosas, para testar sua confiabilidade. Além do EMPOP, existem outros bancos de dados e ferramentas de busca de mtDNA disponíveis na internet, como o MITOMAP (Universidade da Califórnia – EUA), AMBase (Universidade Federal do Pará – Brasil), *FBI Software mtDNA Population database* (FBI – EUA), dentre outros (PANETO, 2010).

Conforme informado por dois peritos criminais brasileiros, Jorge Freitas¹ e Marcelo Malaghini² (comunicação pessoal), os laboratórios forenses no Brasil utilizam os dados do EMPOP e na ausência de informações, alguns laboratórios possuem amostras compiladas em bancos de mtDNA próprios.

O EMPOP é uma ferramenta interessante para avaliar dados de mtDNA populacional, mas ainda necessita de maior número de amostras, principalmente de regiões que, atualmente, são pouco representadas, como o Brasil e o restante da América do Sul. Portanto, mais estudos são necessários para avaliar de forma precisa a frequência e a distribuição das linhagens de mtDNA nas populações desses países (PRIETO et al., 2011).

¹ FREITAS, J. M. de. Utilização do DNA mitocondrial e de bancos de dados nas investigações criminais pelos laboratórios forenses do Brasil. Araraquara, São Paulo, Laboratório de Paternidade da UNESP, 12 nov. 2015. Entrevista realizada pela autora.

² MALAGHINI, Marcelo. Utilização do DNA mitocondrial e de bancos de dados nas investigações criminais pelos laboratórios forenses do Brasil. Araraquara, São Paulo, Faculdade de Ciências Farmacêuticas da UNESP, 17 set. 2015. Entrevista realizada pela autora.

3 OBJETIVOS

3.1 Objetivo geral

Analisar de forma crítica os dados disponíveis sobre os haplogrupos do DNA mitocondrial no Brasil em comparação com outros países.

3.2 Objetivos específicos

- Comparar a deposição de dados brasileiros no EMPOP com outros países;
- Avaliar a influência do tamanho amostral na diversidade genética e no número de haplótipos;
- Analisar a distribuição de haplogrupos e de origem ancestral pelas regiões brasileiras;
- Estimar as distâncias genéticas entre as populações brasileiras e com as populações que deram as suas origens;
- Avaliar as diversidades genéticas e os números de haplótipos, para as regiões hipervariáveis quando analisadas separadamente e em conjunto;
- Sequenciar as três regiões hipervariáveis do DNA mitocondrial de 5 amostras referências fornecidas pela Promega e GHEP.

4 METODOLOGIA

4.1 Análise das quantidades de dados no EMPOP

As quantidades de dados depositados no EMPOP de cada país participante, inclusive do Brasil, foram verificadas no site do EMPOP. Estes dados foram organizados em planilhas de Excel e separados por continentes e subcontinentes. Realizou-se uma comparação das quantidades de sequências entre todos os países, destacando aqueles com maior número de sequências depositadas.

A contribuição de dados brasileiros foi comparada ao de países da América do Sul, ao Continente americano e demais países que depositam seus dados no EMPOP.

4.2 Tamanho amostral em relação à diversidade genética e de haplótipos

Os haplótipos para a região controle do mtDNA de 492 amostras de São Paulo (CARDENA et al., 2013), foram utilizados para observar a influência do tamanho amostral nos cálculos de diversidade genética e de total de haplótipos.

A população de Cardena foi escolhida devido ao seu tamanho, possibilitando que as amostras sejam divididas em vários grupos a partir de 50 indivíduos, crescendo nessa ordem (de 50 em 50 indivíduos) até o valor final de 492. Para cada grupo, foram calculados as diversidades genéticas e o total de haplótipos através do *software* do Arlequin v.3.5.2.2, para HV1 e HV2 separados e em conjunto, assim como de toda região controle. Os resultados foram plotados em um gráfico do Excel 2010.

4.3 Análise de dados brasileiros disponíveis na literatura

Artigos científicos e publicações referentes a várias populações do Brasil foram analisadas. Os haplótipos e haplogrupos foram anotados e organizados por regiões brasileiras (Sudeste, Sul, Norte e Nordeste). Não se encontraram dados referentes à região Centro-Oeste. Estes trabalhos e os demais (nacionais e internacionais) apontados na tabela 1 foram utilizados para comparar essas populações.

Por meio dos dados obtidos, estimou-se a distribuição e a predominância dos haplogrupos pelos territórios brasileiros, assim como as suas origens ancestrais. Também foi possível observar quais regiões estão mais carentes de pesquisas nessa área.

Tabela 1 – Relação dos trabalhos brasileiros utilizados neste estudo

Referências	Regiões geográficas	População estudada	Regiões do mtDNA analisadas
ALVES-SILVA et al., 2000	Sudeste	Minas Gerais	HV1 e HV2
	Sul	Rio Grande do Sul, Santa Catarina e Paraná	HV1
	Nordeste	Pernambuco	HV1
	Norte	Amazonas, Pará, Rondônia e Acre	HV1
AMBROSIO et al., 2015	Sudeste	Espirito Santo	Região controle total
BERNARDO et al., 2014	Sudeste	Rio de Janeiro	Região controle total
CARDENA et al., 2013	Sudeste	São Paulo	Região controle total
FRIDMAN et al., 2014	Sudeste	São Paulo	Região controle total
PANETO, 2010	Sudeste	São Paulo	Região controle total
PRIETO et al., 2011	Sudeste	São Paulo (1)	Região controle total
PRIETO et al., 2011	Sudeste	São Paulo (2)	HV1
PRIETO et al., 2011		Norte de Portugal	HV1
SANCHES, 2013	Sudeste	Espirito Santo	Região controle total
MARRERO et al., 2005	Sul	Veranópolis - RS	HV1
	Sul	General - RS	HV1
PALENCIA et al., 2010	Sul	Santa Catarina	Região controle total
GUERREIRO-JUNIOR et al., 2009 ¹	Sul	Porto Alegre - RS	HV1
BARBOSA et al., 2008	Nordeste	Alagoas	HV1 e HV2
SANTOS et al., 2006	Norte	Pará	HV1
SANTOS et al., 2007	Norte	Região Amazônica, população preta	HV1
GONÇALVES et al., 2007	Sudeste	São Paulo, população homens pretos	HV1
MARRERO et al., 2007	Centro - oeste e sul	Kaingang	HV1
	Centro - oeste e sul	Guarani	HV1
MIKKELSEN et al., 2012	África	Somália	HV1

Fonte: da autora.

4.4 Cálculos estatísticos

As diversidades genéticas de cada população e a distância genética F_{st} entre elas foram calculadas pelo *software* Arlequin v.3.5.2.2. As diversidades genéticas foram obtidas de HV1, HV2 e HV3, separadamente, HV1 e HV2 juntas e das três regiões hipervariáveis juntas, de acordo com cada região analisada nos estudos.

As distâncias genéticas (F_{st}) foram calculadas unicamente para HV1, porque foi a única região em comum utilizada por todos os estudos analisados. As populações brasileiras foram comparadas entre si e com as populações de Portugal, África e de duas tribos indígenas (Guarani e Kaingang). As distâncias genéticas foram plotadas em gráficos pelo *software* “Statistica” v. 13.2.

4.5 Sequenciamentos de amostras referência (Promega e GHEP-ISFG)

Com a finalidade de me proporcionar experiência na técnica e comprovar que uma metodologia padronizada permite obter resultados satisfatórios, foi realizado o sequenciamento de amostras previamente conhecidas, seguindo a metodologia de análise de mtDNA utilizada no Laboratório de Investigação de Paternidade do Núcleo de Atendimento à Comunidade da Faculdade de Ciências Farmacêuticas da UNESP de Araraquara – SP.

4.5.1 AMOSTRAS

Cinco amostras de DNA foram sequenciadas e analisadas, sendo três (M1, M2 e M3) do controle de qualidade do GHEP-ISFG (Grupo de Línguas Espanhola e Portuguesa da Sociedade Internacional de Genética Forense), duas (2800 e 9948) da Promega e o controle negativo de água ultrapura.

As análises foram realizadas no Laboratório de Investigação de Paternidade do Núcleo de Atendimento à Comunidade da Faculdade de Ciências Farmacêuticas da UNESP de Araraquara – SP.

4.5.2 EXTRAÇÃO DE DNA

As amostras fornecidas pela Promega já vieram prontas e não precisaram de extração. O mtDNA da amostra M1 do GHEP foi extraído com o kit do DNA IQ™ System-Small Casework (Promega), seguindo o protocolo do fabricante. As amostras M2 e M3 do GHEP tiveram extração com Chelex 5%, de acordo com o seguinte protocolo:

- Em um tubo de 0,2 ml, adicionou-se um disco do papel FTA contendo a amostra;
- Adicionou-se 50 µl de água ultrapura e agitou-se o tubo no vórtex;
- A água foi descartada e adicionou-se, novamente, 50 µl de água ultrapura;
- A solução foi incubada em temperatura ambiente por 30 minutos;
- O tubo foi centrifugado a 14000 rpm por 1 minuto na centrífuga Eppendorf Centrifugal 5417R;
- A água foi descartada e 100 µl de solução de Chelex 5% foram adicionados;
- A solução foi incubada a 56°C por 1 hora no termociclador *Gene Amp PCR System 9700* (Perkin Elmer);
- Após ser vortexada, a amostra foi incubada por 8 minutos a 100°C no termociclador *Gene Amp PCR System 9700* (Perkin Elmer);
- O tubo foi centrifugado a 14000 rpm por 3 minutos na centrífuga Eppendorf Centrifugal 5417R, e o sobrenadante foi armazenado para a reação em cadeia da polimerase (PCR).

4.5.3 REAÇÃO EM CADEIA DA POLIMERASE (PCR)

O *mix* foi preparado em tubos de 0,2 ml, seguindo o seguinte protocolo:

- 10X *Gold Star Buffer* (Promega);
- 1U de *AmpliTaqGold DNA Polymerase* (Applied Biosystems by Life Technologies);
- 0,2 µM de cada iniciador (tabela 2);
- 1 a 5 ng de DNA;
- Água ultrapura (*Milli-Q*) qsp. 25µL.

Tabela 2 – Iniciadores utilizados na PCR e no sequenciamento

	Iniciador	Orientação	Sequência
HV1	L15997	Forward	5'-CACCATTAGCACCCAAAGCT-3'
	H16401	Reverse	5'-TGATTTACACGGAGGATGGTG-3'
HV2	L29	Forward	5'-GGTCTATCACCCCTATTAACCAC-3'
	H408	Reverse	5'-CTGTTAAAAGTGCATACCGCCA-3'
HV3	F314	Forward	5'-CCGCTTCTGGCCACAGCACT-3'
	639R	Reverse	5'-GGGTGATGTGAGCCCGTCTA-3'

Adaptado de: Mitotyping Technology Protocol *apud* Paneto (2010) e Sanches (2013).

As regiões HV1, HV2 e HV3 foram amplificadas separadamente no termociclador *Gene Amp PCR System 9700* (Perkin Elmer) com a seguinte ciclagem:

95 °C por 11 minutos

96 °C por 1 minuto

94 °C por 30 segundos

56 °C por 30 segundos

72 °C por 90 segundos

60 °C por 11 minutos

4 °C ∞

} 40 ciclos

4.5.4 PURIFICAÇÃO DOS PRODUTOS DA PCR

Após a PCR, as amostras foram purificadas utilizando-se *GFX™ PCR DNA and Gel Band Purification Kit* (GE Healthcare), com a finalidade de retirar os nucleotídeos não incorporados e excesso de iniciadores.

4.5.5 SEQUENCIAMENTO DO DNA

Para o sequenciamento, foram preparadas as soluções em tubos de 0,2 ml, de acordo com o seguinte protocolo:

- 0,25X de *Big Dye Terminator v. 3.1* (Applied Biosystems by Life Technologies);
- 3,2 picomolar de iniciador direto ou reverso;

- 1X de tampão do kit *Big Dye Terminator* v. 3.1 (Applied Biosystems by Life Technologies);
- 2,0 µL do produto de PCR de HV1 e 3 µL para os produtos de HV2 e HV3.
- Água ultrapura (Mili-Q) qsp. 10 µL;

O sequenciamento foi realizado no termociclador *GeneAmp PCR System 9700* (Perkin Elmer) conforme a ciclagem:

96 °C por 1 minuto	}	25 ciclos
96 °C por 10 segundos		
56 °C por 10 segundos		
60 °C por 4 minutos		
4 °C ∞		

4.5.6 PURIFICAÇÃO E PRECIPITAÇÃO

Com a finalidade de se retirar os nucleotídeos não incorporados, os produtos do sequenciamento foram purificados pela técnica de precipitação por etanol, de acordo com o seguinte protocolo:

- Em um tubo de 1,5 ml contendo 1 µL do produto de PCR, foram adicionados 80 µL de isopropanol 75%;
- Após homogeneizar, a mistura foi mantida por 15 minutos no escuro em temperatura ambiente;
- Centrifugou-se a solução por 15 minutos a 14.000 rpm na centrífuga Eppendorf Centrifugal 5417R;
- O sobrenadante foi cuidadosamente descartado sobre papel absorvente;
- Adicionou-se ao precipitado 1 ml de etanol 70%, centrifugando (Eppendorf Centrifugal 5417R) por 5 minutos em temperatura ambiente;
- O sobrenadante foi cuidadosamente descartado e o precipitado foi submetido à secagem a vácuo (DNA Speed Vac – Savant) por 30 minutos;
- Os precipitados secos foram ressuspensos em 10 µL de Formamida HI-DI (Applied Biosystems by Life Technologies).

4.5.7 ELETROFORESE CAPILAR

Os produtos dos sequenciamentos purificados foram transferidos para uma placa. Agitou-se a placa em agitador de tubos *FINEMIXER SH2000*, durante 1 minuto, e centrifugou-se a 14000 rpm por 30 segundos no Eppendorf Centrifugal 5417R. As amostras foram submetidas à temperatura de 95 °C no termociclador *Gene Amp PCR System 9700* (Perkin Elmer) por 3 minutos para a desnaturação do DNA, e logo em seguida, a placa foi depositada sob o gelo por 3 minutos. As amostras foram aplicadas no analisador genético *ABI 3500* (Applied Biosystems by Life Technologies).

Para realizar a eletroforese no *ABI 3500*, utilizou-se o polímero POP-7 (Applied Biosystems by Life Technologies) e capilar de 50 cm (Applied Biosystems by Life Technologies).

4.5.8 ANÁLISE DOS POLIMORFISMOS

As sequências foram analisadas no programa SeqScap v. 3 para confirmação da qualidade dos eletroferogramas.

4.5.9 CLASSIFICAÇÃO EM HAPLOGRUPOS

As sequências foram analisadas com as ferramentas do mtDNAManager (YONSEI DNA PROFILING GROUP, 2007) e classificados nos seus respectivos haplogrupos.

5 RESULTADOS E DISCUSSÕES

5.1 Quantidades de dados no EMPOP

O EMPOP possui 34617 haplótipos depositados até a data da avaliação 20 de setembro de 2015, e um total de 63 países participantes. Não há dados da Oceania.

A pesquisa iniciou-se pela América do Sul, onde foi possível observar maior número de dados da Argentina, seguido pelo Chile. O Brasil é o terceiro país com maior número de haplótipos depositados, sendo responsável por cerca de 25% dos dados sul-americanos (tabela 3).

Tabela 3- Número de haplótipos depositados no EMPOP por países da América do Sul

Países	Número de haplótipos	Porcentagem
Argentina	1088	34,3%
Chile	996	31,4%
Brasil	782	24,6%
Equador	107	3,4%
Venezuela	101	3,2%
Colômbia	98	3,1%
Total	3074	100%

Fonte: European DNA Profiling Group (2015).

De doze países e três territórios ultramarinos, apenas seis países da América do Sul (tabela 3) participam do EMPOP.

Os países da América Central apresentaram os menores índices de dados depositados de todo o continente Americano (tabela 4). De vinte países e dezesseis territórios que compõem a América Central, apenas sete deles estão representados no EMPOP até a data da avaliação.

Tabela 4 – Número de haplótipos depositados no EMPOP por países da América Central

Países	Número de haplótipos	Porcentagem
Guatemala	221	17%
Porto Rico	209	16%
Jamaica	199	16%
Honduras	184	15%
Haiti	163	13%
Nicarágua	162	13%
República Dominicana	130	10%
Total	1268	100%

Fonte: European DNA Profiling Group (2015).

A América do Norte, embora possua apenas dois países contribuintes, é o continente com o maior número de haplótipos depositados no EMPOP, devido principalmente ao fornecimento de dados dos Estados Unidos da América (EUA), que superou todos os outros países americanos (tabela 5).

Tabela 5 - Número de haplótipos depositados por países da América do Norte no EMPOP

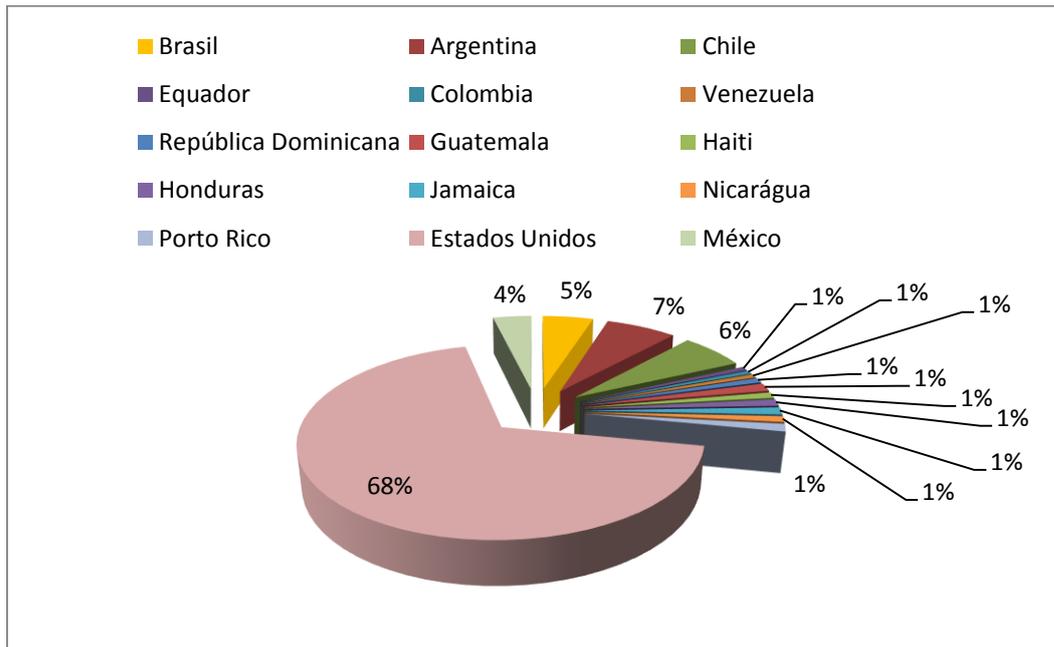
Países	Número de haplótipos	Porcentagem relativa
Estados Unidos	10799	95%
México	593	5%
Total	11392	100%

Fonte: European DNA Profiling Group (2015).

O Canadá não contribuiu até esse momento com dados para o EMPOP, talvez porque utilize outros bancos de dados populacionais.

O gráfico 1 permite correlacionar as participações dos países americanos, sendo os EUA o maior contribuidor, seguido da Argentina, Chile e Brasil, respectivamente.

Gráfico 1 – Porcentagem de haplótipos depositados no EMPOP por países Americanos



Fonte: Autora.

A pequena participação do Brasil no EMPOP possivelmente deve-se ao uso do mtDNA ser relativamente recente no país. Também se deve, segundo o perito criminal Marcelo Malaghini² (comunicação pessoal), ao alto custo dos equipamentos e reagentes utilizados nas análises e maior complexidade do método quando comparado com a análise do DNA nuclear. Além disso, segundo Walther Parson³ (comunicação pessoal) responsável pelo banco de dado do EMPOP, muitos pesquisadores brasileiros não disponibilizam os *raw data* (eletroferogramas) para assegurar a qualidade e confirmar os seus resultados, o que não permite que tais sequências sejam depositadas no EMPOP.

A Europa possui muitos países participantes, sendo a Alemanha e a Espanha os que possuem maiores números de sequências depositados (tabela 6).

² MALAGHINI, Marcelo. Utilização do DNA mitocondrial e de bancos de dados nas investigações criminais pelos laboratórios forenses do Brasil. Araraquara, São Paulo, Faculdade de Ciências Farmacêuticas da UNESP, 17 set. 2015. Entrevista realizada pela autora.

³ PARSON, W. Analysis and interpretation of mtDNA in the forensic field. Rio de Janeiro, Universidade do Estado do Rio de Janeiro. Workshop realizado 10 out. 2016.

Tabela 6 - Número de haplótipos depositados por países da Europa no EMPOP

Países	Número de amostras	Porcentagem
Espanha	1567	16%
Alemanha	1177	12%
Polônia	894	9%
Bulgária	855	9%
Portugal	733	7%
Suécia	634	6%
Hungria	594	6%
Federação Russa	456	5%
Itália	398	4%
Romênia	360	4%
Grécia	319	3%
Áustria	273	3%
Eslovênia	228	2%
Dinamarca	206	2%
Finlândia	200	2%
Macedônia	200	2%
Croácia	187	2%
Bósnia e Herzegovina	156	2%
Letônia	131	1%
Bélgica	104	1%
Holanda	104	1%
Chipre	91	1%
Total	9867	100%

Fonte: European DNA Profiling Group (2015).

Os EUA superam o continente europeu em número de haplótipos depositados e, embora a Europa (750 milhões de habitantes) apresente uma população maior que a dos EUA (324,4 milhões de habitantes), apenas 22 países dos quase 50 que o

compõem depositaram dados no EMPOP. (COUNTRYMETER, 2016; FREITAS, 2016).

A Ásia foi o terceiro continente com maior número de haplótipos depositados (Gráfico 2), com maior participação do Uzbequistão e da China. De 45 países asiáticos, apenas dezesseis contribuem com dados para o EMPOP (tabela 7).

Tabela 7 – Número de haplótipos depositados no EMPOP por países da Ásia

Países	Número de haplótipos	Porcentagem relativa
Uzbequistão	1573	23,1%
China	927	13,6%
Coréia do Sul	692	10,1%
Paquistão	644	9,4%
Japão	564	8,3%
Kuwait	381	5,6%
Indonésia	277	4,1%
Índia	250	3,7%
Emirados Arábes Unidos	249	3,6%
Laos Republica Democrática	214	3,1%
Barein	213	3,1%
Tailândia	190	2,8%
Vietnã	187	2,7%
Iraque	182	2,7%
Filipinas	153	2,2%
Malásia	123	1,8%
Total	6819	100%

Fonte: European DNA Profiling Group (2015).

A África apresentou uma maior contribuição de dados provenientes do Marrocos e do Egito (Tabela 8). De 55 países que compõem esse continente, apenas dez contribuíram com dados para o EMPOP até o momento deste levantamento (20 de setembro de 2015).

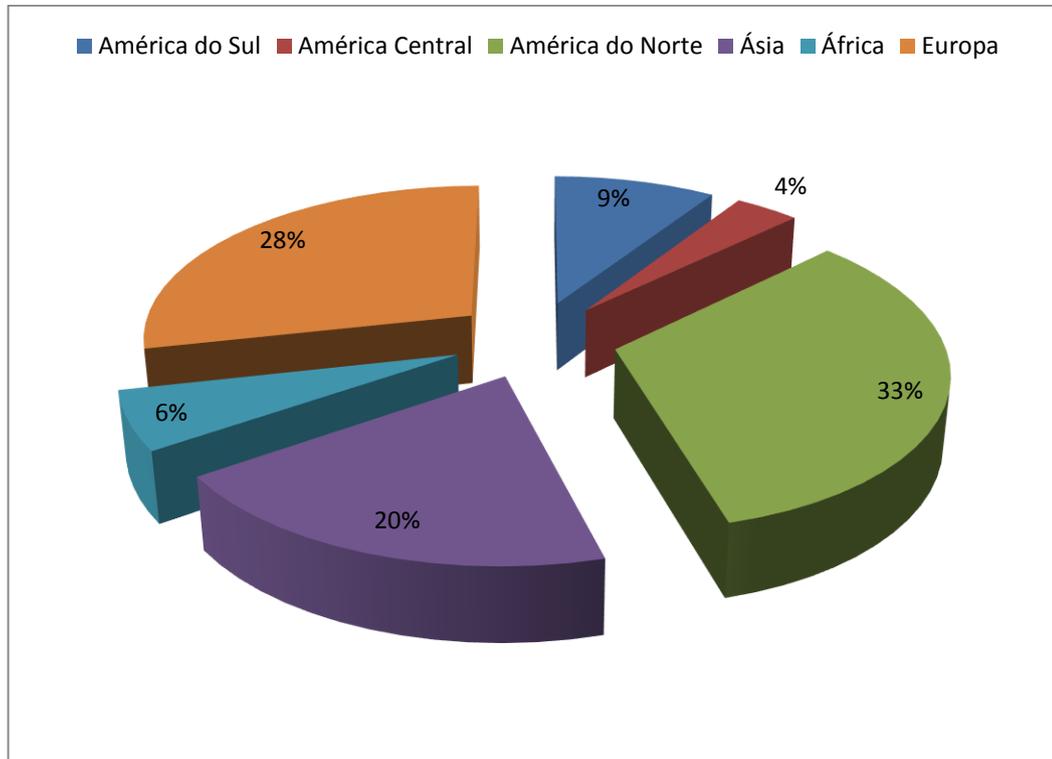
Tabela 8 – Número de haplótipos depositados no EMPOP por países da África

Países	Número de haplótipos	Porcentagem relativa
Marrocos	565	26,9%
Egito	405	19,3%
Uganda	261	12,4%
Gana	191	9,1%
Somália	190	9%
Angola	129	6,1%
Gabão	115	5,5%
Quênia	100	4,8%
Guiné-Bissau	79	3,8%
Tunísia	64	3%
Total	2099	100%

Fonte: European DNA Profiling Group (2015).

Com todos os dados apresentados anteriormente, foi possível comparar os continentes e subcontinentes entre si, obtendo-se o seguinte gráfico de participação no EMPOP:

Gráfico 2 – Comparação da porcentagem de haplótipos fornecidos ao EMPOP entre continentes e subcontinentes. América do Norte (verde) contou com 33% do total dos dados depositados no EMPOP, sendo 2 países depositantes (EUA e México); a Europa (laranja) contou com 22 países; a Ásia (roxo) com 16; a América do Sul (azul escuro) com 6; a África (azul claro) com 10; e a América Central (vermelho) com 7.



Fonte: Autora.

O continente americano foi o que apresentou maior contribuição de haplótipos pela inclusão dos EUA, seguido pelos continentes europeu e asiático, respectivamente. A tabela 9 mostra os países que tiveram a maior quantidade de dados depositados no EMPOP.

Tabela 9 – Países com maior número de dados depositados no EMPOP

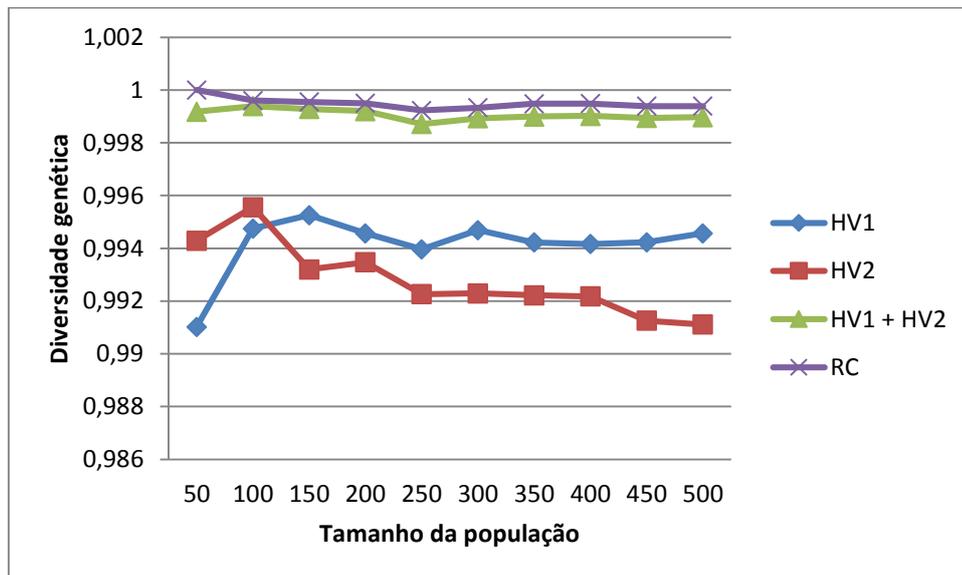
País	Continente	Total de Haplótipos
EUA	América do Norte	10799
Uzbequistão	Ásia	1573
Espanha	Europa	1567
Alemanha	Europa	1177
Argentina	América do Sul	1088

Fonte: European DNA Profiling Group (2015).

5.2 Tamanho amostral, diversidade genética e total de haplótipos

Este estudo utilizou-se dos dados de Cardena (2013) para avaliar a influência do tamanho amostral na diversidade genética. Esta população foi utilizada por ser a que contém a maior quantidade de haplótipos entre os estudos brasileiros. Foi observado que as diversidades genéticas, das regiões de menor poder de discriminação (HV1 e HV2), aumentam junto com o tamanho da amostra, atingem um pico e, logo em seguida, começam a diminuir, o que também foi observado por Pereira e colaboradores (2004). Em relação às diversidades genéticas das regiões mais polimórficas (região controle e HV1+HV2), estas apresentaram menor variabilidade em função do crescimento amostral (Gráfico 3).

Gráfico 3 – Diversidade genética e tamanho da população (492 indivíduos) conforme o haplótipo estudado – HV1, HV2, HV3 e região controle. As regiões menos polimórficas (HV1 e HV2) demonstraram maior variabilidade da diversidade genética em função do tamanho amostral do que as outras regiões mais polimórficas (HV1+HV2 e RC). RC – Região controle.



Fonte: Autora.

Por HV2 ser menos discriminatória, a queda na diversidade genética é mais acentuada. Isto se deve ao aumento do número de haplótipos compartilhados entre indivíduos, uma vez que HV2 possui pouco poder para discriminar estes haplótipos entre si. Portanto, o aumento do número de sequências iguais, na população, resultou na diminuição da sua diversidade genética.

Observando as diversidades genéticas descritas em outros estudos, em relação à região do mtDNA analisada, a região controle demonstrou maiores índices de diversidade genética, seguida por HV1-HV2, HV1, HV2 e HV3 (tabela 10).

Tabela 10 – Diversidade genética e tamanho da população estudada. (1+2) HV1 +HV2; (RC) Região Controle. SP – São Paulo; MG – Minas Gerais; ES – Espírito Santo; RJ – Rio de Janeiro; RS – Rio grande do Sul, SC – Santa Catarina; PR – Paraná; PA – Pará, AM – Amazonas; RO – Rondônia; AC – Acre, AL – Alagoas; PE – Pernambuco. ND – Não determinado.

Autor / População	Tamanho da amostra	HV1	HV2	HV3	1 + 2	RC
PRIETO et al., 2011 (SP)	142	0,9927	0,9907	0,78134	0,99860	0,99910
PRIETO et al., 2011 (SP)	48	0,9929	0,9823	ND	1	ND
PANETO, 2010 (SP)	160	0,9936	0,9912	0,79410	0,99866	0,99921
FRIDMAN et al., 2014(SP)	290	0,9932	0,99	0,82320	0,99831	0,99862
CARDENA et al., 2013 (SP)	492	0,9946	0,9911	0,77168	0,99897	0,99939
ALVES-SILVA et al., 2000 (MG)	99	0,9885	ND	ND	ND	ND
SANCHES, 2013 (ES)	100	0,9946	0,9901	0,76061	0,99798	0,99818
GONÇALVES et al., 2008 (SP – pretos)	120	0,9913	ND	ND	ND	ND
BERNARDO et al., 2014 (RJ)	179	0,9971	0,9892	ND	0,99956	ND
MARRERO et al., 2005 (General – RS)	88	0,9237	ND	ND	ND	ND
MARRERO et al., 2005 (Veranópolis – RS)	31	0,9936	ND	ND	ND	ND
PALENCIA et al.,	80	0,9687	0,9775	0,76076	0,99209	0,99430

2010 (SC)						
GUERREIRO-JUNIOR et al., 2009 (RS)	171	0,9774	ND	ND	ND	ND
ALVES-SILVA et al., 2000 (RS, SC, PR,)	50	0,9804	ND	ND	ND	ND
SANTOS et al., 2006 (PA)	60	0,9853	ND	ND	ND	ND
ALVES-SILVA et al., 2000 (AM, PA, RO, AC)	49	0,9847	ND	ND	ND	ND
SANTOS et al., 2007 (AM – Pretos)	235	0,9895	ND	ND	ND	ND
BARBOSA et al., 2008 (AL)	167	0,9934	0,9907	ND	0,99755	ND
ALVES-SILVA et al., 2000 (PE)	50	0,9943	ND	ND	ND	ND
PRIETO et al., 2011 (Portugal)	132	0,9762	ND	ND	ND	ND
MIKKELSEN et al., 2012 (África)	190	0,9865	ND	ND	ND	ND
MARRERO et al., 2006(Guarani)	200	0,7635	ND	ND	ND	ND
MARRERO et al., 2006 (Kaingang)	78	0,7729	ND	ND	ND	ND

Fonte: Autora.

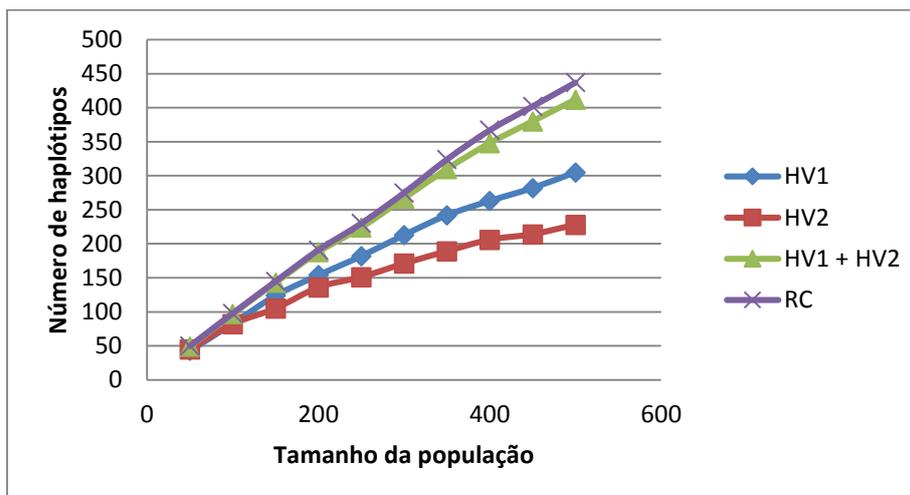
Os valores das diversidades genéticas nos estudos apresentados na tabela 10 apontam altos índices de variabilidade genética nas populações brasileiras estudadas.

A região Sudeste foi a que apresentou os maiores valores de diversidade genética para suas populações. Estes resultados podem ser explicados pelo forte fluxo de imigrantes (suíço-alemães, espanhóis, japoneses, portugueses, africanos e italianos) e pela migração de grupos do Norte e Nordeste, principalmente para essa região (ANTUNES, [entre 1998 e 2008]).

Nos estudos que disponibilizaram haplótipos da região controle, foi possível observar a influencia de cada região (HV1, HV2, HV3, HV1-HV2 e RC) na diversidade genética. As áreas mais polimórficas apresentaram os maiores índices de diversidade.

Estes resultados corroboram com o que já foi relatado em outros estudos. No trabalho de Fridman e colaboradores (2008), a região HV1 apresentou mais sítios polimórficos em relação a HV2, tendo assim um poder de discriminação maior. HV1-HV2 aumenta consideravelmente esse poder de discriminação. Lutz e colaboradores (2000) demonstraram que HV3, embora possua menor poder de discriminação quando comparado com HV1 e HV2, é muito útil quando analisada em conjunto com estas regiões, permitindo diferenciar sequências de HV1 e HV2 idênticas. Por englobar as três regiões hipervariáveis e mais alguns polimorfismos, a região controle é altamente polimórfica. Portanto, atualmente, as recomendações internacionais para análise de mtDNA sugerem o sequenciamento da região controle, principalmente quando houver o intuito de depositar os haplótipos em bancos de dados (PARSON et al., 2014).

Gráfico 4 – Número de haplótipos e tamanho da população (492 indivíduos) conforme o haplótipo estudado – HV1, HV2, HV3 e região controle. As regiões mais polimórficas (HV1+HV2 e RC) discriminaram um maior número de haplótipos em cada grupo populacional, e dessa forma apresentaram maior proximidade entre os valores do número de haplótipos e do tamanho da população, do que as outras regiões menos polimórficas (HV1 e HV2). RC – Região controle.



Fonte: Autora.

No gráfico 4, observa-se que o número de haplótipos aumentou conforme foram acrescentadas mais amostras. Quanto mais polimórfica a região (hipervariável e controle) mais próximos são os valores do tamanho amostral com o total de haplótipos. O mesmo pôde ser observado para outras análises como apresentado na tabela 10.

Assim como foi observado para a diversidade genética, quanto mais polimórfica era a sequência em análise, maior foi o número de haplótipos discriminados entre si. Por exemplo, no gráfico 4, quando a população é de 400 indivíduos, HV2 consegue determinar cerca de 200 haplótipos, enquanto que a região controle permite discriminar, para o mesmo tamanho amostral, mais de 350 haplótipos.

Tabela 11 – Total de haplótipos em função da região analisada e do tamanho da população estudada. (1) HV1; (2) HV2; (1+2) HV1-HV2; (3) HV3; (RC) Região Controle. SP – São Paulo; MG – Minas Gerais; ES – Espírito Santo; RJ – Rio de Janeiro; RS – Rio grande do Sul, SC – Santa Catarina; PR – Paraná; PA – Pará, AM – Amazonas; RO – Rondônia; AC – Acre, AL – Alagoas; PE – Pernambuco. ND – Não determinado.

Autor / População	Tamanho da amostra	Número de haplótipos				
		1	2	3	1 + 2	RC
PRIETO et al., 2011 (SP)	142	110	94	26	130	134
PRIETO et al., 2011 (SP)	48	43	36	ND	48	ND
PANETO, 2010 (SP)	160	123	103	29	145	151
FRIDMAN et al., 2014(SP)	290	197	156	47	249	253
CARDENA et al., 2013 (SP)	492	305	228	63	412	437
ALVES-SILVA et al., 2000 (MG)	99	80	ND	ND	ND	ND
SANCHES, 2013 (ES)	100	81	70	21	92	93
GONÇALVES et al., 2008 (SP – pretos)	120	85	ND	ND	ND	ND
BERNARDO et al., 2014 (RJ)	179	147	108	ND	172	ND
MARRERO et al., 2005 (General – RS)	88	43	ND	ND	ND	ND
MARRERO et al., 2005 (Veranópolis – RS)	31	28	ND	ND	ND	ND
PALENCIA et al., 2010 (SC)	80	52	44	23	61	66
GUERREIRO-JUNIOR et al., 2009 (RS)	171	105	ND	ND	ND	ND
ALVES-SILVA et al., 2000 (RS, SC, PR,)	50	41	ND	ND	ND	ND

SANTOS et al., 2006 (PA)	60	49	ND	ND	ND	ND
ALVES-SILVA et al., 2000 (AM, PA, RO, AC)	49	40	ND	ND	ND	ND
SANTOS et al., 2007 (AM – Pretos)	235	129	ND	ND	ND	ND
BARBOSA et al., 2008 (AL)	167	145	102	ND	145	ND
ALVES-SILVA et al., 2000 (PE)	50	44	ND	ND	ND	ND
PRIETO et al., 2011 (Portugal)	132	89	ND	ND	ND	ND
MIKKELSEN et al., 2012 (África)	190	95	ND	ND	ND	ND
MARRERO et al., 2006(Guarani)	200	9	ND	ND	ND	ND
MARRERO et al., 2006 (Kaingang)	78	12	ND	ND	ND	ND
Total	3211	2053	941	209	1454	1134

Fonte: Autora.

As populações nativo-americanas, Guarani e Kaingang (MARRERO et al., 2006) apresentaram uma diversidade genética menor (tabela 10), assim como o número de haplótipos (tabela 11) quando comparadas com as outras populações. Observando a diferença entre o tamanho da amostra com o número de haplótipos, entende-se que a quantidade de haplótipos compartilhados entre indivíduos foi alta dentro dessas tribos, mostrando uma falta de interação genética com outros povos.

5.3 Relação de haplogrupos distribuídos nos territórios brasileiros

5.3.1 SUDESTE

A região sudeste contou com 1621 haplótipos obtidos de oito populações oriundas de São Paulo, Minas Gerais, Rio de Janeiro e Espírito Santo, conforme descrito anteriormente na tabela 1. Os haplogrupos predominantes foram: os de

origem africana L3 (17,8%), L2 (12,1%), L1 (9,7%), ameríndia B (9,4%), A (8,9%), C (7,1%), e europeia H (8,6%). Os haplogrupos de origem africana foram os de maior número em relação àquelas de ancestralidades nativa americana e europeia, que se apresentaram em níveis similares (tabela 12).

Tabela 12 – Quantidades de haplótipos classificados em haplogrupos da Região Sudeste avaliados na análise da literatura. (1) FRIDMAN et al.,2014; (2) BERNARDO et al., 2014; (3) CARDENA et al., 2013; (4) AMBROSIO et al., 2015; (5) ALVES-SILVA et al., 2000; (6) PRIETO, L. et al., 2011; (7) SANCHES, 2013; (8) PANETO, 2010.

Haplogrupos	1	2	3	4	5	6	7	8	Total	Porcentual
A	28	10	50	11	14	13	7	11	144	8,90%
B	42	18	46	4	10	15	4	14	153	9,40%
C	27	16	25	7	6	13	7	14	115	7,10%
D	10	4	12	4	4	4	5	2	45	2,80%
H	8	21	46	11	14	6	11	22	139	8,60%
HV	6	0	9	7	0	0	7	1	30	1,90%
J	5	1	9	1	1	5	1	5	28	1,70%
K	6	3	10	2	2	6	2	4	35	2,20%
L0	9	9	18	4	0	7	4	6	57	3,50%
L1	28	20	56	11	11	12	11	9	158	9,70%
L2	23	25	67	13	8	24	13	23	196	12,10%
L3	40	54	87	15	13	33	15	31	288	17,80%
R	29	0	0	0	0	3	0	0	32	2,00%
R0	0	0	19	0	0	25	0	0	44	2,70%
T	8	1	9	5	4	9	5	4	45	2,80%
U	11	1	16	3	5	8	2	4	50	3,10%
Indeterminados	1	0	0	0	0	0	0	0	1	0,10%
Outros	9	7	13	2	7	7	6	10	61	4%
Total	290	190	492	100	99	190	100	160	1621	100,00%

Fonte: Autora.

O predomínio de haplogrupos de origem africana pode ser explicado, historicamente, pela vinda de grupos europeus, compostos principalmente por homens portugueses, que no início da colonização relacionaram-se com mulheres indígenas e, subsequentemente com as africanas trazidas para a escravidão

(ANTUNES, [entre 1998 e 2008]). Isto pode ser confirmado por estudos de patrilinearidade (cromossomo Y) que demonstram um predomínio de haplogrupos europeus (56%), principalmente de Portugal, na composição dessas populações brasileiras (PENA, et al., 2000). Uma maior miscigenação com mulheres europeias começou no Brasil a partir da metade do século 19, com a abertura dos portos, e com a chegada de vários imigrantes europeus (ANTUNES, [entre 1998 e 2008]; PENA, et al., 2000).

5.3.2 SUL

A região Sul apresentou 421 haplótipos na literatura sendo que a maior parte dos haplogrupos foi de ancestralidade europeia: H (32,1%), U (8,6%), T (8,1%), seguida dos nativo-americanos: C (8,1%), B (5,0%) e A (3,8%) e, uma pequena parcela de africanos, com destaque para L1 (3,6%) (tabela 13).

Tabela 13 - Quantidades de haplótipos classificados em haplogrupos da Região Sul avaliados na análise da literatura. (1) ALVES-SILVA et al., 2000; (2) MARRERO et al., 2005; (3) PALENCIA et al., 2010; (4) GUERREIRO-JUNIOR et al., 2009.

Haplogrupos	1	2	3	4	Total	Porcentual
A	3	1	2	10	16	3,80%
B	3	3	7	8	21	5,00%
C	3	7	7	17	34	8,10%
D	2	1	1	1	5	1,20%
H	13	50	22	50	135	32,10%
HV	1	0	0	0	1	0,20%
J	6	9	1	15	31	7,40%
K	2	8	5	8	23	5,50%
L0	0	0	2	5	7	1,70%
L1	3	1	3	8	15	3,60%
L2	0	1	6	1	8	1,90%
L3	2	1	1	1	5	1,20%
N	0	0	2	2	4	1,00%
pré-HV ou I	1	2	7	1	11	2,60%
pré-V ou HV0	1	0	3	0	4	1,00%
T	4	7	7	16	34	8,10%
U	4	11	4	17	36	8,60%
V	1	4	0	8	13	3,10%
X	1	0	0	1	2	0,50%
Indeterminado	0	13	0	3	16	3,80%
Total	50	119	80	172	421	100,00%

Fonte: Autora.

Historicamente, com a abertura dos portos em 1808, o Sul foi a região de maior concentração de imigrantes europeus, tanto homens quanto mulheres, principalmente alemães, eslavos, italianos e suíços-alemães. Nessa região, o relacionamento dos homens europeus com as mulheres indígenas foi maior do que com as africanas, pois o tráfico de escravos teve maior intensidade nas regiões Sudeste e Nordeste do que na região Sul (ANTUNES, [entre 1998 e 2008]). Além disso, a coroa portuguesa estimulou fortemente o casamento com mulheres

“brancas”, a partir do ano de 1755, com conseqüente aumento da vinda de mulheres europeias para o Brasil (PENA, et al., 2000).

5.3.3. NORDESTE

Poucos estudos sobre a região Nordeste foram encontrados (tabela 1), apresentando-se apenas 217 haplótipos. Observou-se a prevalência dos haplogrupos africanos L3 (17,1%), L2 (11,5%), L1 (15,2%), seguida de nativo americanos A (10,1%), B (9,2%), C (7,8%) e europeus H (12,9%) (tabela 14).

Tabela 14 - Quantidades de haplótipos classificados em haplogrupos da Região Nordeste avaliados na análise da literatura.

Haplogrupos	BARBOSA et al., 2008	ALVES-SILVA et al., 2000	Total	Porcentual
A	18	4	22	10,1%
B	17	3	20	9,2%
C	16	1	17	7,8%
D	4	3	7	3,2%
H	17	11	28	12,9%
J	2	1	3	1,4%
K	0	2	2	0,9%
L1	26	7	33	15,2%
L2	20	5	25	11,5%
L3	27	10	37	17,1%
pré-V ou HV0	1	0	1	0,5%
T	4	1	5	2,3%
U	7	1	8	3,7%
V	2	1	3	1,4%
W	2	0	2	0,9%
Indeterminado	4	0	4	1,8%
Total	167	50	217	100,0%

Fonte: Autora.

Assim como no Sudeste, o Nordeste teve um predomínio de vinda de escravos africanos. Houve nessa região um forte relacionamento entre os homens europeus com as mulheres indígenas e africanas, e entre os próprios ameríndios com os africanos (ANTUNES, [entre 1998 e 2008]). Dessa forma, analisando-se a linhagem materna nessa região, observa-se uma maioria de origem africana (44%) e uma minoria de europeus (~27%), enquanto que a linhagem paterna aponta a origem europeia como predominante (67%) (PENA, et al., 2000). A origem africana

supera a de indígenas, possivelmente devido à grande quantidade de escravos africanos em relação aos indígenas na região.

5.3.4 NORTE

A região Norte também apresentou baixo número de haplótipos (108) (tabela 15), sendo que mais de 50% representados por haplogrupos de ancestralidade nativo-americana: C (20%), A (15,7%), B (14,8%) e D (8,3%). Na ancestralidade africana, destacou-se o haplogrupo L3 (14,8%), enquanto que na europeia destacou-se o H (5,6%) e o J (4,6%).

Tabela 15 - Quantidades de haplótipos classificados em haplogrupos da Região Norte avaliados na análise da literatura.

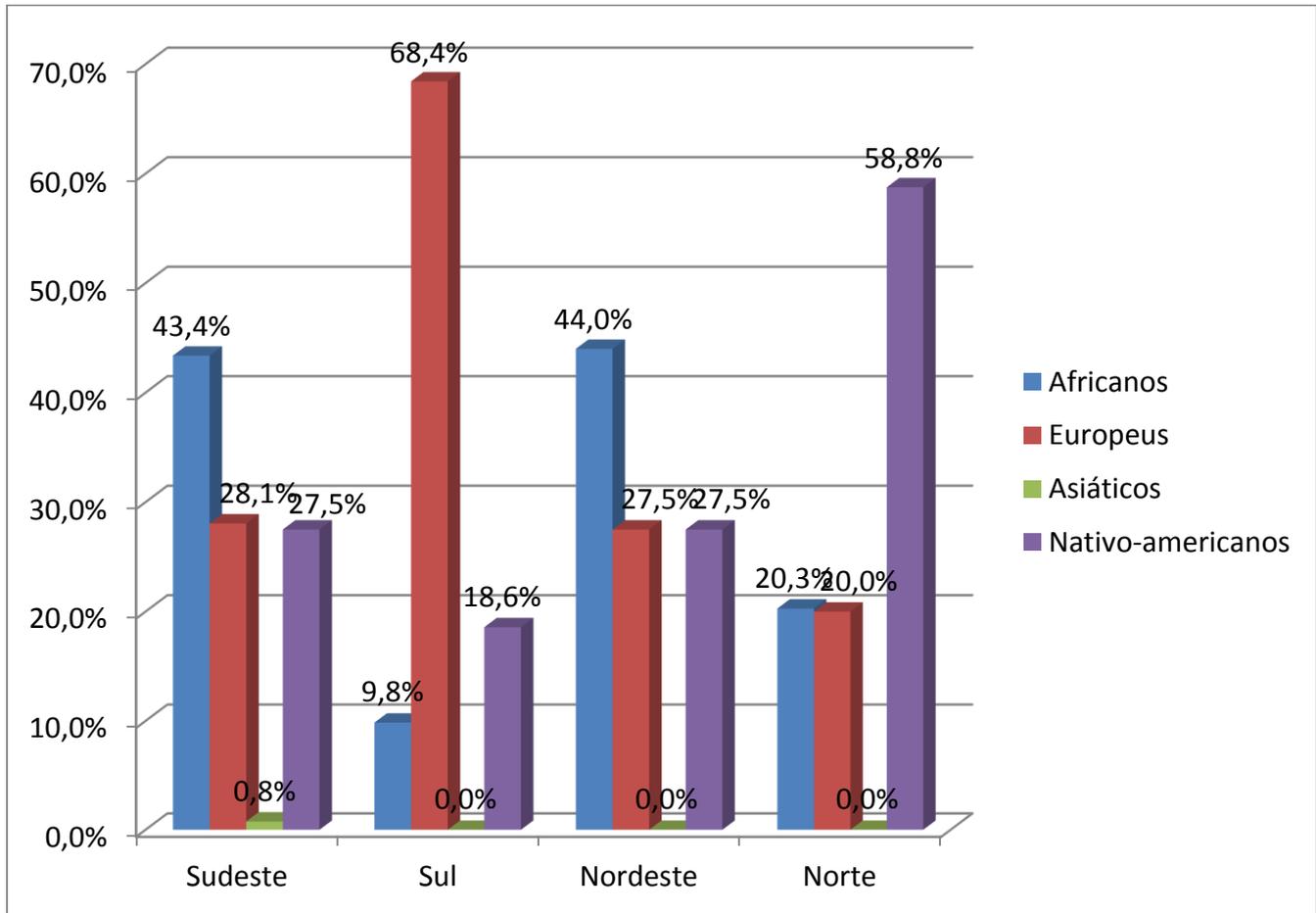
Haplogrupos	ALVES-SILVA et al., 2000	SANTOS et al., 2006	Total	Percentual
A	4	13	17	15,7%
B	8	8	16	14,8%
C	10	12	22	20,4%
D	4	5	9	8,3%
H	4	2	6	5,6%
J	3	2	5	4,6%
K	1	0	1	0,9%
L	1	0	1	0,9%
L0	0	1	1	0,9%
L1	2	0	2	1,9%
L2	1	3	4	3,7%
L3	3	13	16	14,8%
pre-V ou HV0	1	0	1	0,9%
T	4	0	4	3,7%
U	2	0	2	1,9%
Indeterminado	0	1	1	0,9%
Total	48	60	108	100,0%

Fonte: Autora.

Esta região demonstrou linhagem materna predominantemente nativo-americana (58,8%), seguida de africana e europeia com cerca de 20% cada. Isto pode ser explicado pelos relatos históricos das relações dos portugueses e escravos africanos com as americanas nativas (ANTUNES, [entre 1998 e 2008]). Desse modo, assim como nas outras regiões, houve o predomínio de haplogrupos europeus (56%) na patrilinhagem dessa região (PENA, et al., 2000).

O gráfico 5 mostra as proporções de ancestralidade para cada região, de acordo com os dados citados acima.

Gráfico 5 – Porcentagem da composição ancestral nas regiões brasileiras analisadas pelos dados da literatura.



Fonte: Autora.

Observando o gráfico acima, é possível visualizar a similaridade entre as regiões Sudeste e Nordeste. Isto provavelmente se deve ao grande fluxo migratório do Nordeste para o Sudeste, com maior intensidade nas décadas de 60 e 70, devido às melhores condições climáticas, de trabalho e de vida oferecidas pela última (ANTUNES, [entre 1998 e 2008]).

Comparando estes resultados com os de Pena e colaboradores (2000), é possível inferir que a linhagem materna do Brasil tem porcentagens de origens predominantemente africanas (37%), seguida de europeia (35%) e nativo-americana (28%), enquanto que a linhagem paterna tem, em sua maioria, a origem europeia

(57%). Estes dados corroboram com os relatos históricos das imigrações europeia e africana, e da relação desses imigrantes com os povos nativos.

5.3.5 COMPARAÇÃO ENTRE OS DADOS DO BRASIL NO EMPOP E DA LITERATURA

Na tabela 16 é demonstrada a relação entre as quantidades de dados brasileiros encontrados na literatura e que não estão, necessariamente, depositados em bancos de dados, em comparação com a quantidade de haplótipos no EMPOP.

Tabela 16 – Comparação dos dados brasileiros avaliados na literatura com os depositados no EMPOP

Regiões	Literatura	EMPOP
Sudeste	1621 (68,5%)	677 (86,6%)
Sul	421 (17,8%)	79 (10%)
Nordeste	217 (9,2%)	20 (2,6%)
Norte	108 (4,6%)	-
Indeterminado	-	6 (0,8%)
Total	2367	782

Fonte: Autora.

A tabela 16 permite observar que as quantidades de haplótipos encontrados na literatura superam mais que o dobro as quantidades do EMPOP. Isto pode refletir a dificuldade dos laboratórios em atender aos controles de qualidade exigidos pelo EMPOP para submissão das amostras. É possível observar que as quantidades de dados do Sudeste superam aos das outras regiões, demonstrando uma carência de dados e de estudos sobre as populações de outras regiões que estão pouco representadas.

5.4 Distâncias genéticas das populações brasileiras

Foram realizadas 253 análises de comparação pareadas entre as 23 populações estudadas. Sabendo que os valores de *p value* obtidos estão superestimados, estes resultados foram submetidos à correção de Bonferroni para múltiplos testes, através da divisão do nível de significância de 5% pelo número total de análises (253). Dessa forma, as distâncias genéticas (*Fst*) são significativas para valores de *p value* menores que 0,000198.

5.4.1 SUDESTE

Foram estudadas seis populações de São Paulo, sendo uma composta somente por homens autodeclarados pretos, e as outras oriundas de Minas Gerais (caucasianos), Espírito Santo e Rio de Janeiro.

Todas as populações se demonstraram próximas geneticamente, principalmente entre as de São Paulo e estas com a do Espírito Santo e de Minas Gerais (tabela 17). A população do Rio de Janeiro se demonstrou mais próxima de duas populações de São Paulo, a de Cardena (2013) e a de indivíduos pretos. A população de São Paulo de indivíduos autodeclarados pretos se mostrou mais afastada das demais populações e apresentou valores significativos para o *p value*. Também demonstraram distâncias significantes as populações de Minas Gerais (MG) e Espírito Santo, e Rio de Janeiro com a maioria dessas populações, exceto SP2 e SP pretos (tabela 17).

Tabela 17 – Valores de *Fst* e *p value* das distâncias genéticas entre as populações. Valores de *Fst* estão representados abaixo do eixo diagonal. Valores de *p value* estão representados acima do eixo diagonal. Valores estatisticamente significantes para *p value* < 0,000198 (Correção de Bonferroni para múltiplos testes). SP – São Paulo: 1 e 2 (PRIETO et al., 2011), 3 (PANETO, 2010), 4 (FRIDMAN et al., 2014), 5 (CARDENA et al., 2013); SP pretos – população autodeclarada preta de São Paulo (GONÇALVES et al., 2008); MG – Minas Gerais (ALVES-SILVA et al., 2000); ES - Espírito Santo (SANCHES, 2013); RJ - Rio de Janeiro (BERNARDO et al., 2014).

População	SP 1	SP 2	SP 3	SP 4	SP 5	MG	ES	SP pretos	RJ
SP 1	*	0,74775	0,99099	0,45045	0,74775	0,11712	0,12613	0,0000	0,0000
SP 2	0,00103	*	0,7027	0,59459	0,48649	0,73874	0,33333	0,0000	0,01802
SP 3	0,00606	0,00079	*	0,41441	0,54054	0,07207	0,18018	0,0000	0,0000
SP 4	0,00003	0,00053	0,00001	*	0,21622	0,02703	0,17117	0,0000	0,0000
SP 5	0,0003	0,00011	0,00016	0,00028	*	0,02703	0,07207	0,0000	0,0000
MG	0,0011	0,00098	0,00155	0,00175	0,00172	*	0,0000	0,0000	0,0000
ES	0,00096	0,00043	0,00064	0,00083	0,00106	0,00336	*	0,0000	0,0000
SP pretos	0,00489	0,00531	0,00476	0,00505	0,00328	0,00861	0,00483	*	0,03604
RJ	0,00285	0,00263	0,00233	0,00285	0,00172	0,0064	0,00205	0,00131	*

Fonte: Autora.

5.4.2 SUL

Todas as populações do sul também se demonstraram próximas geneticamente. General (RS), Santa Catarina, Porto Alegre e Veranópolis apresentam uma maior proximidade entre si. A população mesclada (RS, SC e PR) ficou mais próxima de General e Porto Alegre. Podem ser observados valores significativos de *p value* entre as populações de Veranópolis e Porto Alegre, e entre a população misturada (RS, SC e PR) com Veranópolis, Santa Catarina e Porto Alegre (tabela 18).

Tabela 18 – Valores de *Fst* e *p value* das distâncias genéticas entre as populações sulistas. Valores de *Fst* estão representados abaixo do eixo diagonal. Valores de *p value* estão representados acima do eixo diagonal. Valores estatisticamente significantes para *p value* < 0,000198 (Correção de Bonferroni para múltiplos testes). RS – Rio Grande do Sul; SC – Santa Catarina.

Referência	População	Veranópolis -	General	SC	Porto	RS, SC e
		RS	- RS		Alegre - RS	Parana
MARRERO et al., 2005	Veranópolis - RS	*	0,00901	0,00901	0,0000	0,0000
	General - RS	0,0261	*	0,11712	0,06306	0,00901
PALENCIA et al., 2010	SC	0,01158	0,00594	*	0,05405	0,0000
GUERREIRO-JUNIOR et al., 2009	Porto Alegre - RS	0,0138	0,00522	0,00453	*	0,0000
ALVES-SILVA et al., 2000	RS, SC e Paraná	0,04615	0,01057	0,02383	0,01643	*

Fonte: Autora

5.4.3 NORDESTE

As duas populações comparadas de Alagoas e Pernambuco (ALVES-SILVA et al., 2000; BARBOSA, et al., 2008) apresentaram os valores de *Fst* e *p value* correspondente à 0,00080 e 0,23423, respectivamente. As duas populações são geneticamente parecidas.

5.4.4 NORTE

Todas as populações demonstraram-se muito próximas e com valores de *p value* significativos de todas as populações com a de indivíduos autodeclarados pretos do Amazonas. (tabela 19).

Tabela 19 – Valores de *Fst* e *p value* das distâncias genéticas entre as populações do Norte. Valores de *Fst* estão representados abaixo do eixo diagonal. Valores de *p value* estão representados acima do eixo diagonal. Valores estatisticamente significantes para *p value* < 0,000198 (Correção de Bonferroni para múltiplos testes). PA – Pará; AM – Amazonas; AM pretos – população de indivíduos pretos da Amazonas; RO – Rondônia.

Referência	População	PA	AM, PA, RO e Acre	AM
SANTOS et al., 2006	PA	*	0,01802	0,000
ALVES-SILVA et al., 2000	AM, PA, RO e Acre	0,00723	*	0,000
SANTOS et al., 2007	AM	0,00734	0,00706	*

Fonte: Autora

5.4.5 DISTÂNCIA GENÉTICA DAS POPULAÇÕES

Todas as populações brasileiras apresentaram-se geneticamente próximas entre si e com as estrangeiras, exceto pelas tribos Guarani e Kaingang, que tiveram as maiores distâncias com *p value* significativos com as outras populações. Comparando essas populações nativo-americanas com as demais, Kaingang teve as menores distâncias (*Fst*) com o Guarani (0,09335), com o Pará (0,09353) e com São Paulo – SP 5 (0,09963); já a tribo Guarani apresentou menores valores de *Fst* com o Pará (0,09933) , além de Kaingang.

A população de São Paulo (SP 5) se demonstrou muito próxima geneticamente de Alagoas (0,00028) e Pernambuco (0,00041). Todas as populações do sudeste tiveram distâncias maiores quando comparadas com Veranópolis e com a população misturada do Rio Grande do Sul, Santa Catarina e Paraná (ALVES-SILVA et al., 2000). Já as populações de pretos de São Paulo e a do Rio de Janeiro ficaram próximas entre si (0,00131) e com Pernambuco (SP – 0,00390 e RJ – 0,00171); ambas tiveram distâncias maiores com Santa Catarina e Porto Alegre, além das citadas anteriormente.

Portugal apresentou menores distâncias com Santa Catarina (0,00146), São Paulo-SP 2 (0,00203), Minas Gerais (0,00205) e Porto Alegre (0,00250). A África apresentou os menores valores de *Fst* com as populações do Rio de Janeiro (0,00769), Espírito Santo (0,00814) e Pernambuco (0,00842).

Os valores da distância genética e do *p value* entre todas as populações podem ser observados na tabela 20.

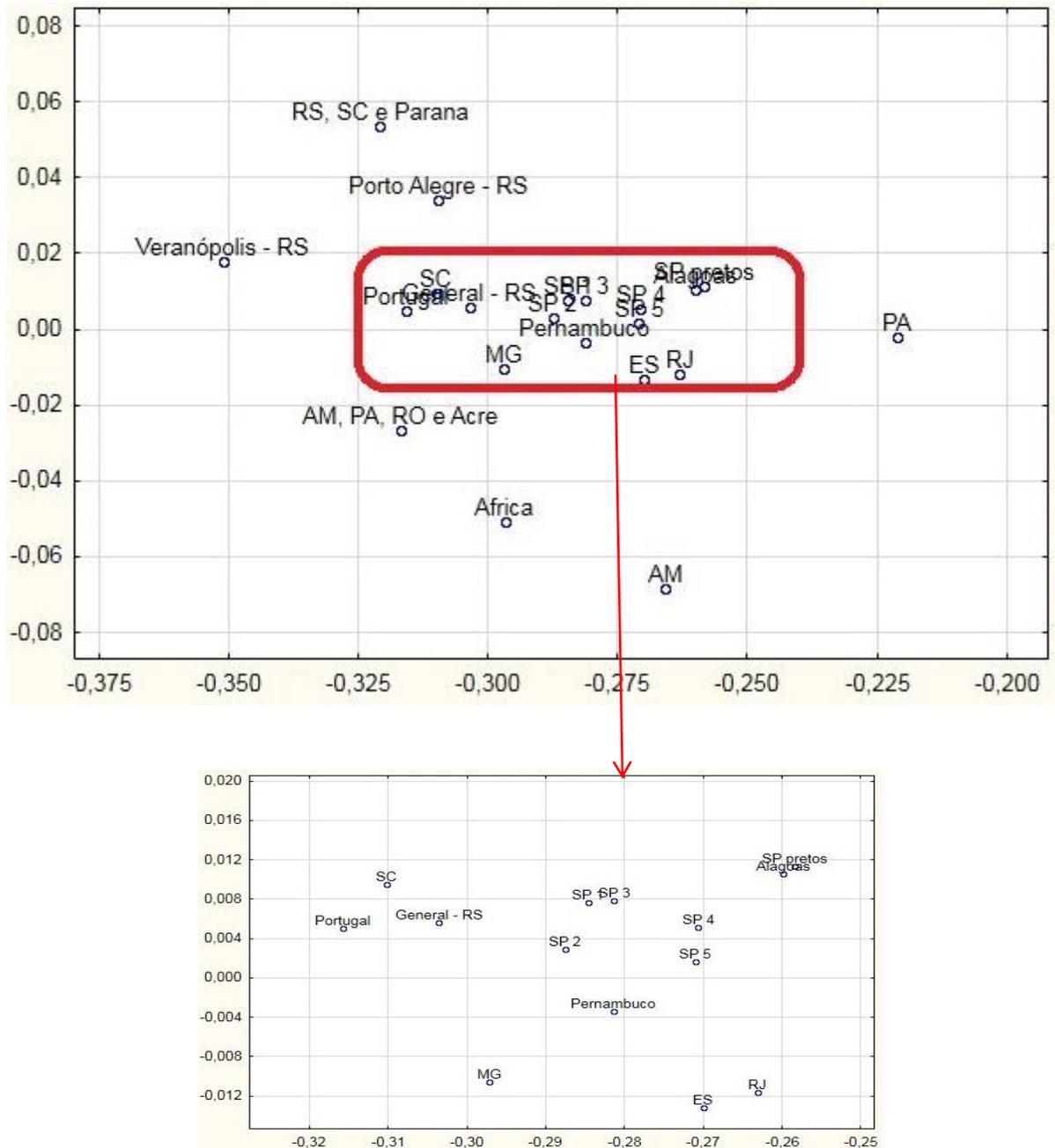
Tabela 20 – Valores de *Fst* e *p value* das distâncias genéticas entre as populações estudadas. Valores de *Fst* estão representados abaixo do eixo diagonal. Valores de *p value* estão representados acima do eixo diagonal. Como relatado anteriormente, foi aplicada a correção de Bonferroni para as 253 análises de distâncias pareadas. Valores estatisticamente significantes para *p value* < 0,000198 (Correção de Bonferroni para múltiplos testes). SP – São Paulo: 1 e 2 (PRIETO et al., 2011), 3 (PANETO, 2010), 4 (FRIDMAN et al., 2014), 5 (CARDENA et al., 2013); SP pretos – população autodeclarada preta de São Paulo (GONÇALVES et al., 2008); MG – Minas Gerais (ALVES-SILVA et al., 2000); ES - Espírito Santo (SANCHES, 2013); RJ - Rio de Janeiro (BERNARDO et al., 2014); RS – Rio Grande do Sul; SC – Santa Catarina; PA – Pará; AM – Amazonas; AM pretos – população autodeclarada preta do Amazonas ; RO – Rondônia.

	SP 1	SP 2	SP 3	SP 4	SP 5	MG	ES	SP pretos	RJ	Veranópolis - RS	General - RS	SC	Porto Alegre - RS	RS, SC e Parana	PA	AM, PA, RO e Acre	AM	Alagoas	Pernambuco	Portugal	Africa	Guarani	Kaingang
SP 1	0	0,75676	0,99099	0,4685	0,7568	0,10811	0,1441	0,00000	0,00000	0,00000	0,42342	0,018	0,018	0,00000	0,009	0,02703	0,00000	0,36937	0,53153	0,01802	0,00000	0,00000	0,00000
SP 2	-0,001	0	0,72072	0,6577	0,4865	0,66667	0,3423	0,00000	0,02703	0,00901	0,36937	0,0631	0,2072	0,00000	0,0631	0,16216	0,009	0,42342	0,21622	0,18018	0,00000	0,00000	0,00000
SP 3	-0,0061	-0,0008	0	0,3604	0,6216	0,09009	0,1712	0,00000	0,00000	0,00000	0,27928	0,00000	0,009	0,00000	0,00000	0,03604	0,00000	0,31532	0,4955	0,00901	0,00000	0,00000	0,00000
SP 4	0,00003	-0,0005	0,00001	0	0,0811	0,03604	0,1171	0,00000	0,00000	0,00000	0,10811	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,01802	0,04505	0,00000	0,00000	0,00000	0,00000
SP 5	-0,0003	-0,0001	-0,00016	0,0003	0	0,01802	0,036	0,00000	0,00000	0,00000	0,24324	0,00000	0,00000	0,00000	0,00000	0,00901	0,00000	0,22523	0,22525	0,00000	0,00000	0,00000	0,00000
MG	0,0011	-0,001	0,00155	0,0018	0,0017	0	0,009	0,00000	0,00000	0,00901	0,16216	0,018	0,009	0,00000	0,00000	0,56757	0,00000	0,00901	0,0991	0,09009	0,00000	0,00000	0,00000
ES	0,00096	0,00043	0,00064	0,0008	0,0011	0,00336	0	0,00000	0,00000	0,00000	0,13514	0,00000	0,00000	0,00000	0,0541	0,00901	0,00000	0,03604	0,18018	0,00901	0,00000	0,00000	0,00000
SP pretos	0,00489	0,00531	0,00476	0,0051	0,0033	0,00861	0,0048	0	0,00000	0,00000	0,02703	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,02703	0,00000	0,00000	0,00000	0,00000
RJ	0,00285	0,00263	0,00233	0,0029	0,0017	0,0064	0,0021	0,00131	0	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,07207	0,00000	0,00000	0,00000	0,00000
Veranópolis - RS	0,02308	0,0197	0,02459	0,0244	0,026	0,01687	0,0296	0,04216	0,03865	0	0,01802	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
General - RS	0,00007	0,00073	0,00061	0,0023	0,0011	0,00223	0,0018	0,00573	0,00336	0,0261	0	0,0811	0,0811	0,00000	0,0451	0,0991	0,00000	0,05405	0,58559	0,12613	0,02703	0,00000	0,00000
SC	0,00628	0,00429	0,00722	0,0079	0,008	0,00418	0,0077	0,01849	0,01624	0,01158	0,00594	0	0,0451	0,00000	0,00000	0,01802	0,00000	0,00000	0,00901	0,18919	0,00000	0,00000	0,00000
Porto Alegre - RS	0,00303	0,00124	0,00355	0,0043	0,0054	0,00262	0,006	0,01392	0,01082	0,0138	0,00522	0,0045	0	0,00000	0,018	0,00901	0,00000	0,00000	0,00901	0,03604	0,00000	0,00000	0,00000
RS, SC e Parana	0,01027	0,01046	0,01017	0,0107	0,0102	0,01312	0,0099	0,01273	0,00997	0,04615	0,01057	0,0238	0,0164	0	0,00000	0,00901	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
PA	0,00511	0,00399	0,00492	0,0057	0,0046	0,00793	0,0032	0,00779	0,00547	0,03525	0,00583	0,0151	0,0099	0,01252	0	0,00901	0,00000	0,01802	0,16216	0,00000	0,00000	0,00000	0,00000
AM, PA, RO e Acre	0,00394	0,00229	0,00433	0,0042	0,0042	-0,001	0,0044	0,0111	0,00762	0,0241	0,00504	0,008	0,0059	0,01058	0,0072	0	0,00000	0,00901	0,11712	0,02703	0,00000	0,00000	0,00000
AM pretos	0,00719	0,00715	0,00678	0,0069	0,0059	0,00785	0,0059	0,00765	0,00472	0,04142	0,00785	0,0191	0,0141	0,01266	0,0073	0,00706	0	0,00000	0,00901	0,00000	0,00000	0,00000	0,00000
Alagoas	0,00005	0,00013	0,00014	0,0009	0,0003	0,00293	0,0015	0,00488	0,00209	0,02788	0,00289	0,0102	0,0073	0,01133	0,004	0,00518	0,0066	0	0,27928	0,00000	0,00000	0,00000	0,00000
Pernambuco	-0,0005	0,00141	-0,00017	0,0021	0,0004	0,00302	0,0012	0,0039	0,00171	0,03103	-0,0004	0,0115	0,0082	0,00988	0,0026	0,00318	0,0058	0,0008	0	0,01802	0,00000	0,00000	0,00000
Portugal	0,00456	0,00203	0,00542	0,0057	0,0059	0,00205	0,0074	0,01596	0,01294	0,01055	0,00566	0,0015	0,0025	0,01863	0,0141	0,00564	0,0163	0,00719	0,00846	0	0,00000	0,00000	0,00000
Africa	0,00973	0,01014	0,00958	0,0099	0,0092	0,01241	0,0081	0,00996	0,00769	0,04418	0,0089	0,0216	0,0178	0,01625	0,0137	0,01426	0,0118	0,00997	0,00842	0,01784	0,00000	0,00000	0,00000
Guarani	0,11752	0,12504	0,117	0,1138	0,1075	0,12316	0,1155	0,12308	0,11661	0,16189	0,13205	0,1356	0,1285	0,12649	0,0993	0,12909	0,1155	0,1054	0,12191	0,13331	0,1254	0	0,00000
Kaingang	0,10768	0,11381	0,10715	0,104	0,0996	0,11369	0,1039	0,11097	0,10536	0,15091	0,12472	0,1244	0,1173	0,11838	0,0935	0,11789	0,1063	0,09736	0,11308	0,12178	0,1144	0,09335	0

Fonte: Autora.

As distâncias genéticas de todas as populações estão apresentadas no gráfico 6. As populações nativo-americanas dos Guaranis e dos Kaingangs (MARRERO et al., 2006) se mantiveram muito afastadas das demais e, portanto, não foram demonstradas.

Gráfico 6 – Distâncias genéticas (F_{st}) das populações brasileiras e estrangeiras. As distâncias foram submetidas ao *software* Statistica v. 13.2, sendo configuradas para uma visualização em duas dimensões (eixo x e eixo y). SP – São Paulo: 1 e 2 (PRIETO et al., 2011), 3 (PANETO, 2010), 4 (FRIDMAN et al., 2014), 5 (CARDENA et al., 2013); SP pretos – população autodeclarada preta de São Paulo; MG – Minas Gerais; ES - Espírito Santo; RJ - Rio de Janeiro; RS – Rio Grande do Sul; SC – Santa Catarina; PA – Pará; AM (Amazonas) – População de indivíduos pretos do Amazonas; RO – Rondônia.



Fonte: da autora.

No quadrante superior esquerdo do gráfico 6, observa-se uma proximidade de Veranópolis (RS) e Porto Alegre (RS) com a população misturada de Santa Catarina, Rio Grande do Sul e Paraná.

O gráfico permitiu observar a proximidade da população africana com a população de autodeclarados pretos na Amazônia (AM) e da população misturada do Amazonas, Pará, Rondônia e Acre.

A região central destacada do gráfico 6 englobou todas as populações do sudeste e nordeste, e um pouco do sul. Observa-se uma maior proximidade: dos autodeclarados pretos de São Paulo com Alagoas; de Santa Catarina e General (RS) com Portugal; e das populações de São Paulo 1 e 3 (PANETO, 2010; PRIETO et al., 2011). A proximidade entre as populações de pretos de São Paulo e de Alagoas reforça a hipótese da alta influência genética na primeira gerada pelos movimentos migratórios massivos do Nordeste para o Sudeste.

5.5 Sequenciamento do mtDNA em amostras referência

As amostras foram amplificadas e sequenciadas, confirmando-se a sequência de mtDNA. Foi possível classificar todas as sequências em haplogrupos, conforme apresentado na tabela 20.

Tabela 21 – Sequências analisadas e seus respectivos haplogrupos

Amostras	Sequência	Haplogrupo
2800	152C 263G 315.1C 477C	H1c
9948	73G 152C 315.1C 16126C 16163G 16186T 16189C 16294T	T1
M1	263G 315.1C	H
M2	150T 239C 263G 315.1C 16362C	H6
M3	73G 263G 315.1C 497T 523DEL 524DEL 16224C 16311C	K1a

Fonte: Autora.

Todos os haplogrupos corresponderam à ancestralidade europeia.

Nos eletroferogramas visualizaram-se picos bem definidos e com sinal de intensidade adequada. As sequências apresentaram linhas de base com ruídos, mas que não influenciaram nas análises.

Os controles negativos não demonstraram presença de outros mtDNAs, assegurando que todas os reagentes e equipamentos utilizados estavam livres de contaminação.

Essas análises permitiram compreender como uma metodologia complexa, porém padronizada em suas várias etapas, pode ser realizada e fornecer resultados satisfatórios, mesmo que o operador não tenha muita experiência técnica, mas com supervisão pode obter os dados necessários para avaliação dos haplótipos.

6 CONCLUSÕES

- A América do Norte e a Europa são os continentes com maior número de haplótipos depositados no EMPOP. A América do Sul contribui com 9% dos dados mundiais, dos quais 5% correspondem a amostras brasileiras, ou seja, 2,26% de todos os dados do banco, um número bastante reduzido.
- A quantidade de dados no banco, assim como as regiões analisadas são importantes, pois a diversidade genética e o número de haplótipos estão diretamente relacionados com o tamanho amostral e poder de discriminação, sendo a região controle a mais polimórfica, seguida de HV1 + HV2, HV2 e HV3, nessa ordem.
- As populações brasileiras demonstraram altos valores de diversidade genética, o que aponta a grande miscigenação do país.
- Muitos estudos diferiram quanto à escolha da região a ser analisada e, portanto, padronizar o sequenciamento de toda a região controle é o mais adequado, seja para estudos populacionais quanto para a deposição de dados no banco.
- As distâncias genéticas demonstraram que populações brasileiras estão muito próximas. O Sudeste e o Nordeste demonstraram uma predominância de ancestralidade africana; o Norte apresentou ancestralidade nativo americana e o Sul europeia, o que está de acordo com outros estudos e relatos históricos.
- A região Centro-Oeste ainda carece de mais estudos, assim como a Norte e a Nordeste que apresentaram poucos trabalhos publicados sobre o assunto.
- Embora, neste estudo, não foi possível encontrar uma equação que permita estabelecer um valor ideal de quantidade de amostras de mtDNA para estabelecer-se um banco de dados, pode-se inferir que, embora o Brasil apresente 24,6% dos dados sul americanos, 782 amostras é um número muito pequeno para representar de forma confiável uma população com mais de 200 milhões de habitantes. Pela literatura, há mais que o dobro de haplótipos (HV1) do total contido no EMPOP. Além disso, cada região brasileira possui haplótipos característicos na composição de sua população, e os dados brasileiros do EMPOP representam muito mais a população do

sudeste do que as outras, sendo 677 haplótipos da região Sudeste, 79 da região Sul, 20 da região Nordeste e 6 não especificados.

- A pequena participação do Brasil no EMPOP pode estar relacionada a vários fatores, como: introdução relativamente recente da análise do mtDNA no país; alto custo de equipamentos e reagentes; relativa complexidade da metodologia quando comparada a outros métodos de identificação humana; e, principalmente, as dificuldades dos estudos em atender aos controles de qualidade exigidos para submissão das amostras ao EMPOP.
- A metodologia utilizada neste estudo permitiu obter as sequências corretas das regiões HV1, HV2 e HV3, livre de contaminações ou outras interferências técnicas. Também demonstrou que mesmo sendo complexa, sua padronização permite que um operador com pouca experiência na técnica, mas com orientação adequada, possa obter resultados satisfatórios.

REFERÊNCIAS

AGENDA Cultural Brasil. Disponível em:

<<http://agendaculturaldobrasil.blogspot.com.br/>>. Acesso em: 17 jan. 2017.

ALVES-SILVA, J. et al. The ancestry of brazilian mtDNA lineages. **American Journal of Human Genetics**, v. 67, p. 444-461, 2000.

AMBROSIO, I. B. et al. Mitochondrial DNA 30-SNP data confirm high prevalence of African lineages in the population of Espirito Santo, Brazil. **Forensic Science International: Genetics Supplement Series**, v. 5, p. 346-347, 2015.

ANDERSON, S. et al. Sequence and organization of the human mitochondrial genome. **Nature**, v. 290, p. 457-465, 1981.

ANDREWS, R. et al. Reanalysis and revision of Cambridge reference sequence for human mitochondrial DNA. **Nature Genetics**, v. 23, n. 2, p. 147, 1999.

ANTUNES, V. L. da C. Movimentos migratórios. In:_____. **Geografia do Brasil: quadro natural e humano**. [S.l.]: CERED, [entre 1998 e 2008]. Cap. 10, p. 272-277.

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, p. 142-146, 2008.

BENFICA, F. S.; VAZ, M. Identidade e identificação. In:_____. **Medicina legal**. 3. ed. Porto Alegre: Livraria do Advogado, 2015. Cap. 2, p. 21-32.

BERNARDO, S. et al. mtDNA ancestry of Rio de Janeiro population, Brazil. **Molecular Biology Reports**, v. 41, p. 1945-1950, 2014.

BUDOWLE, B. et al. Mitochondrial DNA regions HVI and HVII population data. **Forensic Science International**, v. 103, p. 23-35, 1999.

BUTLER, J. M. Mitochondrial DNA analysis. In:_____. **Advanced topics in forensic DNA typing: methodology**. Elsevier Academic Press, 2011. Chap. 14, p. 405-456.

CANN, R. L. Mitochondria and evolution. In: SMELSER, N. J.; BALTES, P.B. (Ed.). **International encyclopedia of the social & behavioral sciences**. Amsterdam: Elsevier, 2001. p. 9906-9910.

CANN, R. L.; STONEKING, M.; WILSON, A. C. Mitochondrial DNA and human evolution. **Nature**, v. 325, p. 31-36, 1987.

CARDENA, M. M. S. G. et al. Assessment of the relationship between self-declared ethnicity, mitochondrial haplogroups and genomic ancestry in brazilian individuals. **PLoS One**, v. 8, 2013. [doi: 10.1371/journal.pone.0062005](https://doi.org/10.1371/journal.pone.0062005).

CARRACEDO, A. et al. DNA Commission of the International Society for Forensic Genetics: guidelines for mitochondrial DNA typing. **Forensic Science International**, v. 110, p. 79-85, 2000.

COUNTRYMETERS. **População dos Estados Unidos da América**. Disponível em: <[http://countrymeters.info/pt/United_States_of_America_\(USA\)](http://countrymeters.info/pt/United_States_of_America_(USA))>. Acesso: 23 jul. 2016.

EUROPEAN DNA PROFILING GROUP. **EMPOP mtDNA database**. Version 3/R11. Disponível em: <<http://empop.online/populations>>. Acesso em: 20 set. 2015.

EUROPEAN DNA PROFILING GROUP. **EMPOP mtDNA database**. Version 3/R11. Disponível em: <<http://empop.online/>>. Acesso em: 17 jan. 2017.

FOSWIKI. **Human mtDNA migrations**. 2016. Disponível em: <<http://www.mitomap.org/foswiki/pub/MITOMAP/MitomapFigures/WorldMigrations2012.pdf>>. Acesso em: 17 jan. 2017.

FREITAS, E. de. Aspectos da população da Europa. **Mundo Educação**. Disponível em: <<http://mundoeducacao.bol.uol.com.br/geografia/aspectos-populacao-europa.htm>>. Acesso: 23 jul. 2016.

FRIDMAN, C. et al. Mitochondrial HVI and HVII polymorphisms and heteroplasmies inheritance in Brazilian pairs of mother/child. **Forensic Science International: Genetics**, v. 1, n. 1, p. 277-278, 2008.

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

GILES, R. E. et al. Maternal inheritance of human mitochondrial DNA. **Proceedings of the National Academy of Sciences of the United States of America**, v. 77, n. 11, p. 6715-6719, 1980.

GODINHO, N. M. O. Banco de dados de DNA: uma ferramenta a serviço da justiça. **Revista Brasileira de Estudos em Segurança Pública**, v. 7, n. 2, p. 20-30, 2014.

GONÇALVES, V. F. et al. The phylogeography of african brazilians. **Human Heredity**, v. 65, p. 23-32, 2008.

GREENBERG, B. D.; NEWBOLD, J. E.; SUGINO, A. Intraspecific nucleotide sequence variability surrounding the origin of replication in human mitochondrial DNA. **Gene**, v. 21, p. 33-49, 1983.

GUERREIRO-JUNIOR, V. et al. Genetic signatures of parental contribution in black and white populations in Brazil. **Genetics and Molecular Biology**, v. 32, p. 1-11, 2009.

HOLLAND, M. M.; PARSONS, T. J. Mitochondrial DNA sequence analysis – validation and use for forensic casework. **Forensic Science Review**, v. 11, n. 1, p. 21-50, 1999.

IVANOV, P. L. et al. Mitochondrial DNA sequence heteroplasmy in the Grand Duke of Russia Georgij Romanov establishes the authenticity of the remains of Tsar Nicholas II. **Nature Genetics**, v. 12, n. 4, p. 417-420, 1996.

LUTZ, H. et al. Is it possible to differentiate mtDNA by means of HVIII in samples that cannot be distinguished by sequencing the HVI and HVII regions? **Forensic Science International**, v. 113, p. 97-101, 2000.

MARRERO, A. R. et al. Heterogeneity of the genome ancestry of individuals classified as white in the state of Rio Grande do Sul, Brasil. **American Journal of Human Biology**, v. 17, p. 496-506, 2005.

MARRERO, A. R. et al. Demographic and evolutionary trajectories of the Guarani and Kaingang natives of Brazil. **American Journal of Physical Anthropology**, v. 132, p. 301-310, 2007.

MIKKELSEN, M. et al. Forensic and phylogeography characterization of mtDNA lineages from Somalia. **International Journal of Legal Medicine**, v. 126, p. 573-579, 2012.

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, p. 331-336, 2010.

PANETO, G. **Análise de polimorfismos do DNA mitocondrial em indivíduos residentes na Grande São Paulo para aplicação na identificação humana**. 2010. 120 f. Tese (Doutorado em Biociências e Biotecnologia Aplicadas à Farmácia) - Faculdade de Ciências Farmacêuticas, Universidade Estadual Paulista, Araraquara, 2010.

PARSONS, T. J.; COBLE, M. D. Increasing the forensic discrimination of mitochondrial DNA testing through analysis of the entire mitochondrial DNA genome. **Forensic Sciences**, v. 42, n. 3, p. 304-309, 2001.

PARSON, W. et al. The EDNAP mitochondrial DNA database (EMPOP) collaborative exercises: organization, results and perspectives. **Forensic Science International**, v. 139, p. 215-226, 2004.

PARSON, W. et al. DNA commission of the International Society for Forensic Genetics: revised and extend guidelines for mitochondrial DNA typing. **Forensic Science International: Genetics**, v.13, p. 134-142, 2014.

PENA, S. D. J. et al. Retrato molecular do Brasil. **Ciência Hoje: Genética**, v. 27, n. 159, p. 16-25, 2000.

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, p. 132-136, 2004.

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

SALAS, A.; LAREU, M. V.; CARRACEDO, A. Heteroplasmy in mtDNA and the weight of evidence in forensic mtDNA analysis: a case report. **International Journal of Legal Medicine**, v. 114, p. 186-190, 2001.

SANCHES, N. M. **Análise de polimorfismos da região controle do DNA mitocondrial em indivíduos nascidos e residentes no estado do Espírito Santo para utilização na identificação humana**. 2013. 83 f. Dissertação (Mestrado em Biociências e Biotecnologia Aplicadas à Farmácia) - Faculdade de Ciências Farmacêuticas, Universidade Estadual Paulista, Araraquara, 2013.

SANTOS, A. C. F. dos et al. Nucleotide variability of HV-I in admixed population of the Brazilian Amazon Region. **Forensic Science International**, v. 164, p. 276-277, 2006.

SANTOS, A. K. R. dos et al. Nucleotide variability of HV-I in Afro-descendants populations of the Brazilian Amazon Region. **Forensic Science International**, v. 167, p. 77-80, 2007.

SANTOS, V. dos. Mitocôndria. **Escola Kids**. Disponível em: <<http://escolakids.uol.com.br/mitocondrias.htm>>. Acesso em: 17 jan. 2017.

SCIENTIFIC WORKING GROUP ON DNA ANALYSIS METHODS. **Interpretation guidelines for mitochondrial DNA analysis by forensic DNA testing laboratories**. 2013. Disponível em: <http://media.wix.com/ugd/4344b0_c5e20877c02f403c9ba16770e8d41937.pdf>. Acesso em: 23 nov. 2016.

SHOKOLENKO, I. N.; ALEXEYEV, M. F. Mitochondrial DNA: a disposable genome? **Biochimica et Biophysica Acta: Molecular bases of Disease**, v. 1852, n. 9, p. 1805-1809, 2015.

SOUZA, A. A. e. **Noções gerais de antropologia forense e identificação**. maio 2013. Disponível em: <<http://profalessandraareas.blogspot.com.br/2013/05/noco-es-gerais-de-antropologia-forense-e.html>>. Acesso em: 17 jan. 2017.

SULLIVAN, K. M.; HOPGOOD, R.; GILL, P. Identification of human remains by amplification and automated sequencing of mitochondrial DNA. **International Journal of Legal Medicine**, v. 105, p. 83-86, 1992.

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.

UNITED STATES DEPARTMENT OF AGRICULTURE. Beltsville Agricultural Research Center. **Single nucleotide polymorphisms**. 2007. Disponível em: <<https://sgil.ba.ars.usda.gov/cregan/snp.html>>. Acesso em: 17 ago. 2016.

VAN OVEN, M. **Phylotree.org – mtDNA tree build17**. 2016. Disponível em: <<http://www.phylotree.org/tree/index.htm>>. Acesso em: 17 ago. 2016.

VAN OVEN, M.; KAYSER, M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. **Human Mutation**, v. 30, p. 386-394, 2008.

YONSEI DNA PROFILING GROUP. **mtDNAMANAGER**. 2007. Disponível em: <http://mtmanager.yonsei.ac.kr/search_sample.php>. Acesso em: 17 ago. 2016.